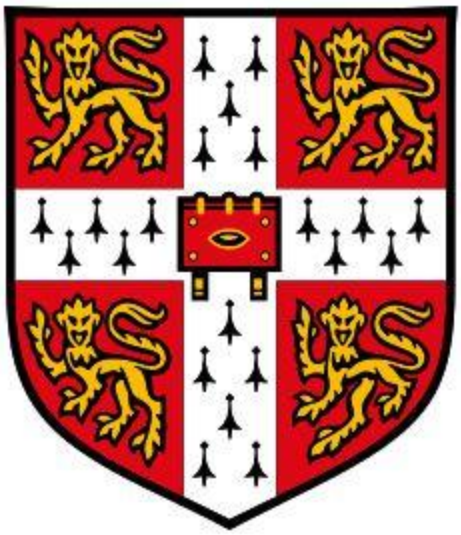
Investigation of natural genetic modifiers of meiotic crossover frequency in *Arabidopsis thaliana*



Catherine Helen Griffin

Christs College

University of Cambridge

This dissertation is submitted for the degree of *Doctor of Philosophy*

September 2017

Name: Catherine Helen Griffin

Title: Investigation of natural genetic modifiers of meiotic crossover frequency in *Arabidopsis thaliana*

Summary:

Meiotic recombination, known as crossover, is a vital mechanism for generating genetic diversity in sexually reproducing populations. Recombination events are non-uniform across the genome, due to a variety of influences including chromatin structure, DNA-sequence, epigenetic marks and interference from other recombination events. These known factors do not fully explain the distribution of recombination events, and additionally do not account for all the variability in recombination frequency observed both between and within species. Furthermore, of the mechanisms that have been identified, many are not yet fully understood. In *Arabidopsis thaliana*, considerable variation is observed in recombination frequency and distribution between natural accessions. By investigating recombination events in *A.thaliana*, this project aimed to identify *trans*-acting modifiers of recombination frequency that varied between natural accessions.

Identification of meiotic recombination modifiers was performed through Quantitative Trait Loci (QTL) mapping in *A.thaliana* natural-accession cross populations. Populations were generated from crosses between two accessions which differed significantly for recombination frequency as measured across a defined region of the genome flanked by a fluorescent-reporter system. F_1 plants were then self-fertilised to produce segregating mosaic F_2 populations for mapping. Recombination frequency for specific genomic intervals was determined for each individual in the population through measurement of the segregation of flanking fluorescence-genes expressed in the products of meiosis - seeds or pollen. Individuals were also genotyped using accession-specific markers across the genome, at a marker density of one marker per 2-5Mb, depending on the chromosome.

Association of variation in recombination frequency with specific sections of the genome differing between the parental accessions through QTL mapping revealed significant modifiers of meiotic recombination segregating within the populations. This resulted in the identification of three significant large-effect modifiers that differed between Col-0 and Cvi-0 accessions, on chromosomes 1 ,2 and 5, affecting recombination in an interval in the sub-telomere region of chromosome 3. An additional modifier on chromosome 4 affecting the same sub-telomeric interval was identified that differed between the Col-0 and Can-0 accessions.

Further fine-mapping of modifiers to improve location resolution was performed by repeated backcrosses into the Col-0 genetic background to remove the influence of other large-effect QTL and possible unknown small-effect modifiers. Improving the resolution provided a number of potential candidates for genes underlying the recombination phenotype for each QTL. Candidate testing was then performed, either through transformation of different accession alleles into the fluorescent-reporter system, or through analysis of T-DNA insertion lines that interrupted candidate genes. Preliminary results from T-DNA insertion mutants crossed to the fluorescent-reporter system suggest a potential role for the AT2G31510 gene in modification of meiotic recombination frequency, though the mode of action remains unknown.

These results demonstrate the presence of large-effect modifiers of meiotic recombination frequency that vary between the natural *A.thaliana* accessions Col-0, Cvi-0 and Can-0. Confirmation of underlying genes or sequence elements and characterisation of their mechanism of action are opportunities for exploration in future experiments.

Preface

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text

It does not exceed the prescribed word limit for the relevant Degree Committee.

Table of Contents

List of acronyms

Acronym	Term
CAPS	Cleaved Amplified Polymorphic Sequence
CFP	Cyan Fluorescent Protein
cM	CentiMorgan
CO	Crossover
DCO	Double Crossover
dHJ	Double Holliday Junction
DNA	DeoxyriboNucleic Acid
DSB	Double-Strand Break
EMS	Ethyl MethaneSulfonate
epiRIL	Epigenetic Recombinant Inbred Line
FTL	Fluorescent Transgenic Line
FDR	False Discovery Rate
GFP	Green Fluorescent Protein
IBD	Isolation By Distance
LD	Linkage Disequilibrium
LND	Low Nucleosome Density
LOD	Logarithm of Odds
MAS	Marker Assisted Selection
MMR	Mismatch Repair
mRNA	Messenger RiboNucleic Acid
NCO	Non-Crossover
NIL	Near-Isogenic Line
PCR	Polymerase Chain Reaction
PMC	Pollen Mother Cell
PSB	Pollen Sorting Buffer
QTL	Quantitative Trait Loci
RFP	Red Fluorescent Protein
RIL	Recombinant Inbred Line
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SC	Synaptonemal Complex
SDSA	Synthesis Dependent Strand Annealing
SIC	Synapsis Initiation Complex
SNP	Single Nucleotide Polymorphism
SSLP	Simple Sequence Length Polymorphism
T-DNA	Transfer DNA
TE	Transposable element
TF	Transcription factor
YFP	Yellow Fluorescent Protein

Chapter 1 – Introduction – Meiosis and meiotic recombination

1.1 Introduction: Sexual reproduction and Meiosis

The survival of a species over time depends on its propagation through generations, either through the simpler forms of cell division found in many unicellular prokaryotes, or a more complex form of reproduction such as those found in higher organisms (Campbell *et al.* 2008). Most eukaryotic species reproduce sexually using a specialised cell division known as meiosis, which halves the chromosome complement of an organism to produce haploid gametes, allowing the restoration of the original level of ploidy during fertilisation (Villeneuve and Hillers 2001; Mercier *et al.* 2015). This process promotes the accurate transmission of genetic information to the next generation in complex organisms, in addition to creating further genetic diversity in offspring, and is therefore a key step in the life cycle of many animals, plants and fungi (Mercier *et al.* 2015). Meiotic cell division appeared early during the evolution of eukaryotes and many core mechanisms are widely conserved between species, although subsequent genome divergence has created several interesting differences between groups (Villeneuve and Hillers 2001; Ramesh *et al.* 2005; Gerton and Hawley 2005; Mercier *et al*. 2015).

Traditionally, yeast was used for many early studies of meiosis, as *Saccharomyces cerevisiae* has a well characterised system with clear cytological protocols that allow visual inspection of meiosis (Roeder 1995). Furthermore, all four products of an individual meiotic division can be identified and analysed in yeast, which is not the situation in most plant or animal systems (Roeder 1995; Copenhaver *et al*. 2000). Additionally, meiosis can be easily induced in *S.cerevisiae*, resulting in a synchronised assortment of cells that facilitates analysis of specific meiotic stages, making yeast ideal for investigation of meiosis (Roeder 1995). However, a considerable number of experiments have also been performed in plant and animal species, particularly crop or livestock species that are useful for agricultural applications and mammalian models that can potentially provide insight into human systems, in addition to studies performed directly in humans (Baudat *et al.* 2013; Crismani *et al.* 2013; Mercier *et al*. 2015). Plants in particular have been a growing source of information about meiosis, as their large chromosomes facilitate cytological studies and meiotic mutants in plants are frequently more viable than those in animals, allowing the characterisation of many meiotic processes and components (Mercier and Grelon 2008; Mercier *et al.* 2015). The expansion of genetic resources and mutant collections in plants in recent years, particularly in Arabidopsis, rice and maize has aided the functional characterisation of numerous meiotic genes and greatly improved our understanding of meiosis and sexual reproduction (Mercier and Grelon 2008; Mercier *et al.* 2015).

While the understanding of the process of sexual reproduction has improved in recent decades, the reasons for the evolution of sexual reproduction, and its persistence in a majority of eukaryotic

species, are still not fully understood (Hörandl 2009; Mercier *et al*. 2015). This is because sexual reproduction has a number of inherent costs that would be expected to reduce the fitness of an organism, thereby minimising the extent to which this strategy might proliferate in a population and be transmitted to the next generation (Otto 2009). These costs include the requirement of a mating partner and the reduction in the number of genes an organism transmits to the next generation compared to asexual reproduction strategies – only 50% of an organism's genes are transmitted during sexual reproduction, resulting in what is known as 'transmission disadvantage' (Bell 1982; Otto 2009). Additionally, the process of meiotic recombination, which breaks genetic linkage and shuffles alleles, in conjunction with the characteristic chromosome segregation of meiosis can break apart favourable allele combinations that have evolved due to past selection (Otto 2009; Stapley *et al*. 2017; Dapper and Payseur 2017). Given these disadvantages, it could be expected that sexual reproduction would be a rare strategy, especially as many lineages show an ability to perform asexual reproduction (Otto 2009). Nevertheless, sexual reproduction is found in the vast majority of eukaryotes.

Theoretical models posit that sex and recombination evolved to generate and rearrange genetic variation that is required for selection, as variation gradually becomes depleted in finite populations with no recombination over time leaving a mix of linked beneficial and deleterious genetic variants in the genome (Hill and Robertson 1966; Otto and Lenormand 2002; Marais and Charlesworth 2003; Otto 2009; Stapley *et al.* 2017). The reintroduction of genetic variation and the breaking of unfavourable associations, which would affect selection and adaptation, may provide sufficient advantage over asexual reproduction under certain conditions to overcome the costs of sexual reproduction and allow its proliferation (Otto 2009; Stevison *et al.* 2017; Stapley *et al.* 2017). In support of this, there is empirical evidence that asexual populations do not adapt as rapidly as those that undergo sexual reproduction, as meiosis combines independent mutations faster than is possible in clonal species and promotes more efficient natural selection through the breaking of genetic associations (Barton and Charlesworth 1998; Colegrave 2002; Goddard *et al*. 2005; Cooper 2007; Otto 2009; Stapley *et al*. 2017). While early theoretical models had difficulty in determining how these advantages could overcome the costs of sex to allow the ubiquity of sexual reproduction among eukaryotes, later models have shown that this is likely because they did not account for features found in natural populations, instead making assumptions about migration, mutation and population size (among other factors) that are not accurate for most natural populations (Otto 2009). When models acknowledge that **i.** selection varies over time and space due to changes in conditions; **ii.** populations evolve and adapt; **iii.** migration introduces novel genetic associations and; **iv.** populations are finite which makes genetic drift an important factor, advantages of sex and recombination can be found

under a much wider range of circumstances (Muller 1964; Hill and Robertson 1966; Otto 2009; Salathe *et al*. 2009; Agrawal 2009; Dapper and Payseur 2017; Stapley *et al.* 2017).

Interestingly, despite the prevalence of sexual reproduction, and the importance of recombination in particular, there is still considerable variation in recombination observed in natural populations (Dapper and Payseur 2017; Stapley *et al.* 2017). There is some debate about the reason for this variation, but it is generally considered likely to relate to differences in genetic associations and selection in populations under differing conditions creating different requirements for diversity and adaptation and other advantages to recombination (Dapper and Payseur 2017; Stapley *et al.* 2017). Genetic modifiers of recombination have been identified in many eukaryotic species and can affect the rate and distribution of recombination events to differing degrees (Baudat *et al*. 2013; Mercier *et al.* 2015; Lawrence *et al*. 2017). Identification and characterisation of these modifiers is of interest to both geneticists and evolutionary biologists, as they can offer insight into the basic processes involved in meiosis and also provide insight into factors affecting variation in recombination, which are likely to influence evolutionary processes such as selection and adaptation. Furthermore, recombination modifiers in plants may also have uses for more practical purposes, particularly in crop improvement where skewed recombination patterns can prevent the breaking of deleterious variant combinations in particular areas of the genome and limit the generation of optimised agricultural lines in breeding programmes (Bauer *et al.* 2013; Crismani *et al.* 2013). An improved understanding of the factors involved in the control of recombination could enable manipulation of recombination rate and distribution, promoting effective use of genetic variation in trait improvement in crops.

Numerous studies have investigated meiosis and recombination in a variety of plant species, but a considerable amount of this work in recent years has been performed in the model *Brassicacea Arabidopsis thaliana* (Meinke *et al.* 1998; Mercier and Grelon 2008; Mercier *et al*. 2015). The variation in recombination rate between many natural inbred lines of *A.thaliana* has been well characterised, and evidence has been found of modifier genes in the species through the use of both mutagenic screens and Quantitative Trait Locus (QTL) mapping of natural modifiers (Sanchez-Moran *et al*. 2002; Esch *et al*. 2007; Crismani *et al*. 2012; Lopez *et al*. 2012; Ziolkowski *et al*. 2015; Seguela-Arnaud *et al.* 2015; Fernandes *et al.* 2017; Ziolkowski *et al*. 2017). However, additional variation in recombination rate remains unexplained, suggesting that further mapping and use of alternative inbred lines may identify additional modifiers, which could offer further insight into the recombination process or provide different advantages in experimental manipulation (Ziolkowski *et al*. 2015). This project aimed to test the hypothesis that there were additional modifiers of meiotic recombination frequency that had yet to be identified varying between natural inbred lines of *A.thaliana*, and investigate whether

these were considerably different between populations, which is likely as the differences in natural habitat may have resulted in differing selection pressures and differential adaptation.

This project aimed to characterise variation in recombination between *A.thaliana* natural inbred lines and use this variation to identify genetic modifiers. I utilised lines previously demonstrated to differ significantly in recombination rate to create segregating mapping populations that could be used to identify modifiers of recombination rate through QTL mapping (Ziolkowski *et al.* 2015; Ziolkowski *et al.* 2017). Identification of modifiers and characterisation of the underlying genes and their mechanism of action could explain more of the variation observed in recombination between populations. Investigation of the differences between modifiers found in different populations could aid understanding of how variation in recombination may develop, providing insight into a fundamental evolutionary process that is likely to have considerable effects on natural selection and the adaptation of populations to different conditions. In this thesis I will provide a review of the process of meiosis and meiotic recombination, in addition to some background information concerning variation in recombination in eukaryotes and the factors and modifiers that are currently known to be responsible. I will also address the evolutionary implications of natural variation in recombination and the potential benefits of improving the understanding of factors influencing recombination. The characterisation of variation in recombination between *A.thaliana* natural inbred lines and the identification of recombination modifiers and potential candidate genes will be detailed, followed by a preliminary characterisation of the effects of these modifiers and a discussion of their importance alongside potential directions for future research.

1.2 Meiotic cell division

Meiosis is a specialised cell division performed in sexually reproducing eukaryotes by diploid cells known as meiocytes, to produce haploid gametes for fertilisation (Villeneuve and Hillers. 2001). Unlike somatic cell division (mitosis) where DNA replication is followed by one round of cell division, thereby returning the chromosome number back to its original count, meiosis entails two sequential rounds of division. These consecutive reductions of the chromosome complement create a haploid product, which then fuses with another haploid gamete during fertilisation to produce an organism with a full diploid genome (Figure 1).

Meiosis is preceded by a round of DNA replication which produces identical sister chromatids that become linked together in a bivalent (Cai *et al.* 2003; Chelysheva *et al*. 2005). During prophase I, the first stage of meiosis, these chromatids condense, and chromatin becomes compacted into loop structures, linked at the base by a protein axis composed of cohesin proteins and other meiosis-specific proteins (Figure 2) (Kleckner. 2006; Mercier *et al.* 2015). The cohesins act to hold sister

Schematic representation of meiosis and mitosis removed for copyright reasons.

Figure 1. Schematic representation of chromosomes during plant mitosis and meiosis, using only two pairs of chromosomes for simplicity. Adapted from Crismani *et al*. (2013). **Top**: Mitosis through stages, resulting in two identical diploid daughter cells. **Bottom**: Meiosis I through stages, demonstrating pairing and recombination between homologous chromosomes in Prophase I, alignment on the metaphase plate and subsequent separation of homologous chromosomes. Meiosis II follows the same pattern, and is represented by the separation of sister chromatids, generating four recombinant haploid gametes.

chromatids closely together along their entire length, through association with other axis proteins (Cai *et al.* 2003). Homologous pairs of sister chromatids then synapse, becoming linked by a conserved 100nm wide proteinaceous structure known as the synaptonemal complex (SC) (Costa and Cooke 2007). This tripartite structure is formed of two lateral elements, comprised of the chromosome axes, and a linking central element holding homologous chromosomes together in a tetrad (Moses 1968; Westergaard and Wettstein 1972) (Figure 2). The SC performs two vital functions during meiosis – it holds homologous chromosomes together for correct alignment to promote accurate segregation during division, and it interacts with proteins involved in the process of meiotic recombination, providing a scaffold for the assembly of recombination intermediates and interactive processes (Page and Hawley 2004; Sanchez-Moran *et al.* 2008; Miao *et al.* 2013). This facilitates the exchange of genetic information between homologous chromosomes by promoting crossover (CO) formation, thereby encouraging the formation of stable links in the form of chiasmata between homologous chromosomes to ensure accurate segregation and allowing meiosis to generate variation between gametes.

The synapsis of homologous chromosomes is essential for the second stage of meiosis, metaphase I, where homologous chromosome pairs are aligned on the metaphase plate between the poles of the cell. Alignment and linkage of homologous chromosomes encourages correct attachment of microtubules from the centrioles at either end of the cell to kinetochores, thereby ensuring separation of homologous chromosomes to opposite poles during anaphase I (Villeneuve and Hillers 2001; Roeder 1997). This guarantees that each daughter cell will contain a copy of every chromosome. Release of cohesins responsible for sister chromatid cohesion along the chromosome arms also occurs to allow resolution of chiasmata and homolog separation, although cohesins remain within the pericentromere to ensure that sister chromatids segregate together during the first meiotic division (Cai *et al.* 2003). Unsynapsed chromosomes often fail to form chiasmata, and therefore do not align correctly as bivalents on the metaphase plate, resulting in ectopic attachments of microtubules to univalent chromosomes and unbalanced segregation of chromosomes to the poles of the cell, culminating in cells with incorrect chromosome complements (Ross *et al*. 1997). The segregation of homologous chromosomes also produces a source of genetic variation – each daughter cell will receive a copy of each chromosome, but which copy they receive will vary. Random segregation results in each daughter cell having a mix of maternal and paternal chromosomes, generating variation between the gametes (Villeneuve and Hillers. 2001; Mercier *et al*. 2015). After chromosome separation in anaphase I, the cell enters telophase I where the nuclear envelope reforms and the chromosomes decondense, concluding the first meiotic division (Griffiths *et al.* 2008; Mercier *et al*. 2015).

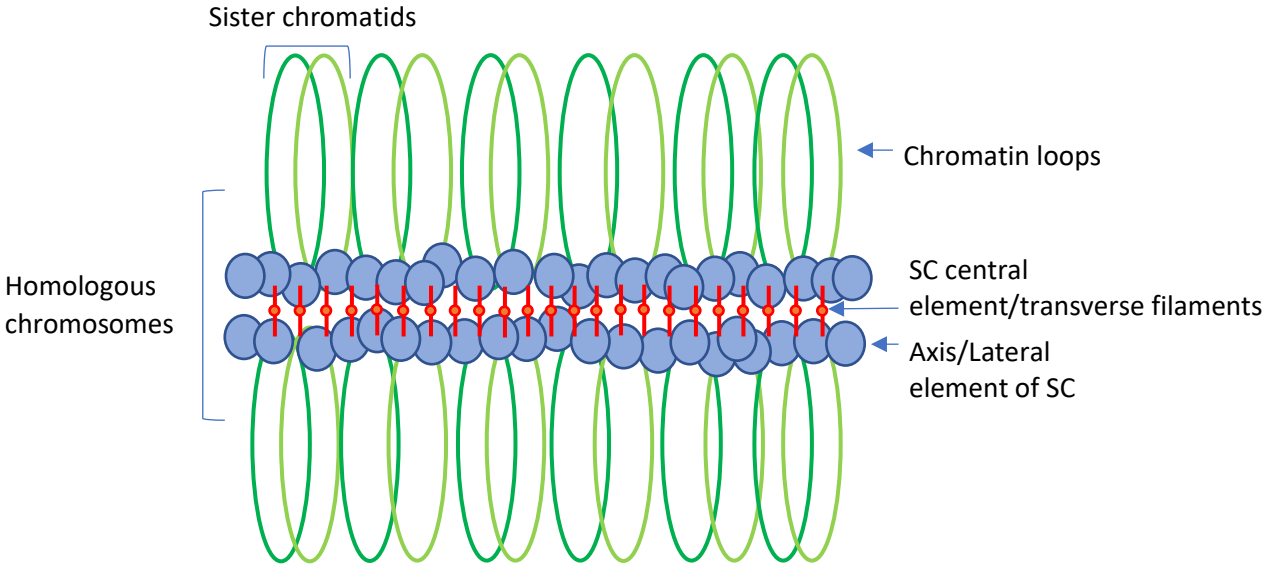


Figure 2: Schematic of synapsed homologous chromosome in prophase I of meiosis. Condensed chromatin loops (green) extend from the protein axis (blue). The transverse filaments and central element of the synaptonemal complex are shown in red, linking the axes of homologous chromosomes.

The daughter cells of the first division then proceed through a second division, meiosis II, without an intervening round of DNA replication. This second division holds more similarity to mitosis than to meiosis I. Sister chromatids condense in a bivalent, align on the metaphase plate and are separated to opposite poles of the cell after release of pericentromeric cohesins (Mercier *et al*. 2015). The nuclear envelope reforms, producing four haploid gametes. The reduction of chromosomes, and subsequent reconstitution after fertilisation, forms the basis of single-allele inheritance from each parent, and is widely conserved across the eukaryotic domain (Mercier *et al*. 2015).

1.3 Meiotic recombination

A major advantage of sexual reproduction and meiosis is the introduction of genetic variation into subsequent generations and breaking of deleterious genetic linkage through the processes of random chromosome segregation and homologous recombination, thereby providing additional mechanisms to facilitate adaptation and evolution (Barton and Charlesworth 1998; Wijnker *et al*. 2013; Stapley *et al.* 2017). Although there are differences in the processes of meiosis and recombination between species, with several factors and mechanisms still unknown or poorly understood, a considerable amount of information is known about the general progression of meiotic recombination and the proteins involved (Mercier *et al.* 2015; Lawrence *et al.* 2017). The importance of meiotic recombination as a fundamental evolutionary process is also well known, although empirical evidence is still required to test many hypotheses (Stapley *et al*. 2017).

Homologous recombination, otherwise known as crossing over, is the physical exchange of genetic information between homologous chromosomes (Villeneuve and Hillers 2001). This permits the reshuffling of genetic variants to provide new allele combinations, and the breaking up of linkage groups that may otherwise result in linkage drag of unfavourable variants through generations due to proximity to a positively selected beneficial variant (Wijnker *et al*. 2013). The combination of recombination and intermixing of chromosomes through segregation allows meiosis to rapidly combine independent mutations that arise within populations in a way that is impossible in clonal species (Barton and Charlesworth. 1998). The shuffling of variants and reduction of genetic linkage caused by recombination has been shown to aid adaptation by increasing the efficiency of natural selection (Marais and Charlesworth 2003; Morrell *et al*. 2004; Roze and Barton 2006; Webster and Hurst 2012). Variation in recombination, therefore, could have important implications for differential adaptation in natural populations (Stapley *et al.* 2017).

The overall process of homologous recombination is highly conserved between species, with high levels of similarity found in core steps and components (Villeneuve and Hillers. 2001; Mercier *et al*. 2015; Lambing *et al*. 2017). The process of homologous recombination is initiated by the formation of

a double-strand break (DSB) in the DNA, catalysed by the SPO11 (Sporulation protein 11) endonuclease in conjunction with an associated complex of proteins (Keeney at al. 1997; de Massy. 2013). While the SPO11 protein is highly conserved among eukaryotes, as is the process of DSB formation, the accessory proteins show some variation between species (Mercier at al. 2015). In the budding yeast *Saccharomyces cerevisiae* there are nine SPO11-related accessory proteins that form a variety of sub-complexes (de Massy. 2013), but these are not conserved functionally or at the sequence level across species. A number of plant-specific factors required for DSB formation have been identified, including DFO (DSB forming), PRD1 (putative recombination initiation defect 1), PRD2, PRD3 and CRC1 (central region component 1) (Zhang *et al*. 2012; Nonomura *et al*. 2004; De Muyt *et al*. 2009; De Muyt *et al*. 2007; Miao *et al*. 2013) that show some similarities with yeast proteins, however functional differences remain (Kumar *et al*. 2010; Ronceret *et al*. 2009). These differences may contribute to observed differences between yeast and plant DSB patterns across the genome, possibly by affecting DSB colocalization with other chromosome features. For example, in *S. cerevisiae* the Spp1 protein, which is involved in deposition of the H3K4^{me3} histone post-translational modification, interacts with the Mer2 yeast SPO11 accessory protein at the chromosome axis (Sommermeyer *et al.* 2013; Mercier *et al*. 2015). This interaction promotes DSB formation at H3K4^{me3} sites and could lead to a stronger correlation between DSBs and H3K4^{me3} in yeast than in many plant species, for example maize, where no equivalent interaction has been identified (Sommermeyer *et al*. 2013; Borde *et al.* 2009; Mercier *et al*. 2015; Sidhu *et al.* 2015).

After SPO11 catalyses the DSB, it remains covalently bound to the 5' end of the DNA at the break site, where it is subsequently liberated by resection of the break by the MRN complex (Mre11-Rad50-Nbs1)(Keeney and Neale. 2006; Neale *et al*. 2005). The MRN complex resects the DNA in collaboration with the Exo1 exonuclease, releasing SPO11 and the covalently bound DNA fragment, and leaving 3' single-stranded DNA overhangs (Sun *et al*. 1991). These overhangs are originally bound by replication protein A (RPA), but this is then displaced and replaced by the RAD51 and DMC1 proteins, which form a nucleoprotein filament to promote strand-invasion (Figure 3)(Brown and Bishop. 2015).

While RAD51 and DMC1 proteins perform similar functions, there are minor differences in their activities - DMC1 is meiosis specific, while RAD51 is required for both mitotic and meiotic recombination (Bishop *et al*. 1992; Neale and Keeney. 2006). However, loss of either protein causes major meiotic defects – in *A.thaliana, rad51* knockout mutants are sterile and exhibit chromosome fragmentation, while *dmc1* mutants show intact chromosomes but vastly reduced fertility (Li *et al*. 2004; Couteau *et al*. 1999). Loss of both proteins creates an even more severe phenotype in DSB repair, indicating that they are performing different functions and cannot directly compensate for one another (Da Ines *et al*. 2013).

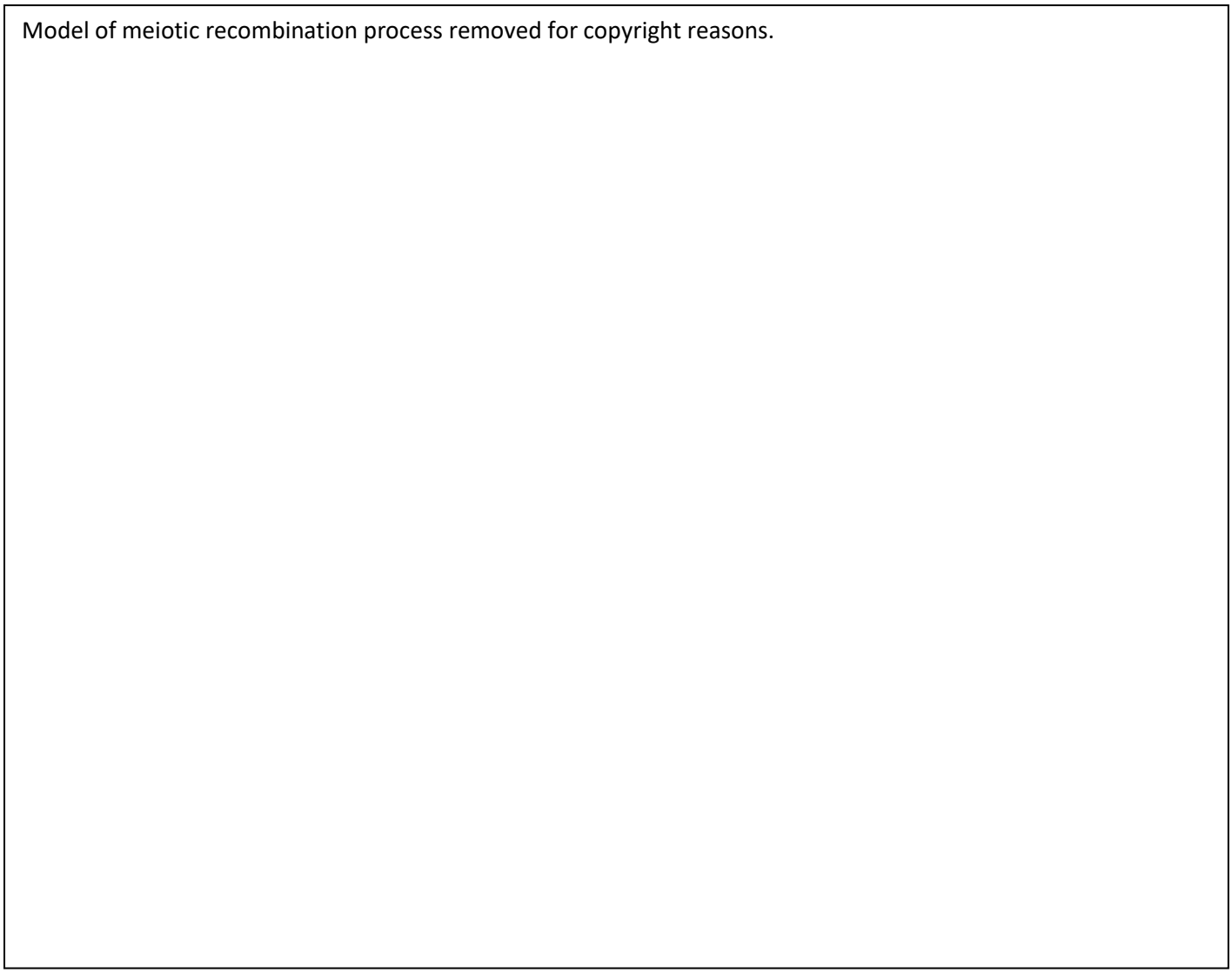


Figure 3: Model of meiotic recombination process, adapted from Mercier *et al*. (2015). Proteins involved are indicated next to the relevant step – in black to indicate that the protein promotes this step, or in red to indicate that the protein suppresses this step. Double-strand breaks (a) are followed by resection to leave 3'-overhangs (b) that are bound by proteins and invade either the sister chromatid (c) or the homologous chromosome to find a repair template (d). Displacement-loops can be developed into dHJ with the aid of ZMM proteins (e), and subsequently resolved as Class I crossovers (CO)(f), or non-crossovers (NCO) (h). Alternatively, intermediates can be resolved as NCOs through additional pathways including Synthesis Dependent Strand Annealing (SDSA)(g and i). An additional ZMM-independent pathway produces Class II COs from alternative intermediates (j). Numbers indicate the quantity of each event expected in a single meiosis in *Arabidopsis thaliana*. For references, see Mercier *et al*. (2015)

An important decision in the recombination pathway is the choice of DSB repair template, as DSBs can be repaired either from the sister chromatid or the homologous chromosome (Youds and Boulton 2011). Inter-homolog recombination is required for the shuffling of variants and breaking of genetic linkage and is often referred to simply as 'recombination', while inter-sister repair does not result in any changes to the genome as the sequences involved are identical (Youds and Boulton 2011; Mercier *et al*. 2015). As DMC1 has been previously implicated specifically in inter-homolog recombination (Schwacha and Kleckner. 1997), it was believed that the two proteins were operating in tandem, with RAD51 promoting repair from the sister chromatid, and DMC1 from the homologous chromosome. This was supported by work in *S.cerevisiae* where RAD51 activity is restricted by the Hed1 protein, thereby favouring DMC1-dependent strand invasion which can progress to inter-homolog recombination (Busygina *et al*. 2012). While no corresponding protein has been identified in plants, it is possible that a similar system regulates the nucleoprotein filament and affects the probability of inter-homolog repair. However, work from Da Ines *et al*. (2013) has shown that while RAD51 is required for DMC1 localisation during meiosis in Arabidopsis, its catalytic function in strand invasion is non-essential. This suggests that DMC1 is principally responsible for the strand-invasion step in meiotic recombination, promoting both inter-homolog and inter-sister repair, and therefore the balance between the activity of RAD51 and DMC1 is not the deciding factor in repair template choice.

During inter-homolog recombination, after the nucleofilament has formed, it locates and invades a complementary sequence in the homologous chromosome, forming a single-end strand invasion intermediate (Paques and Haber 1999). The single strand displaces the strand on the homologous chromosome, forming a displacement-loop (D-loop, see Figure 3). At this point, multiple outcomes are possible through different pathways: i. the invading strand can be displaced resulting in a non-crossover (NCO) event; ii. there may be limited DNA synthesis extending the invading strand before dissociation and re-ligation with the original strand (synthesis dependent strand annealing, SDSA, see Figure 3); iii. the second end of the original DSB can also bind the homologous chromosome, forming a double-Holliday Junction (dHJ) (Paques and Haber 1999; Schwacha and Kleckner 1995). dHJs can then be resolved as COs or NCOs, dependent on their interactions with recombination machinery, or NCOs can form from dHJ dissolution (Youds and Boulton. 2011; Bishop and Zickler. 2004; Schwacha and Kleckner. 1995).

1.4 Meiotic recombination in the context of the chromosome axis

When discussing meiotic recombination and its component proteins, it is important to consider that the formation of recombination intermediates does not occur independently from other meiotic processes and structures. DSBs and recombination occur within the context of chromatin connected

to the chromosome axis, both in relation to the accessibility of the condensed loop DNA to recombination proteins, as the chromosome axis serves to compact the DNA further and limit protein access, and recruitment of interacting proteins (Kleckner 2006; Lawrence *et al*. 2017; Sanchez-Moran *et al*. 2007; Ferdous *et al*. 2012). Association of axis proteins with SPO11 accessory proteins has been shown to tether sites in chromatin loops to the axis for DSB formation in yeast and mice (Panizza *et al.* 2011; Kumar *et al.* 2010; Lambing *et al*. 2017). While this tethering has not yet been fully investigated in plants, there is evidence to suggest that some plant axis proteins are required for wild-type DSB numbers, indicating that the axis is likely to be involved in DSB formation in plants (Ferdous *et al*. 2012; Lambing *et al*. 2017). Further protein interactions also tether the ensuing recombination complex to the axial elements (Kleckner 2006; Sanchez-Moran *et al*. 2007; Ferdous *et al*. 2012; Börner *et al*. 2004).

Numerous protein interactions are required to assemble the recombination complex on the chromosome axis and allow progression of recombination. This means that axis proteins, that form the lateral elements of the SC, are vital for recombination processes. Disruption of these structures usually results in loss of COs and failure to complete meiosis due to ectopic chromosome segregation (Ferdous *et al*. 2012; Ross *et al*. 1997). For example, the key Arabidopsis axis proteins ASY1 and ASY3 are required for both synapsis, bringing the homologous chromosomes into close apposition for recombination, and activity of the DMC1 recombination protein (Sanchez-Moran *et al*. 2007; Ferdous *et al*. 2012). Direct loss of the ASY1 protein itself in an *asy1* T-DNA insertion line results in a decline in DMC1 localisation and a loss of the bias towards inter-homolog recombination over inter-sister repair (Figure 3)(Ferdous *et al*. 2012; Sanchez-Moran *et al*. 2007). The ASY3 protein recruits ASY1 to form the axis, therefore loss of the ASY3 protein in an *asy3* knockout line disrupts ASY1 organisation in the axis, thereby reducing CO rates (Ferdous *et al*. 2012). Loss of the meiosis specific cohesin REC8 in Arabidopsis also disrupts ASY1 distribution and axis formation, resulting in defects in DSB repair, and subsequent chromosome fragmentation (Chelysheva *et al*. 2005; Lambing *et al.* 2017). This demonstrates that the presence of axis proteins and their organisation is essential for the progression of recombination and meiosis. Disruption of the central element of the SC can also result in loss of COs and segregation problems - the SC transverse filament protein ZYP1 is also required for recombination in Arabidopsis, as its loss is characterised by ectopic recombination in addition to synapsis defects (Ferdous *et al*. 2012).

Interestingly, just as DSB and recombination processes depend on components of the SC, the process of synapsis depends on recombination proteins. In many species, for example in yeast and mice, DSB formation is required for homologous chromosome pairing and synapsis (Roeder. 1997; Baudat *et al*. 2000; Romanienko and Camerini-Otero. 2000). SC formation in *Arabidopsis thaliana* is dependent on

formation of recombination intermediates, as no synapsis is observed in *spo11-1* or *dmc1* loss of recombination mutants (Mercier *et al*. 2015). This interplay between processes can complicate analysis of recombination mutant phenotypes, as it can be unclear if effects on recombination are direct, or if they are the result of effects on synapsis.

Chromosome axis proteins are also believed to play an additional role in recombination, by regulating the choice of DSB repair template – a mechanism that appears to be consistent between yeast and plants. In yeast, phosphorylation of the Hop1 axis protein by the Tel1 and/or Mec1 kinases leads to activation of the Mek1 kinase, which acts to suppress RAD51-mediated inter-sister repair, thereby establishing a bias towards inter-homolog repair (Hollingsworth and Byers 1989; Hollingsworth and Ponte 1997; Niu *et al*. 2005; Carballo *et al*. 2008; Niu *et al*. 2009). Similarly, the Arabidopsis homolog of Hop1, ASY1, is also required for establishment of the inter-homolog recombination bias, acting through inhibition of DMC1-mediated strand invasion of the sister chromatid (Kurzbauer *et al*. 2012). Work from Kurzbauer *et al*. (2012) indicated that this occurs through a phosphorylation pathway homologous to the pathway operating in yeast, by showing that the ATR kinase (homolog of yeast Mec1) regulates deposition of DMC1 at DSBs. A possible model was proposed whereby RPA-coated ssDNA at DSBs activates ATR signalling (as it is known to do in various organisms), which leads to phosphorylation of H2A.X, recruitment of DNA repair factors and prevention of DMC1 protein deposition at the DSB. Replacement of RPA with RAD51 on one side of the DSB could attenuate ATR signalling, thereby allowing deposition of DMC1 on the opposite side of the DSB, and subsequent DSB repair from the homologous chromosome (Kurzbauer *et al*. 2012). Although additional work is required to determine whether ATR affects DMC1 activity or DNA-binding directly, or through an alternative indirect process, this model demonstrates a mechanism for nucleofilament regulation and enforcement of the inter-homolog recombination bias, through a phosphorylation signalling cascade, that may be widespread among eukaryotes.

1.5 Pro- and Anti-recombination factors

While the initial formation of the recombination intermediate is dependent on a complex array of proteins and interactions, the downstream repair of an intermediate as a CO or NCO event is perhaps even more complex (Youds and Boulton 2011; Mercier *et al*. 2015). The development of each pathway is modulated by a multitude of proteins that either promote or oppose resolution of intermediates as COs, and changes in the relative activity or interactions of these proteins, or in factors influencing them, can affect the resolution outcome and thereby, the rate of recombination observed in the genome (Youds and Boulton 2011; Mercier *et al.* 2015; Lawrence *et al*. 2017).

Resolution of the dHJ as a CO is promoted by multiple factors, including a set of proteins from what is known as the ZMM pathway. In plants this includes the MSH4/5 heterodimer (Snowden *et al*. 2004; Higgins *et al*. 2008), the MER3 helicase (Mercier *et al*. 2005), the MLH1/MLH3 heterodimer (Dion *et al*. 2007; Jackson *et al*. 2006), the ZIP4 synapsis initiation complex (SIC) protein (Chelysheva *et al*. 2007), the SHOC1 XPF nuclease and associated protein PTD (Macaisne *et al.* 2008; Macaisne *et al.* 2011) and the HEI10 E3 ubiquitin ligase (Chelysheva *et al*. 2012). Loss of any of these proteins in *A.thaliana* results in a significant reduction in COs to 15-40% of wild-type levels (Higgins *et al*. 2004; Jackson *et al*. 2006), though analysis of combined mutations indicates they all act at different points within the same pathway and provide different functions (Mercier *et al*. 2015; Chelysheva *et al*. 2007; Chen *et al*. 2005; Higgins *et al.* 2008; Macaisne *et al*. 2011, Mercier *et al*. 2005). The MSH4/MSH5 MutS-homologue heterodimer is proposed to stabilise dHJ to promote COs (Snowden *et al*. 2004; Higgins *et al*. 2008), while the MER3 helicase is believed to play a role in strand exchange by extending the DNA heteroduplex (Mercier *et al*. 2005). The SHOC1 XPF endonuclease and interacting protein PTD, an ERCC1-like protein, are suggested to be involved in the maturation of recombination intermediates, though their exact role is unclear (Macaisne *et al*. 2008; Macaisne *et al*. 2011), and the ZIP4 protein is involved in the initiation of synapsis, though it is not required for its completion (Chelysheva *et al*. 2007). The HEI10 protein is functionally related to yeast Zip3 which is involved in synapsis initiation and required for ZMM-dependent CO formation (Chelysheva *et al*. 2012). The MLH1/MLH3 MutL-homologue heteroduplex is involved in the later stages of dHJ resolution, potentially imposing a conformation that promotes resolution as a CO. This is supported by loss of AtMLH3 causing a bias in resolution towards NCO (Dion *et al*. 2007; Jackson *et al*. 2006).

The ZMM pathway is responsible for a majority of COs in plants and other species, including approximately 80-85% of COs in *A.thaliana* (Lam *et al*. 2005; Berchowitz *et al*. 2007; Hollingsworth and Brill. 2004), but the presence of a small number of COs in *zmm* mutant backgrounds demonstrates that there is at least one ZMM-independent pathway of CO production. At least a proportion of these are dependent on the MUS81 endonuclease, which is conserved between plants, animals and fungi, as *zmm mus81* mutants exhibit even lower recombination levels than *zmm* single mutants (Higgins *et al*. 2008). These two pathways are classified as Class I (ZMM-dependent) and Class II (MUS81-dependent) COs, and between them account for most of the observed recombination events. However, analysis of an Arabidopsis *msh4 mus81* double mutant showed some residual COs (Higgins *et al*. 2008), suggesting there may be further unknown minor pathways. Additionally, while these pathways are conserved in many eukaryotic species, there are exceptions – in *Drosophila melanogaster mus81* mutants show no reduction in recombination, suggesting that all COs come from the Class I pathway. Similarly, in *Caenorhabditis elegans msh4* mutants have no COs, supporting the

absence of the Class II pathway (Hollingworth and Brill. 2004), and in *Schizosaccharomyces pombe mus81* mutants show complete loss of recombination, which is indicative of a lack of Class I COs (Berchowitz *et al*. 2007). Therefore, while these processes are generally consistent between eukaryotes, some variation does exist.

In most plant species, only a small proportion of DSB events form COs (Mercier *et al*. 2015; De Muyt *et al.* 2009) - in *A.thaliana* approximately 200 DSBs results in just 10 COs (Copenhaver *et al.* 2002; Chelysheva *et al.* 2010; Ferdous *et al.* 2012, Wjinker *et al.* 2013; Giraut *et al.* 2011; Salome *et al.* 2012). The majority of DSBs are repaired as NCOs, with the decision to repair via CO or NCO made early in the process, around the time of strand invasion (Allers and Lichten. 2001; Boerner *et al*. 2004). This decision is likely partly related to crossover interference, a phenomenon where the presence of a CO reduces the probability of a second CO occurring nearby (Berchowitz and Copenhaver 2010). This inhibition weakens with distance from the original recombination event, resulting in COs being more widely and evenly spaced than expected by chance (Copenhaver *et al*. 2002; Berchowitz and Copenhaver 2010). CO interference has been observed in multiple species, though it can differ in strength. For example, in *C.elegans* the interference signal appears to cover entire chromosomes as each chromosome has only one CO, whereas in *A.thaliana* the chromosomes with the longest physical length frequently have at least two recombination events per meiosis (Hammarlund *et al*. 2005; Giraut *et al*. 2011) though they are widely spaced (Copenhaver *et al*. 2002; Lam *et al.* 2005; Drouaud *et al.* 2007). This may contribute to the observation that total CO frequency is generally low, regardless of physical genome size (Figure 4)(Mercier *et al*. 2015; Smukowski and Noor. 2011). However, not all COs exhibit interference – the phenomenon is limited to Class I COs, with class II COs showing a more random distribution. This is clearly demonstrated in Arabidopsis *zmm* mutants, where residual COs are randomly distributed (Higgins *et al*. 2004), and the fit of models of CO distribution are improved by including two classes of CO – one being interference-insensitive (Copenhaver *et al*. 2002).

The chromosome axis, which plays a role in the regulation of recombination through protein recruitment, is also believed to be of importance for CO interference, as it is thought to regulate CO number by transmitting the CO interference signal (Kleckner 2006; Berchowitz and Copenhaver 2010). Recent models have posited that the axis implements interference through a mechanical stress model, where expansion and contraction of chromatin loops tethered to the axis during meiosis generates localised stress that can be relieved bi-directionally by DSB designation as COs (Berchowitz and Copenhaver 2010; Kleckner *et al*. 2004; Kleckner 2006; Zhang *et al*. 2014). Reduction of tension proximal to the CO reduces the likelihood of a second CO occurring nearby. While there is

Figure 4: Number of crossovers per chromosome per meiosis in a selection of eukaryotic species, adapted from Mercier *et al*. (2015). The number of COs, inferred from male/female-average genetic maps, is plotted against the physical size of each autosomal chromosome (Mb, log scale). See Mercier *et al.* (2015) for data and references.

circumstantial evidence to support this, such as the timing of DSB formation and CO resolution coinciding with expansion and contraction cycles of chromatin, the model has not been definitively proven (Kleckner *et al*. 2004). However, analysis of *topoisomerase II* mutants in budding yeast suggests that Topoisomerase II is necessary for CO interference, and that it may catalyse the effect by adjusting spatial relationships between sections of DNA during chromatin compaction, which could provide a mechanism for alleviating mechanical stress and lend further support to the theory (Zhang *et al*. 2014). Interestingly, it appears that SUMOylation, a post-translational modification, of both topoisomerase II and the axis protein Red1 is also required for wild-type interference levels in yeast (Zhang *et al.* 2014). As interference is also influenced by the ubiquitin-mediated removal of SUMOylated proteins, this suggests that a SUMO/ubiquitin relay involving the chromosome axis could be part of a molecular mechanism involved in the implementation of interference (Zhang *et al.* 2014).

While CO interference may provide a proximate explanation for the excess of DSBs over COs, it does not explain why so many DSBs are formed in plants to begin with. The relationship between DSBs and chromosome pairing and synapsis has provided a potential explanation for the excess in DSBs over COs observed in plant meiosis (De Muyt *et al*. 2009). It is possible that this excess in DSBs could provide additional interactions to ensure efficient homolog pairing in large plant chromosomes. This is consistent with smaller budding yeast chromosomes, which don't require so many interactions, not showing the same excess in breaks (Zickler and Kleckner. 1999; de Massy. 2013). The additional breaks can then be repaired via a different mechanism, or recombination intermediates can be resolved by anti-CO mechanisms to limit the number of genetic exchanges between chromosomes. A number of anti-CO mechanisms have been identified in plants, many through a series of experiments in *A.thaliana* performing a mutant screen in a *zmm* background to identify genes whose loss restores fertility (Crismani *et al*. 2012). Knockout of these anti-CO mechanisms can produce a boost in CO levels that compensates for the loss of Class I COs and allows accurate segregation of chromosomes and completion of meiosis.

A variety of distinct anti-CO mechanisms have been identified, suggesting that the outcome of recombination is tightly regulated by overlapping pathways. The FANCM helicase and its cofactors were discovered to suppress COs in *A.thaliana*, as disruption of the FANCM gene in a *zmm* mutant background leads to a three-fold increase in CO number over wild-type levels. These events are MUS81-dependent, indicating that FANCM disrupts Class II CO intermediates (Crismani *et al*. 2012). FANCM is believed to unwind D-loops by displacing the invading strand, thereby promoting NCOs through SDSA, which is supported by evidence that yeast orthologs of FANCM unwind somatic D-loops (Sun *et al*. 2008; Prakash *et al*. 2009). However, even in the *fancm* loss of function mutant DSBs considerably outnumber COs, signifying that additional anti-CO pathways are still active (Crismani *et*

al. 2012; Seguela-Arnaud *et al*. 2015). A second anti-CO mechanism system was identified in Arabidopsis by the same *zmm* mutant suppressor screen, involving the BLM helicase homologues RECQ4A and RECQ4B (Seguela-Arnaud *et al*. 2015). The two paralogs, formed from a duplication event in Brassicaceae, have some functional redundancy but loss of both proteins results in a six-fold increase in COs, showing that they act as CO suppressors in the wild-type (Seguela-Arnaud *et al*. 2015). DSB numbers are unaffected in the *recq4ab* double mutant compared to wild-type, and the increase in COs is independent of FANCM but dependent on MUS81, suggesting a second mechanism of unwinding Class II CO intermediates (Seguela-Arnaud *et al*. 2015). A model was proposed whereby RECQ4A/B act in a complex with Top3α, RMI1 and RMI2 (a BTR-complex ortholog)(Seguela-Arnaud *et al.* 2015; Tang *et al.* 2015; Kaur *et al.* 2015; Fasching *et al.* 2015) to promote D-loop displacement, as orthologs in yeast and humans are known to promote intermediate unwinding (De Muyt *et al*. 2012; Zakharyevich *et al*. 2012; Daley *et al*. 2014). While AtRECQ4A/B suppress excess COs, they are not required for resolution of recombination intermediates. Alternatively, loss of AtTop3α is lethal, and as Top3α is also known to promote resolution of dHJs as NCOs (Seguela-Arnaud *et al*. 2016; Wu and Hickson 2003; Bocquet *et al.* 2014), this implies an additional function, separate from the RECQ4A/B complex, with the RMI1 co-factor that is essential to resolve intermediates and prevent subsequent chromosome entanglement and breakage (Seguela-Arnaud *et al*. 2016). Therefore, Top3α offers two mechanisms of anti-CO activity – unwinding of D-loops in a complex with RECQ4 helicases, and dissolution of dHJs with RMI1.

A third anti-CO pathway in plants involves the AAA-ATPase FIDGETIN-LIKE-1 (FIGL1) and its partner FLIP which form a complex that's conserved in eukaryotes (Fernandes *et al*. 2017). Combination of *figl1* and *fancm* loss of function mutations results in a larger increase in recombination than that observed in either single mutant, demonstrating that these proteins act independently to suppress CO in Arabidopsis (Girard *et al.* 2015). FIGL1 has been shown to control the dynamics of the DMC1 and RAD51 strand invasion proteins, suggesting that it limits invasion of the homologous chromosome, acting before FANCM in the recombination process (Girard *et al*. 2015). The final anti-CO pathway that has been identified prevents ectopic recombination between divergent sequences and involves the mismatch repair (MMR) protein MSH2 (Chakraborty and Alani 2016). In Arabidopsis MSH2 is known to suppress recombination between polymorphic sequences, such as those found between different accessions (Emmanuel *et al*. 2006).

The interaction of pro- and anti-CO factors creates a balance between CO and NCO outcomes of DSB repair that is believed to be necessary to maintain genome stability, while still providing sufficient CO to generate genetic variation and link homologous chromosomes by chiasmata for accurate segregation. While a minimum requirement for CO number appears to have been established in many

eukaryotic species, as low levels of CO result in achiasmate chromosomes and mis-segregation during meiosis (as seen in ZMM pathway mutants: Mercier *et al.* 2015; Higgins *et al.* 2008; Mercier *et al*. 2005; Chelysheva *et al*. 2007), it has often been unclear whether a maximum limit exists. It was believed that an overabundance of chiasmata may cause entanglement of the chromosomes and ultimately breakages when they attempted to separate, particularly in plants due to the large physical size of their chromosomes – a hypothesis that seemed to be supported by observations that natural CO rates in plants are usually constrained far below the hypothetical maximum posited by the number of DSBs (Mercier *et al*. 2015; De Muyt *et al.* 2009; Louis and Borts 2003). However, increasing the recombination frequency to up to nine-fold wild-type levels in *A.thaliana* does not cause any observable meiotic defect (Seguela-Arnaud *et al*. 2015), consistent with yeast where high recombination rates do not result in meiotic defects (Mercier *et al*. 2015) (Figure 4), suggesting that suppression of COs below maximum levels is due to the long term cost of high recombination, potentially in the breaking up of beneficial linkage groups (Seguela-Arnaud *et al*. 2015; Hadany and Comeron 2008; Otto 2009). Further experimentation is required to fully clarify the recombination process, the influences on its outcome, and the evolutionary reasoning behind the balance.

1.6 Variation in meiotic recombination frequency between species

Meiotic recombination is a crucial component for the correct segregation of chromosomes during meiotic division, in addition to providing a mechanism to generate further variation in sexually reproductive populations (Bauer *et al* 2013; Barton and Charlesworth 1998). However, despite its necessity in creating physical linkages between homologous chromosomes, meiotic recombination rate has been shown to be incredibly variable between species, and even within species, populations and across chromosomes. This variation is observable both in the frequency of recombination events in a single meiotic division, and in the distribution of these events across the genome (Lawrence *et al*. 2017).

Meiotic recombination frequency is not equal between all species. Even accounting for differences in physical genome size, large disparities in average genetic distances across the genome are observed (Mercier *et al*. 2015). For example, while the average genome wide recombination rate in wheat has been measured at 0.2cM/Mb in the 17,000Mb genome, the mustard weed *Arabidopsis thaliana* has an average recombination rate of 5cM/Mb over its 125Mb genome – 25-fold higher (Choulet *et al*. 2014; Salome *et al*. 2012) - and maize lines show an average recombination rate of 0.73cM/Mb (Bauer *et al*. 2013) (Figure 5).

This phenomenon is not unique to plant species – recombination is variable between most species, regardless of close evolutionary relationships or genome sequence similarities (Mercier *et al.* 2015).

Graphs showing recombination distribution (cM/Mb) across a chromosome (Mb) in wheat and *Arabidopsis thaliana* removed for copyright reasons.

Figure 5. Recombination distribution across the chromosome in wheat and *Arabidopsis thaliana*, adapted from Choulet et al. (2014) and Yelina et al. (2013). A. Wheat chromosome 3B - meiotic recombination rate (cM/Mb sliding window of 10Mb in black and 1Mb in red). **B.** Crossover frequency along chromosome 1 of Arabidopsis. Mean crossover frequency denoted by horizontal red dashed line, centromere denoted by vertical black dashed line and locations of FTL T-DNAs denoted by vertical black lines.

Looking at other model organisms, *Drosophila melanogaster* has a sex-averaged genome wide recombination rate of 3cM/Mb, humans have a sex-averaged rate of 1.3cM/Mb, and budding yeast is known to have a recombination rate per unit of physical distance that is much higher than in humans (Nachman 2002; Kauppi, Jeffreys and Keeney 2004).

Studies performed in less commonly examined species show similar variation - despite a highly conserved genome structure and gene synteny, chickens and reed warblers also differ significantly for average recombination rates across the genome (Dawson *et al*. 2007), as do closely related grasshopper species (Hewitt 1964). A study looking at eight closely related rodent species from the Muridae family also demonstrated that recombination rate is highly variable between mammalian species, and that these differences are not proportional to DNA sequence divergence (Dumont *et al*. 2011). Recombination patterns aren't even conserved between chimps and humans which share 99% sequence identity (Winckler *et al*. 2005; Ptak *et al*. 2005), suggesting that these differences are not due to overall divergence of genomes between species but rather due to changes in specific features. Figure 4, taken from a review by Mercier *et al*. (2015) shows that the number of COs per chromosome per meiosis is variable between many species from different families, and that this is weakly positively correlated with physical genome size, though this does not explain much of the observed variation, again suggesting the presence of specific features influencing recombination rates.

However, despite all these differences, some similarities do remain, particularly in plant species domesticated as crops – recombination frequencies in barley correlate with wheat estimates, and recombination over gene clusters in wheat is comparable with estimates made over similar regions in rice and tomato (Kunzel *et al*. 2000; Gill *et al*. 1996) – which is encouraging for work aiming to alter recombination in crop species, as it suggests some conservation of mechanisms and features may be present.

1.7 Variation in meiotic recombination distribution

In addition to variability in recombination frequency, the distribution of recombination events across the genome is also often uneven (Figure 5) (Drouaud *et al*. 2013, Koehler *et al*. 2002) – some areas have consistently higher rates of recombination, and some areas are frequently crossover suppressed (Nachman 2002). The distribution of events, or recombination pattern, has also been observed to vary between species, and between individuals of the same species, although some general rules do account for broad scale chromosomal patterns in most organisms. For example, recombination is usually suppressed in heterochromatic areas of the chromosome, particularly around the centromeres, and elevated in euchromatic sub-telomeric regions, although this is not the case for all chromosomes or individuals (Nachman 2002). These common features often relate to conserved

meiotic mechanisms and may reflect adaptation of recombination patterns to maintain fertility – suppression of recombination around the centromere is vital for proper chromosome segregation in many organisms, which explains its ubiquity across a range of species (Ellermeier *et al.* 2010; Stapley *et al*. 2017).

However, considerable variations in this pattern do exist in nature. Over broad scales humans have been shown to have high recombination closer to the telomeres of chromosomes, and lower recombination around the centromeres (Myers *et al*. 2005). This contrasts with the distribution of events observed in budding yeast, where double-strand breaks (DSBs), the precursors of crossovers, are low around the centromere and sub-telomere and most events are detected in the interstitial chromosome arms (Kauppi, Jeffreys and Keeney 2004). While these observations were made of DSBs and not COs, in yeast COs account for a far higher proportion of DSB repair products than in many plant or animal species, with estimates ranging from 40-68% of DSBs repaired as COs in *S.cerevisiae* compared to approximately 5% observed in Arabidopsis, and 10% in mammals (Chen *et al*. 2008; Mancera *et al.* 2008; Mercier *et al*. 2015; Lambing *et al.* 2015; De Muyt *et al.* 2009; Giraut *et al.* 2011; Sarbakna *et al.* 2012; Guillon *et al.* 2005). Consequently, the DSB distribution is more representative of the subsequent distribution of COs across the genome in *S.cerevisiae* than it is in other species, which means that DSB patterns can be used as a rough approximation of the broad scale pattern of recombination in yeast, suggesting that COs are also likely to be found predominantly in interstitial regions. Similar recombination patterns have been observed in *D.melanogaster*, where recombination events are also frequently detected in the chromosome arms, although the distribution across an individual chromosome arm is highly variable, with regions varying in recombination frequency from 0 to over 5 cM/Mb (Nachman 2002).

In plants, the broad scale pattern remains of centromeric suppression and recombination in euchromatic regions of the chromosome, but again, variations can be detected between species and populations. A study of *A.thaliana* (hereafter Arabidopsis) chromosome 4 revealed a pattern of minimal recombination in the centromeric region, followed by alternating regions of higher and lower recombination frequency along the chromosome arms and into the sub-telomere, while analysis of segregating F_2 populations generated from Arabidopsis inbred population crosses reveals some variation in crossover distribution, but also conservation of the pericentromeric region as an area of high recombination (Drouaud *et al*. 2006; Salome *et al*. 2012). Similarly, tomato, wheat, rice and barley all show centromeric suppression of recombination, although the distance over which this extends is smaller in rice (Cheng *et al.* 2001). While tomato recombination is broadly confined to euchromatin, recombination in wheat follows a more distinct pattern of being primarily located in the distal portions of the chromosome arms - 90% of crossovers occur in the 40% of the chromosome around the sub-

telomere (Sherman and Stack 1995; Gill *et al*. 1996; Saintenac *et al*. 2009). This distal skew is shared by maize and barley (Tenaillon *et al*. 2002). Recombination in barley is heterogeneous along the chromosome arms with small highly recombinogenic areas surrounded by large recombinationally suppressed regions, creating an alternating pattern similar to that seen in Arabidopsis (Kunzel *et al*. 2000). Therefore, while some broad-scale features, such as suppression of recombination in the centromere, are conserved in plants, additional features of the pattern are variable between species, and cereal crops favour more distal recombination than other plant species.

In addition to broad-scale recombination patterns, fine-scale variations in recombination distribution are also observed across the genome. Small areas of the genome with recombination rates magnitudes higher than the genome average, known as recombination hotspots, have been identified in a multitude of species, and in many cases account for a large proportion of crossovers (Myers *et al*. 2005). In human populations extensive rate variation is seen across the genome, and hotspots are common, occurring every 200kb or less, preferentially outside of genes (Crawford *et al*. 2004; McVean *et al*. 2004). Approximately 50% of human recombination is found to occur in less than 10% of the genome sequence, and recombination ranges from below average to up to 370-fold higher than the genome average in the hottest hotspots, creating a dynamic recombination pattern (Kauppi, Jeffreys and Keeney 2004). Myers *et al*. (2005) identified over 25000 hotspots in the human genome, and found that they did not occur randomly, but were constrained to appear in sites with specific features such as recurring sequence motifs and sequence contexts. While these motifs are not sufficient to fully explain hotspot distribution, they do provide a framework of sites upon which additional factors can act.

Recombination hotspots are also found in plants – multiple hotspots have been identified in *Zea mays*, in both genic and non-genic contexts (Yao *et al*. 2002; Dooner 1986), and hotspots have also been identified and characterised in Arabidopsis, clustering in regions less than 8kb in length and preferentially occurring in nucleosome free regions at the 5’ and 3’ ends of genes (Yelina *et al*. 2012; Drouaud *et al*. 2013). This indicates a link with accessibility of DNA to the recombination machinery that corresponds with observations made in *S.cerevisiae* where DSB hotspots, which frequently parallel CO hotspots in yeast, are often found in the low nucleosome density (LND) regions around gene promoters (Pan *et al*. 2011).

While recombination occurs predominantly in hotspots in many species, including humans, mice, and chimps, other species such as *D.melanogaster* do not contain hotspots (Nachman 2002). Additionally, even between closely related species, fine-scale recombination and hotspot usage is variable. For example, despite high sequence similarity and close evolutionary links, hotspots are not conserved

between humans and chimps, and their recombination landscapes are significantly different at a fine scale (Ptak *et al*. 2005; Winckler *et al*. 2005). No species pairs are yet known with either complete divergence or complete conservation of hotspots between them (Smukowski and Noor 2011), suggesting that the evolution of the recombination distribution pattern has been a gradual process, and is likely influenced by many factors.

The considerable variation in crossover distribution found in nature could be the result of several processes, and it is likely to have substantial implications for genome evolution in different species, as the rate of adaptation and generation of genetic diversity throughout the genome is believed to vary depending on the level of recombination (Lawrence *et al.* 2017; Stapley *et al.* 2017; Dapper and Payseur 2017). Variation in patterns of molecular features such as sequence motifs and chromatin state, that are known to affect recombination and vary between populations, is likely to be the proximate cause of variation in crossover distribution (Lawrence *et al.* 2017). Differences in recombination patterns may develop between species as an indirect result of selection acting to alter factors that have a downstream effect on recombination but also have additional functions within the cell (Otto and Barton 2001; Smukowski and Noor 2011; Morgan *et al.* 2017; Stapley *et al*. 2017; Dapper and Payseur 2017).

It is also possible that changes in crossover distribution may be directly selected for in certain species as specific recombination patterns may offer a fitness advantage (Otto and Barton 2001; Smukowski and Noor 2011; Webster and Hurst 2012; Stapley *et al*. 2017; Dapper and Payseur 2017). For example, skews towards distal crossovers could be selected for in plant species with large chromosomes and polyploid genomes as they may facilitate accurate chromosome segregation by limiting chromosome entanglements (Saintenac *et al*. 2009; Tenaillon *et al*. 2002; Mercier *et al.* 2015; Bomblies *et al*. 2016). Alternatively, there may be selection for redistribution of crossovers towards gene-rich areas of the genome to promote shuffling of alleles and aid adaptation in organisms that need to adapt to a change in conditions (Otto 2009; Stapley *et al*. 2017; Dapper and Payseur 2017). In parallel, recombination may be directed away from gene-rich regions in other populations that are well adapted to their conditions, as the breaking of linkage between beneficial alleles would be detrimental to the organism (Lenormand and Otto 2000; Stapley *et al*. 2017; Dapper and Payseur 2017). This would therefore result in differences in crossover distribution between populations. While the causes of much of the variation in crossover distribution between populations are unknown, the importance of recombination distribution for meiosis and genome evolution means that characterisation of differences in crossover distribution between populations could provide insight into fundamental biological processes.

1.8 Sex-specific features of the recombination landscape

Interestingly, another source of variation in recombination in eukaryotes is the observation of differences in recombination between the sexes. Sex-specific variation in recombination frequency and distribution has been observed in many species, notably in humans where the average recombination rate in male meiosis is 0.9cM/Mb, whereas in female meiosis it is 1.7cM/Mb (Nachman 2002). The human female to male ratio of genetic distance varies across the chromosome, peaking around the centromeres as male recombination is higher at distal ends of chromosome arms (Broman *et al*. 1998). Variation in recombination between the sexes can also be observed at a finer scale in mammals. Sex-specific differences are also observed between hotspots in mice – while hotspot locations are shared between males and females, their overall recombination distributions still differ considerably, leading to the hypothesis that activity levels at individual hotspots may differ between sexes, creating a sexually dimorphic recombination pattern (Baudat *et al*. 2007). In other non-mammalian species this dimorphism is even more severe – *D.melanogaster* males do not undergo any meiotic recombination (Nachman 2002).

Sex-specific recombination features are also found in plants, where they often manifest as considerable differences in recombination distribution. In tomato, male and female gametes from the same plant have substantially different recombination patterns, and dramatic differences are observed in Arabidopsis, both in genetic map length and distribution – similarly to humans, Arabidopsis male recombination increases towards distal ends of the chromosome where female recombination is low (Figure 6) (de Vicente and Tanksley 1991; Giraut *et al*. 2011; Drouaud *et al*. 2007).

Although a clear explanation for the substantial differences in recombination between the sexes has not yet been found, it is believed that they may relate to general differences between the processes of male and female meiosis (Hunt and Hassold 2002; Drouaud *et al.* 2007; Kong *et al.* 2008; Mercier *et al.* 2015). Differences have been observed in the control of the meiotic cell cycle, and in the length of the SC, between male and female meiosis in both plants and mammals, and these differences could have an effect on recombination patterns (Hunt and Hassold 2002; Drouaud *et al.* 2007; Mercier *et al.* 2015). However, the mechanisms underlying sex-specific recombination phenotypes are currently unknown, making it difficult to determine if differences are the consequence of direct selection on recombination, or a by-product of sex-specific changes to other features (Drouaud *et al.* 2007). Interestingly, while the mechanisms underlying the disparity in recombination between the sexes are still unclear, a genetic determinant underlying the difference in humans has been identified. Specific variants in the *RNF212* gene, an ortholog of the *ZHP-3* gene that is required for crossover formation in *C.elegans*, are associated with higher recombination in males and lower recombination in females

Graph showing crossover distribution (cM/Mb) and GC% along chromosome 1 (Mb) of *Arabidopsis thaliana* removed for copyright reasons.

Figure 6. Crossover distribution and GC% along chromosome 1 of *Arabidopsis thaliana*, adapted from Giraut *et al.* (2011). Blue line denotes crossover rates in female meiosis, red line denotes crossover rates in male meiosis. Dotted black line denotes GC%. Black stars indicate regions with significantly different sex-specific crossover rates. Black box in bar under graph denotes heterochromatic regions.

(Kong *et al*. 2008). This results in a relatively consistent sex-averaged rate between *RNF212* alleles, but male and female patterns are significantly different (Kong *et al*. 2008). However, it is not yet understood how this variation in *RNF212*, a gene known to affect recombination rates in mammals, is involved in the establishment of sex-specific recombination patterns, therefore additional work is still required to understand this phenomenon.

Although the mechanism responsible for the disparity is currently unknown, the variation in male to female recombination ratios between even closely related species is still an important factor to consider in comparison of recombination patterns, as it is likely to contribute significantly to observations of recombination differences between organisms (Drouaud *et al*. 2007). This means that characterisation of differences in recombination between sexes in different populations could provide additional depth to the understanding of the causes of crossover variability between natural populations.

1.9 Intraspecific variation in recombination

Interspecies variation in recombination is well known and characterised, but additional variation is also found within species. Fine-scale recombination patterns in humans show heritable variability in hotspot activity between individuals (Coop *et al*. 2008; Baudat *et al*. 2007), in addition to broad-scale genome wide recombination rate variation (Kong *et al*. 2008). Recombination distribution has also been shown to vary between humans in gametes from different mothers (Broman *et al*. 1998). Similarly, genomic recombination rates of mouse sub-species vary significantly, and mouse strains of the same species show similar regional crossover distributions but weaker correlation of fine-scale recombination and hotspot usage (Dumont *et al*. 2011; Paigen *et al*. 2008; Baudat *et al*. 2010; Baudat *et al*. 2007).

A substantial number of studies have addressed the issue of intraspecific recombination within plant species, looking at inbred crop species and natural ecotypes. For example, the average recombination level between markers was found to be highly variable between both maize inbred lines, looking at 22 lines from the Dent and Flint gene pools, and maize x teosinte hybrids (Bauer *et al*. 2013; Williams, Goodman and Stuber 1995). Variation in genomic recombination rate has also been observed between near isogenic lines of rye, between barley genotypes, between *Brassica oleracea* populations and between wheat populations (Rees 1961; Sall 1990; Nilsson and Pelger 1991; Sebastian *et al*. 2000; Zarchi *et al*. 1972). Consistently, different species demonstrate variability in recombination between populations and individuals, indicating ongoing divergence of recombination phenotypes and segregating factors influencing crossover frequency and distribution.

1.10 Intraspecific variation in *Arabidopsis thaliana*

While many plant species demonstrate intraspecific variation, few have a system of genetically distinct populations as well studied and characterised as *Arabidopsis thaliana*. Arabidopsis, a mustard weed from the Brassicaceae, is native to Africa though more recent expansion and colonisation after the last glacial period has given it a range that extends across Europe, central Asia and North America (Figure 7) (1001 Genomes Consortium 2016; Durvasula *et al*. 2017). This large range covers a variety of habitats and conditions, predominantly disturbed habitats like rocky slopes, making Arabidopsis suitable for analysing variation in adaptive traits. Its predominantly self-fertilising lifestyle has produced a vast range of natural inbred lines, known as accessions, from different habitats, that exhibit extreme pairwise-sequence divergence (Figure 7) (Koorneef *et al*. 2004; Alonso-Blanco *et al*. 2016). Genetic and phenotypic characterisation of these accessions has produced a plant system that is optimal for the study of genetic variation in traits (Alonso-Blanco and Koornneef 2000). Furthermore, unlike many other model systems where immortalized genotypes are often laboratory samples, Arabidopsis accessions represent adapted natural genotypes, thereby providing additional information about adaptation and selection (1001 Genomes Consortium 2016). When added to other features such as its fast life cycle and robust genome pliable to genetic manipulation, this makes Arabidopsis a popular system for mapping genes involved in a variety of processes in plants.

Arabidopsis accessions show substantial variation in DNA sequence and phenotypic characteristics, and sequencing data available for many accessions from the 1001 genomes project has facilitated studies of the genetic variation underlying traits of interest (Alonso-Blanco *et al*. 2016; Alonso-Blanco and Koorneef 2000). Numerous studies have been performed using Quantitative Trait Locus (QTL) mapping in accession cross populations to identify genes underlying a variety of traits. Recombinant-inbred populations generated from accession crosses were used to identify lineage-specific alleles underlying natural variation in trichome density, enzyme activity in primary and secondary metabolism, and seed oligosaccharide content in Arabidopsis (Vaughan Symonds *et al*. 2005; Mitchell-Olds and Pederson 1998; Bentsink *et al*. 2000). Multiple significant QTL have also been identified for seed dormancy behaviour and floral morphology traits, where several putative candidate genes have been proposed based on gene position and function (Juenger *et al*. 2000; Alonso-Blanco *et al*. 2003).

Variation in meiotic recombination has also been observed between Arabidopsis accessions, through the use of several different techniques to measure crossovers. Cytological analysis of chiasma frequency in pollen-mother cells (PMCs) of geographically diverse accessions has demonstrated significant variation in recombination - Lopez *et al*. (2012) identified a variation in chiasma frequency of 22% between 17 different accessions, and the distribution of chiasmata along the chromosome

Graph showing distribution of pairwise genetic distances among *Arabidopsis thaliana* accessions and map of geographic locations of accessions removed for copyright reasons.

Figure 7. Genetic variation and geographic distances between *Arabidopsis thaliana* accessions, adapted from Alonso-Blanco *et al.* (2016). Accessions denoted as 'relicts' show extreme pairwise sequence divergence from other accessions. **A.** The trimodal distribution of pairwise genetic distances among accessions. The mode near zero reflects very close relationships of nearly identical accessions. The mode near 0.007 includes comparisons between relicts and non-relicts. **B**. Geographic locations of relicts (red) and non-relicts (blue) in Eurasia and North Africa. Pairs of near-identical accessions at least 1km apart are connected by green lines.

arms has also been shown to vary significantly between accessions (Sanchez-Moran *et al*. 2002). The magnitude of variation in recombination between these lines suggests that there could potentially be several genes varying between these accessions that have substantial effects on recombination patterns. Interestingly, the Cvi (Cape Verde Islands) and L*er* accessions were shown to have significantly lower chiasma frequencies than the other accessions analysed, suggesting that they possibly contain rare alleles of genes affecting recombination (Lopez *et al*. 2012).

Significant variation in recombination has also been observed between different F_1 accession cross hybrids in Arabidopsis, both when measured by genetic marker segregation in backcross progeny (Barth *et al*. 2001), and when measured by segregation of fluorescent markers expressed in pollen or seed (Ziolkowski *et al*. 2015). F_1 lines were shown to have recombination frequencies that often transgressed the parental phenotype, suggesting the presence of genetic modifiers of recombination that differ between accessions and combine to varying effects in F_1 hybrids. Measurement of the same F_1 hybrids over a range of intervals across the genome demonstrated that these accessions also differ in recombination distribution – F_1 crosses that had comparatively higher recombination in some intervals had lower recombination in others, suggesting that redistribution of crossovers may contribute to the observed variation as much as overall alterations in genome wide recombination frequency (Ziolkowski *et al*. 2015). Consideration of the recombination rate of all five intervals measured by Ziolkowski *et al*. (2015) reveals that the Cvi-0 and Can-0 accessions, when crossed to the Col-0 accession, produce the highest overall recombination rates in F_1 hybrids. While this contrasts somewhat with the evidence that Cvi has a lower chiasma frequency than many other accessions and could therefore be thought to have lower overall recombination (Sanchez-Moran *et al*. 2002), it is possible that the synergistic effect of combining Cvi and Col alleles of recombination modifiers in an F_1 hybrid has a different effect. In either case, it is apparent that recombination varies more between certain accessions, making them optimal choices for crossing to generate segregating mapping populations for the identification of recombination modifiers (Alonso-Blanco and Koornneef 2000), as demonstrated by recent identification of recombination modifiers by QTL mapping in segregating accession-cross populations (Esch *et al*. 2007; Ziolkowski *et al*. 2017).

Several Arabidopsis F_2 populations generated from accession crosses have shown significant variation in recombination (Salome *et al.* 2012; Ziolkowski *et al*. 2017). This variation does not correlate with the level of sequence variation between the parental accessions, indicating that sequence polymorphism is not a major influence on recombination frequency in these populations, which suggests that these differences are probably often due to the segregation of different alleles of recombination modifier genes (Salome *et al*. 2012). An additional point of interest that was derived from this data is that hotspots inferred from F_2 data don't correlate with hotspots deduced from

analysis of linkage disequilibrium (LD) in the global Arabidopsis population which is derived from an accession mixture (Drouaud *et al.* 2006; Salome *et al.* 2012). This could be interpreted as hotspots being accession specific, meaning that the hotspots inferred from the F_2 data probably relate to those found in the parental accessions, suggesting that the fine-scale distribution of recombination events also shows intraspecific variation in Arabidopsis (Salome *et al*. 2012).

1.11 Molecular and cellular factors influencing meiotic recombination frequency and distribution

Variation in recombination frequency has been well described between species, populations, individuals, sexes and across the genome. However, the factors underlying much of this variability are not well understood, and could belong to a range of molecular mechanisms, including variation in the expression or activity of components of the recombination machinery or differences in regulatory networks. Broad-scale functional conservation of the core recombination machinery in eukaryotes suggests that it is under strong selective constraints, although sequence divergence is common and could result in minor functional modifications that could potentially have an effect on recombination rate or distribution while maintaining overall protein function (Villeneuve and Hillers 2001; Mercier *et al.* 2015; Wright *et al*. 2015; Bomblies *et al*. 2015). Alternatively, other factors affecting recombination rate have been identified as factors influencing DNA accessibility or chromatin structure, or as components of minor recombination pathways, such as the FANCM protein which is specifically involved in the disruption of Class II CO intermediates (Wu and Lichten 1994; Fu *et al*. 2001; Fu *et al.* 2002; Merker *et al.* 2008; Crismani *et al*. 2012; Choi *et al.* 2013; Seguela-Arnaud *et al.* 2015; Lawrence *et al*. 2017).

Varying combinations of these factors could be responsible for a significant proportion of the observed variation in recombination. A multitude of studies in a range of organisms have provided considerable insight into the factors influencing meiotic recombination. However, this information is not sufficient to allow accurate prediction of the number of recombination events or their location, suggesting that there may be additional unknown factors affecting recombination that have yet to be identified (Lawrence *et al*. 2017). Observed recombination events do not always occur in what would seem to be the most probable location based on factors known to influence recombination – for example, recombination in *S.cerevisiae* is known to be predominantly associated with open chromatin and specific histone modifications, yet DSB precursors to recombination can often be found outside of open chromatin domains and some open chromatin sites are still recombination suppressed, indicating that additional unknown factors are also affecting the recombination landscape (Berchowitz *et al.* 2009; Ohta *et al.* 1994).

These factors can be roughly separated into *cis*-acting and *trans*-acting modifiers of recombination. *Cis*-acting modifiers act at the site of recombination, or on the same chromosome, where *trans*-acting modifiers exert an influence over recombination on the same and/or other chromosomes by encoding diffusible molecules (Ziolkowski *et al*. 2017; Yandeau-Nelson *et al*. 2006; Baudat and de Massy 2007), although there is an overlap between these definitions, as some modifiers involve interaction of *cis*- and *trans*-acting factors. Therefore, these definitions are used loosely to denote local and non-local effects on recombination. Variation in these features between natural populations can contribute to natural variation observed in meiotic recombination frequency.

1.12 *Cis*-acting modifiers of meiotic recombination

Cis-acting factors exert an influence over local recombination patterns, resulting in variation in crossover distribution across chromosomes in many eukaryotes. These effects range from smaller local effects, such as that of specific sequence determinants and polymorphism heterozygosity affecting recombination over a few to a few hundred base pairs of DNA, to the effects of large scale structural variants and variation in chromatin structure and gene density spanning the entire chromosome (Lawrence *et al*. 2017). A substantial amount of information is known about *cis*-acting factors affecting recombination, perhaps more than is known about those acting in *trans*, and this must be carefully considered when attempting to manipulate recombination experimentally (Lawrence *et al*. 2017). Manipulation of *cis*-acting factors themselves is likely to be difficult and have pleiotropic effects on the cell as many factors have essential roles beyond recombination, whereas experimental manipulation of factors acting in *trans* to alter recombination patterns would have to take into account that these factors act within the context of *cis*-acting factors and may therefore not have the expected effect when manipulated.

Starting at the smallest scale, the local DNA sequence itself has considerable influence over the formation of DSBs and development of recombination intermediates. In a wide range of species, the probability of a DSB occurring and being repaired at any one position in the genome is strongly influenced by the level of sequence heterozygosity, or interhomologue polymorphism, both at crossover hotspot scale (Borts and Haber 1987; Cole *et al*. 2010; Drouaud *et al*. 2013) and at a broader scale across chromosome regions (Ziolkowski *et al*. 2015; Yao and Schnable 2005). Elevated levels of local sequence polymorphisms have been shown to inhibit meiotic recombination in yeast, mice and Arabidopsis at the hotspot scale. Experimental addition of single-nucleotide polymorphisms (SNPs) to a recombination hotspot in *S.cerevisiae* and the presence of insertion-deletion (indel) polymorphisms in a hotspot in mice are associated with reductions in crossovers, although DSBs appear to be unaffected, suggesting a downstream effect on the processing of recombination intermediates (Borts

and Haber, 1987; Cole *et al*. 2010). Analysis of recombination hotspots between different Arabidopsis accession hybrids, each differing in the level of interhomologue polymorphism, revealed that an increase in polymorphism was also associated with a decrease in crossovers in *A.thaliana* (Drouaud *et al*. 2013). Additionally, crossovers have been shown to be suppressed between divergent sequences in a variety of plant hybrids, both natural and domesticated, including tomato, hexaploid wheat and maize (Rick 1969; Sears 1976; Yao and Schnable 2005; Dooner and He 2008). These sequence polymorphisms can directly affect the recombination process independently of variation in *trans*-acting modifiers, as analysis of polymorphic sequences introgressed into a near-isogenic background in maize still demonstrated differences in recombination, which must be attributed to the effect of local polymorphisms in *cis* (Yao and Schnable 2005).

The effect of sequence polymorphism on recombination is predicted to be mediated by anti-recombination activity of the mismatch-repair machinery (MMR), which involves binding and recognition of base-pair mismatches in heterozygous regions which form after strand invasion, or in later joint-molecule recombination intermediates, and recruitment of proteins which then influence the outcome of recombination intermediate resolution (Figure 8)(Harfe and Jinks-Robertson 2000; Chakraborty and Alani 2016). The theory postulated by Modrich and Lahue (2003) suggests that MMR unwinds or limits strand exchange or extension of the heteroduplex in the presence of mismatches, which is supported by evidence that some MMR mutants show extension of mitotic and meiotic conversion tracts (Chen and Jinks-Robertson 1998; Duroc *et al.* 2017). Consistent with this proposed anti-crossover role, mutants of MMR proteins show increased recombination in divergent yeast hybrids and at polymorphic hotspots in yeast, almost to the rate observed at the locus in the absence of polymorphisms (Borts *et al*. 1990; Martini *et al*. 2011; Hunter *et al*. 1996; Borts *et al.* 1989). Evidence for this effect has also been found in plants, where *msh2* MMR mutants show an increase in recombination in divergent Arabidopsis accession-cross lines when compared to wild-type, suggesting a role for MMR-mediated suppression of recombination between naturally divergent sequences (Emmanuel *et al.* 2006).

In addition to small-scale local effects, levels of heterozygosity can also affect recombination patterns at the megabase scale across entire chromosomes. Interestingly, examination of chromosomes containing a juxtaposition of regions of heterozygosity and homozygosity in Arabidopsis has revealed that crossovers preferentially form in large heterozygous regions at the expense of adjacent homozygous regions - an effect that appears to contradict the suppressive effect of heterozygosity on recombination observed at kilobase scales (Ziolkowski *et al*. 2015). This bias of recombination towards heterozygous regions is dependent on the Class I ZMM-dependent interfering crossover pathway, and the remodelling of crossovers across the chromosome is consistent with crossover interference

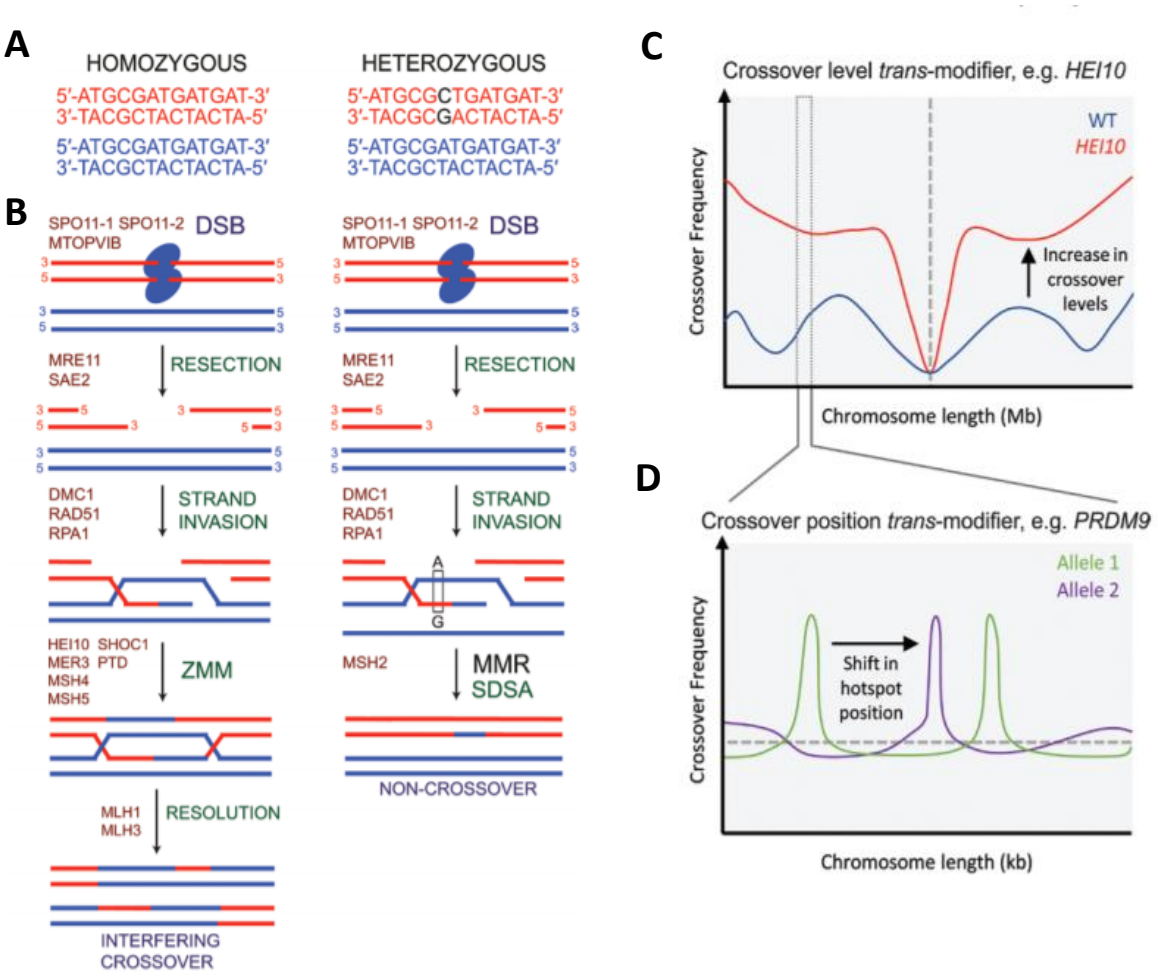


Figure 8. *Cis*-acting (A and B) and *trans*-acting (C and D) modifiers of meiotic recombination, adapted from Lawrence et al. (2017). **A**. Left: two homozygous alleles of a loci, Right: two heterozygous alleles. **B**. Schematic of DSB formation at loci with homozygous (left) and heterozygous (right) sequences. Associated recombination proteins are shown at each step. Heterozygous chromosomes form a mismatch after strand invasion, which is then targeted by MMR machinery, including MSH2, to dissolve the mismatched event and promote non-crossover repair. In the absence of mismatches (left) the ZMM pathway promotes crossover repair. **C**. Elevated recombination throughout chromosome arms in HEI10 overexpressor lines (red) compared to wild-type (blue) (Ziolkowski et al. 2017). Centromere denoted by grey dashed line. **D.** Different mammalian PRDM9 alleles (green and purple) produce distinct recombination hotspot landscapes (Baudat et al. 2010). Genome wide mean recombination rate denoted by grey dashed line.

causing the reciprocal decrease in homozygous regions. The mechanism underlying the bias in crossovers towards heterozygous regions, and the resulting decrease in homozygous regions, is unknown, although it is possible that it involves recruitment of a ZMM pathway component to mismatched substrates. Alternatively, as local heterozygosity appears to predominantly affect processing of recombination intermediates into crossovers rather than the formation of DSBs (Cole *et al.* 2010), it is possible that slower processing caused by mismatches could trigger signalling feedback pathways that result in additional 'late' DSBs in heterozygous regions. A higher number of DSBs in heterozygous regions could account for a higher probability of resolution as crossovers (Ziolkowski *et al*. 2015).

The same study also observed a decrease in crossover coincidence in heterozygous regions, which could potentially be attributed to mismatch-induced alteration of SC or axis properties that influence the crossover interference signal, although further experiments would be required to support this theory (Ziolkowski *et al.* 2015). Additionally, analysis of *fancm* mutants which show a considerable increase in Class II crossovers, demonstrated that the formation of non-interfering crossovers is less efficient in areas of sequence heterology, thereby further contributing to the bias towards interfering Class I crossovers and lower crossover coincidence in heterozygous regions (Girard *et al*., 2015; Ziolkowski *et al*., 2015). However, elevated non-interfering crossovers that form in *recq4a, recq4b* and *fidgl1* mutants are less sensitive to sequence heterologies (Girard *et al*. 2015; Seguela-Arnaud *et al*. 2015), suggesting that multiple distinct non-interfering crossover pathways may exist that are differentially affected by mismatches. The observation that different modifiers of recombination may be differentially sensitive to sequence polymorphism has important implications for the use of such modifiers in experimental manipulation of recombination, particularly in hybrid crops, and could affect the utility of specific modifiers under certain conditions.

While the importance of the influence of sequence polymorphism on recombination should not be underrated, especially in relation to its effects on recombination in hybrid crop species, polymorphisms on a larger scale can have even more substantial effects on recombination. Large structural polymorphisms, such as inversions, translocations, insertions and deletions, have also been shown to strongly inhibit recombination in many species (reviewed in Schwander *et al*. 2014; Thompson and Jiggins 2014). Observations of meiotic chiasmata in cytological experiments have demonstrated that heterozygous chromosome inversions, insertions and translocations all have direct effects on chiasma frequency and distribution (Rhoades 1968; Goldman and Hulten 1993; Gorlov *et al*. 1993; White and Morley 1955), potentially by inhibiting interaction with the homologous sequence. In addition to this direct effect, the observations that have been made in some species of recombination events within structural variants, particularly inversions, show that they are frequently

deleterious and result in unbalanced chromosome segregation and inviable gametes (Dobzhansky and Sturtevant 1938; Maguire and Riess 1994; Thompson and Jiggins 2014; Lowry and Willis 2010; Fransz *et al.* 2016; Rieseberg 2001). As measurements of recombination are frequently made by studying the segregation of loci or phenotypes in progeny, which requires the production of viable gametes, the underrepresentation of recombination events that generate inviable gametes in analysis may contribute to the apparent suppression of recombination documented within these structural variants. The deleterious consequences of these recombination events can also create selective pressure for suppression of recombination across these structural rearrangements. This effect on recombination can potentially alter chromosome pairing and synapsis in heterozygotes (Yandeau-Nelson *et al*. 2006), and therefore ultimately affect fertility, providing a mechanism to facilitate differentiation of populations, and eventually, speciation.

Interestingly, recombination within a different type of structural variant, sequence repeat arrays, is not always deleterious – unequal recombination between repeat sequences can generate insertions or deletions in the genome, which is believed to contribute to genetic diversity at plant disease resistance loci encoding NBS-LRR proteins, where sequence diversity is beneficial for host-pathogen co-evolution (Sasaki *et al*. 2010; Sudupak *et al*. 1993; Parniske *et al*. 1997; Michelmore and Meyers 1998). However, the effect of repetitive sequences on recombination rates within these loci is unclear, as in Arabidopsis, some NBS-LRR genes are recombination hotspots, but others are crossover suppressed (McDowell *et al*. 1998; Choi *et al*. 2016).

In Arabidopsis, natural structural variants between accessions have been shown to affect recombination in hybrid plants – an inversion on chromosome 4 in the Sha-0 accession suppresses recombination in Sha-0/Col-0 hybrids (Salome *et al*. 2012), an inversion encompassing the heterochromatic knob on chromosome 4 prevents local recombination in Col-0/L*er*-0 hybrids (Drouaud *et al*. 2006; Fransz *et al*. 2016), and tandem duplication of the RPP8 disease resistance gene in L*er*-0 relative to Col-0 has a suppressive effect on recombination (Choi *et al*. 2016). Similar rearrangements and associated recombination suppression and redistribution have been detected in maize, mouse, grasshoppers, *Caenorhabditis elegans* and humans (Rodgers-Melnick *et al*. 2015; Gorlov *et al*. 1993; White and Morley 1955; Zetka and Rose 1992; Hammarlund *et al*. 2005; Brown *et al*. 1998; Laurie *et al*. 1984).

This widespread phenomenon of suppression of recombination across structural variants has been linked to adaptation, as this can promote maintenance of genetic linkage between beneficial allele combinations (Thompson and Jiggins 2014), thereby creating 'supergenes' that increase fitness under specific conditions. Many examples of supergenes are available in nature, including nested inversions

causing linkage among butterfly mimicry genes, an inversion causing two recombination suppressed rearrangements in *Mimulus guttatus* (Yellow Monkeyflower) linked to adaptation of inland and coastal ecotypes, and an inversion in fire ant *Solenopsis invicta* resulting in linkage of genes regulating social behaviours (reviewed by Schwander *et al* 2014). Perhaps most significantly, evolution of heteromorphic sex chromosomes in mammals is thought to have involved a series of inversions on the Y chromosome, thereby limiting recombination with the X chromosome and allowing diversification of sex-specific genes (Marais and Galtier 2003).

Another form of structural variant, the insertion of transposable elements (TEs), also has an effect on recombination that may have evolutionary implications for some species, as the prevalence of TEs in some genomes is likely to significantly alter the distribution of recombination events and thereby affect patterns of genetic variation and linkage. Insertion of transposable elements can suppress recombination, both within the TEs themselves, and frequently within any surrounding genes or sequence elements as well. This is important for maintaining genome stability because ectopic recombination between TEs, which are frequently structurally polymorphic, can cause deleterious structural rearrangements, which would be a particularly pronounced problem in species which have high levels of TEs in the genome, such as many cereal crops (Caceres *et al*. 2001; Slotkin and Martienssen 2007; Choulet *et al.* 2014)**.** Studies in maize examining the *a1-sh2* and *bronze* loci showed that crossovers were skewed towards genic sequences and were not found in transposons (Brown and Sundaresan 1991; Yao *et al*. 2002; Yao and Schnable 2005; Fu *et al.* 2002; Dooner and He 2008; He and Dooner 2009). Furthermore, higher levels of TEs in maize are associated with lower levels of recombination, and individuals that are heterozygous for transposon insertions have lower local recombination in specific intervals, with transposons having a larger suppressive effect on recombination than that observed for SNPs (Dooner 1986; Dooner 2002; Wright *et al.* 2003; Yao *et al*. 2002; Dooner and He 2008). This effect is believed to be caused by the local condensation of chromatin and presence of DNA methylation around TEs affecting access of the recombination machinery to flanking genes, as TEs promote the formation of heterochromatin (Caceres *et al*. 2001; Fu *et al.* 2002).

Interestingly, there are some TEs in humans, mice and maize that are observed to be highly recombinogenic (Lichten and de Massy 2011; Smagulova *et al*. 2011; Liu *et al*. 2009). While this may seem contradictory, there is no evidence of a causal effect in these studies, and it seems likely that these TEs have higher recombination due to association with other local features that affect recombination, such as low-nucleosome density and *trans*-factor binding sites for recombination machinery or modifiers. Generally, studies in Arabidopsis indicate that there is no real correlation between crossovers and TEs (Drouaud *et al.* 2006), and that any co-occurrence observed over a broad-

scale is probably due to coincidence with chromatin structure (Wright *et al.* 2003), although it is possible that different classes of TEs could have different effects on recombination.

Beyond sequence heterologies and structural variants, specific sequence-motifs have also been found to influence crossover and DSB distributions. In plants, crossover hotspots are associated with AT-rich, CTT-repeat and CNN-repeat motifs (Horton *et al*. 2012; Choi *et al*. 2013; Choi *et al*. 2016; Wjinker *et al*. 2013; Shilo *et al*. 2015). AT-rich sequences are correlated with low nucleosome occupancy, which is more accessible to the recombination machinery (Segal and Widom 2009). These sequences are often found around gene promoters and terminators and also correspond to strong DSB hotspots in yeast (Lichten and Goldman 1995; Petes 2001; Schultes and Szostak 1991, Baudat and Nicolas 1997, Pan *et al*. 2011). This is consistent with evidence in plants that indicates hotspots are frequently found around AT-rich sequences in gene promoters and terminators (Horton *et al*. 2012; Choi *et al*. 2013; Choi *et al*. 2016; Wjinker *et al.* 2013), suggesting that this is a conserved mechanism of hotspot designation. A second class of recombination hotspots identified in fungi affect DSB frequency via an effect on transcription factor (TF) binding (Lichten and Goldman 1995; Petes 2001). While no equivalent observations have been made in plants, it is possible that this feature is also conserved, and that polymorphisms in TF-binding motifs could contribute to variation in recombination patterns.

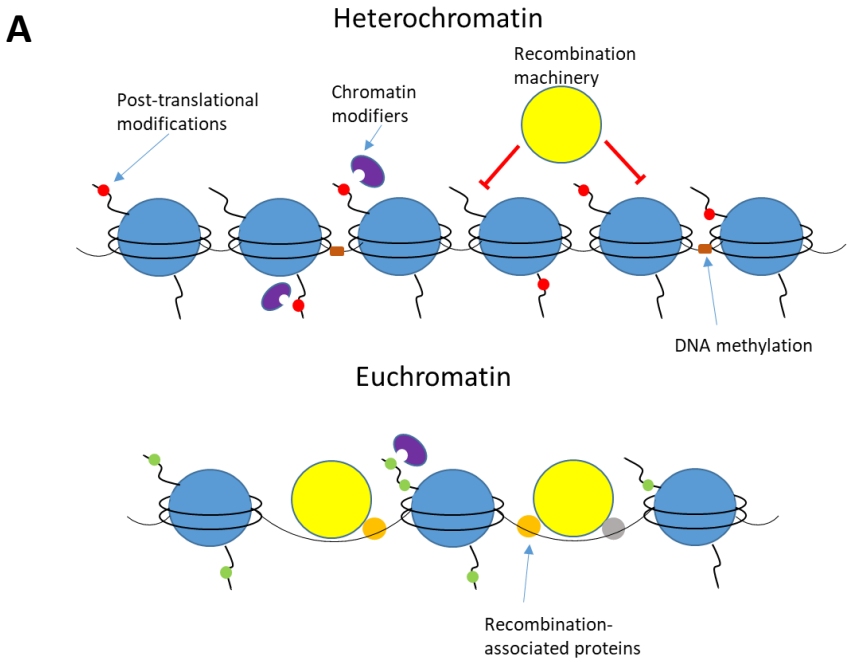
Plant CTT and CNN-motifs are also enriched around promoters, specifically at the +1 nucleosome relative to the transcription start site, thereby overlapping hotspots and areas of high historical recombination (Choi *et al.* 2013, Choi *et al.* 2016). In contrast to AT-rich motifs, these sequences are associated with H2A.Z histone variant occupancy and the H3K4^{me3} histone mark, factors that are associated with active recombination in diverse eukaryotes (Borde *et al*. 2009; Acquaviva *et al*. 2013; Choi *et al*. 2013), which could suggest a link between hotspot motifs and chromatin organisation at the 5' end of genes.

CTT-motifs also bear a marked resemblance to C-rich hotspot motifs in mammals which are bound by the PRDM9 protein which promotes DSB formation (Baudat *et al*. 2010; Myers *et al*. 2010; Parvanov *et al*. 2010). This interaction between *cis* and *trans* factors demonstrates that factors cannot be assessed in isolation – local recombination frequency is the result of many interacting factors. However, no homolog of *PRDM9* has been identified in yeast or plants, suggesting that CTT-motifs perform a different function in non-mammalian species (Zhang *et al*. 2012). Yeast hotspots do not generally demonstrate specific consensus sequences that would suggest binding of a specific modifier, and replacement of the sequence underlying the break still results in DSBs, indicating that in yeast higher order chromatin control has an over-riding effect above sequence variation (de Massy and Nicolas 1993; Baudat and Nicolas 1997). Interestingly, vertebrates that have lost *PRDM9* revert to

hotspots found in accessible gene promoters (Petes 2001; Brick *et al*. 2012; Auton *et al*. 2013), suggesting that the hotspot designation based on sequence-motifs promoting accessibility that is found in plants, birds and fungi could be an ancestral pattern (Pan *et al*. 2011; Choi *et al*. 2013; Shilo *et al*. 2015).

While variation in the DNA-sequence underlies a large proportion of variation in recombination observed across the genome and between populations, particularly if heterozygosity is present in the population, additional factors also affect local recombination, and could therefore be considered *cis*-acting modifiers of recombination. Broad-scale recombination patterns across chromosomes are affected by domains of heterochromatin - condensed, heavily methylated DNA - and euchromatin - gene-rich, repeat-poor domains with open chromatin (Figure 9) (Yelina *et al*. 2012; Choi *et al*. 2013). Generally, these features correlate with gene-density, and both components are linked to recombination. Recombination hotspots in mammalian species, particularly mice, are known to overlap genes (Smagulova *et al* 2011), and a large proportion of recombination in yeast also occurs in genes and adjacent to promoters (Alani *et al.* 1994, Oliver *et al.* 1992, Buard *et al.* 2009). Plants in general exhibit recombination rates in genes that are 10 to 100-fold higher than the genome average (Schnable *et al.* 1998). Gene-rich regions are significantly more recombinationally active in maize, wheat and barley (He and Dooner 2009, Fu *et al.* 2001; Civardi *et al*. 1994; Gill *et al*. 1996), and most known recombination hotspots in Arabidopsis are found in gene promoters (Drouaud *et al*. 2013). This does not correlate with transcription levels but is believed to be related to the accessibility of the chromatin to protein factors (Fu *et al*. 2001).

Heterochromatin is generally suppressed for recombination, therefore regions of the chromosome with significant proportions of it, such as the centromere, are largely recombinationally inert. Centromeric heterochromatin has been shown to be suppressed for crossovers in Arabidopsis, maize, rye and other eukaryotes (Yelina *et al.* 2015, Ziolkowski *et al.* 2015, Yao and Schnable 2005, Rodgers-Melnick *et al.* 2015, Kagawa *et al.* 2002). Consistent with this, recombination hotspots in multiple species have been found to associate with the open features of euchromatin – work in mice, yeast, humans and plants all show higher recombination in areas of low nucleosome density and open chromatin (Yelina *et al.* 2015, reviewed in Smukowski and Noor 2011). This is predominantly due to an effect of chromatin-accessibility on DSBs – in yeast, disruption of nucleosomes in the *PHO5* promoter increases DSBs locally, which is likely to have a knock-on effect on recombination as a large proportion of DSBs are repaired as COs in yeast (Wu and Lichten 1994; Chen *et al.* 2008; Mancera *et al*. 2008). This increase in DSBs is consistent with evidence that meiosis-specific alteration of chromatin in yeast affects the position and frequency of DSBs (Ohta *et al*. 1994). However, while heterochromatic suppression of recombination can be used as a general predictor of broad



Graphs showing recombination rate (cM/Mb) and chromatin landscape around promoters in *Arabidopsis thaliana* removed for copyright reasons.

Figure 9. ***Cis*-acting factors associated with variation in meiotic recombination. Partially adapted from Choi *et al.* (2013). A.** Simplified schematic of euchromatin vs heterochromatin accessibility to recombination machinery. Blue circles denote histones, green and red circles denote histone modifications associated with active and repressive chromatin states respectively. Yellow circles denote recombination proteins. DNA methylation is denoted by brown rectangles on the DNA (black line). **B.** Recombination rate and chromatin landscape around promoters in *Arabidopsis thaliana*, relative to the transcription start site (TSS). Recombination rate (cM/Mb) shown in black, H2A.Z levels shown in red, $H3K4^{me3}$ levels shown in pink, low nucleosome density (LND) shown in green and DNA methylation shown in blue. The graph on the far left shows an overlay of all variables.

recombination patterns, it does not provide an accurate picture of crossover distribution. DSBs in yeast, the necessary precursors to recombination, are frequently found outside of open chromatin domains. This suggests that other factors, in some cases thought to be histone modifications, can override the heterochromatin effect on DNA accessibility which is also likely to affect the activity of recombination complexes (Berchowitz *et al.* 2009; Mancera *et al.* 2008; Drouaud *et al*. 2013). Additionally, open chromatin is not always sufficient to ensure recombination - some MNase sensitive sites are still suppressed for DSBs, and therefore for recombination, again indicating interaction of additional factors influencing recombination (Ohta *et al.* 1994).

DSBs and recombination are also related to more specific chromatin features – including histone variants, histone modifications and associated epigenetic modifications to DNA (Figure 9B). DNA methylation, which is associated with heterochromatin and repressive chromatin marks, has been shown to affect the distribution of recombination events (Yelina *et al*. 2012). Loss of DNA methylation in the *A.thaliana met1* and *ddm1* mutants results in a redistribution of crossovers away from the euchromatic chromosome arms and towards the centromere, although the total number of recombination events remains the same (Mirouze *et al.* 2012; Yelina *et al.* 2012; Melamed-Bessudo and Levy 2012). These changes cannot be wholly attributed to changes in chromatin accessibility as transcription remains unaffected in hotspots where recombination increases, and crossovers are still suppressed in pericentromeric heterochromatin, suggesting that multiple layers of epigenetic modification and histone marks contribute to this phenotype (Yelina *et al.* 2012 Melamed-Bessudo and Levy 2012). In regards to specific histone marks affecting recombination, in yeast DSB hotspots are associated with histone hyperacetylation (Yamada *et al*. 2004), and an increase in histone acetylation at the *HIS4* hotspot in *S.cerevisiae* results in a concomitant increase in DSBs (Merker *et al*. 2008). Consistent with this, hypoacetylation is associated with suppression of recombination in telomeres (Perrella *et al.* 2010), and alterations to histone acetylation in *C.elegans* and Arabidopsis result in altered distributions of DSBs and crossovers (Wagner *et al*. 2010; Perrella *et al*. 2010). Arabidopsis recombination hotspots have also been shown to associate with the H2A.Z histone variant in gene promoters (Figure 9B)(Choi *et al*. 2013), and decreases in H2A.Z deposition affect recombination processing.

Histone methylation marks can also be used as predictors of DSB and crossover activity – the $H3K4^{me3}$ mark, a feature of active chromatin, is enriched around DSB sites in *S.cerevisiae*, and recombination hotspots in Arabidopsis and mice (Borde *et al*. 2009; Choi *et al*. 2013; Smagulova *et al*. 2011). In mice, this association can be partially explained by the PRDM9 protein, which deposits $H3K4^{me3}$ and is believed to recruit the DSB machinery. However, $H3K4^{me3}$ alone is not sufficient to generate a hotspot, and it is believed that only a subset of the marks have a recombination function, an assertion which is

supported by evidence of testis-specific H3K4^{me3} marks in mouse (Smagulova *et al*. 2011). However, no *PRDM9* homolog is known in plants or yeast, and it is likely that in these species this association is linked to the generation of open accessible chromatin (Zhang *et al*. 2012; De Massy 2013). While individual histone marks may be poor predictors of recombination distribution, collectively they explain more of the variation in recombination distribution than they do individually, facilitating identification of regions of the genome that are likely to have high or low recombination.

1.13 *Trans*-acting modifiers of meiotic recombination

Modifiers of recombination also exist which act on recombination in areas of the genome away from their immediate locus (Yandeau-Nelson *et al*. 2006; Lawrence *et al*. 2017), through the medium of diffusible molecules. These are known as *trans*-acting modifiers, as they can affect recombination on other chromosomes, although they act within the *cis*-context of sequence and chromatin features at specific sites. Relatively few *trans*-acting modifiers of recombination are known, particularly in plants where recombination distribution is broadly affected by factors acting in *cis* affecting DNA accessibility. However, modifiers have been identified that can significantly alter global recombination patterns and frequencies, and variation within these modifiers is likely to contribute to natural variation observed in recombination.

Trans-acting modifiers of meiotic recombination could conceivably belong to many different molecular pathways, including the recombination machinery itself, or possibly regulators of protein stability or interactions that could influence the recombination complex and affect the outcome of recombination intermediate resolution. The proteins involved in the formation of the recombination complex include many pro- and anti-crossover factors that can modulate the levels of recombination in *trans* (See Chapter 1.4). Mutation of the anti-crossover proteins *fancm, recq4a/recq4b*, and *figl1* in Arabidopsis result in significant increases in recombination across the genome, as more recombination intermediates are resolved as crossovers (Crismani *et al*. 2012; Seguela-Arnaud *et al*. 2015; Fernandes *et al*. 2017). Although these modifiers aren't known to be involved in natural variation in recombination, they do influence recombination frequency in *trans* across the genome, having a substantial effect when mutated, resulting in changes in recombination frequency of several fold. While natural alleles of these genes varying in their effect on recombination have not been identified, with the exception of a null allele of *recq4b* in L*er* that is compensated for by a functional *RECQ4A* homolog, it is possible that different alleles could exist that had minor differences in expression level or substrate binding (Seguela-Arnaud *et al*. 2015). As recombination can be massively increased in mutants with no apparent ill-effects (Crismani *et al*. 2012; Mercier *et al.* 2015), variants upregulating recombination would not necessarily be selected against and could exist in natural populations.

Evidence for *trans*-acting modifiers of meiotic recombination has been identified in many diverse eukaryotes, including both plant and animal species (Timmermans *et al*. 1997; Yandeau-Nelson *et al*. 2006; Esch *et al*. 2007; Baudat *et al*. 2010, Parvanov *et al*. 2010; Bauer *et al*. 2013; Reynolds *et al*. 2013; Qiao *et al*. 2014; Li *et al.* 2016; Ziolkowski *et al*. 2015, 2017). Extensive work in mammals has identified and characterised the genes underlying several *trans*-acting recombination modifiers, including the *PRDM9*, *RNF212* and *HEI10* genes (Baudat *et al*. 2010, Parvanov *et al*. 2010; Reynolds *et al*. 2013; Qiao *et al*. 2014; Rao *et al*. 2017), however, in plants many of the modifiers remain unknown – while evidence of their effect on recombination has been observed, the underlying genes have not been identified. For example, experiments in maize have shown clear evidence of *trans*-acting modifiers and demonstrated their effects in diverse backgrounds on an identical *a1-sh2* interval sequence in the absence of polymorphic *cis*-acting factors (Timmermans *et al*. 1997; Yandeau-Nelson *et al*. 2006), however the underlying modifier genes remain unknown.

A common strategy for identification of *trans*-acting loci affecting crossovers involves using recombination as a quantitative trait for QTL mapping. This has allowed identification of QTLs affecting recombination in Arabidopsis, maize, mouse and wheat (Esch *et al*. 2007; Li *et al*. 2016; Ziolkowski *et al*. 2015; Ziolkowski *et al*. 2017; Pan *et al*. 2017), although almost invariably the underlying genes were not identified. One study in plants that has managed to identify and confirm a *trans*-acting modifier of meiotic recombination underlying a QTL, utilised Arabidopsis accession-cross populations to identify two QTLs affecting recombination that differed between Col and L*er* accessions (Ziolkowski *et al*. 2017). One of these QTL was confirmed as the HEI10 E3 ubiquitin ligase which has a role in the class I crossover pathway, although the causal variant and its effect on the gene are unknown. *HEI10* was shown to have a dosage dependent effect on recombination, with overexpression lines exhibiting an increase in recombination up to 2-fold across the genome (Figure 8) (Ziolkowski *et al*. 2017; Serra *et al*. bioRxiv). This is consistent with observations of haploinsufficiency in *hei10* and related *rnf212* mutants in Arabidopsis and mice, which contrast with observations made in other *zmm* mutants and supports a dosage-dependent effect of *HEI10* on recombination.

In mammals, HEI10 and RNF212 participate in a SUMO-ubiquitin relay that encourages stabilisation of recombination intermediates to promote crossover resolution (Reynolds *et al*. 2013; Qiao *et al*. 2014; Rao *et al*. 2017). While the targets of HEI10 in plants are unknown, it is possible that it performs a similar function. In support of *HEI10* being a significant influence on natural variation in recombination in Arabidopsis, variation in mammalian *HEI10* and *RNF212* genes is associated with variation in genome wide recombination rates in humans and cattle (Kong *et al*. 2008; Sandor *et al*. 2012).

Another significant *trans*-acting modifier that causes differences in recombination between natural populations is the PRDM9-zinc finger protein in mammals, which binds specific sequence motifs and deposits a H3K4^{me3} mark believed to be involved in recruitment of DSB machinery (Baudat *et al*. 2010, Parvanov *et al*. 2010). Different alleles of *PRDM9* with different DNA-binding profiles are associated with variation in hotspot usage in humans, cattle and mice (Figure 8), and while overall global recombination levels do not appear to change between *PRDM9* alleles, significant variation in the distribution of recombination events is observed (Kong *et al*. 2008; Baudat *et al*. 2010; Fledel-Alon *et al*. 2011; Sandor *et al*. 2012; Davies *et al*. 2016). Further work needs to be performed in plants to identify modifiers affecting the distribution of recombination events, as no modifiers of a similar function to PRDM9 have been identified, leaving large amounts of variation in crossover distribution unaccounted for.

Another point of interest in the study of *trans*-acting modifiers, that may influence the evolutionary implications of a modifier and affect its utility in experimental manipulation, is that while the HEI10 and PRDM9 modifiers have shown an effect across the entire genome, this is not the case for all identified *trans*-acting modifiers. Region-specific modifiers have been identified in other eukaryotes, including *S.pombe* (De Veaux *et al.* 1992), and *Neurospora crassa* (Catcheside 1977), and have also been indicated in maize, through work showing that different regions of the genome vary in recombination frequency at differential rates in different genetic backgrounds (Yandeau-Nelson *et al* 2006). This suggests that while variation in *trans*-acting factors may account for a significant proportion of the variation in recombination observed between natural populations, the effects may not be consistent across the genome. This would in turn affect the pattern of genetic variation and linkage across the genome, and thereby the rate of adaptation, which could have implications for the differential rates of adaptation observed between genes (Stapley *et al.* 2017).

Although many recombination modifiers may have region-specific effects, there are other factors responsible for variation in recombination that have more expansive effects on crossovers. An additional source of *trans*-acting effects on recombination is polyploidy, which can act to suppress recombination across entire chromosomes, thereby aiding the stabilisation of polyploid lines. Polyploidy is observed across most eukaryotic phyla, but it is especially prevalent within plants, which causes challenges to completion of meiosis with the correct chromosome complement (Otto 2007, Yant and Bomblies 2015). Recombination between homeologous chromosomes must be suppressed to promote pairing of homologues and limit the formation of multivalents that can result in unbalanced chromosome segregation (Yant and Bomblies 2015; Bomblies *et al*. 2016). The *Ph1* locus acts in *trans* to regulate homeologous pairing and crossovers in hexaploid bread-wheat (Roberts *et al*. 1999) as although homeologous chromosomes are frequently divergent, they also contain related

genes which may share sequence homology and allow ectopic recombination to occur. The *Ph1* locus, which contains several linked genes, affects several parts of the recombination pathway, including chromosome pairing, chromatin and axis dynamics and licensing of MLH1 foci to form crossovers (Mikhailova *et al*. 1998; Martin *et al*. 2014). Separation of the linked genes at the *Ph1* locus could reveal whether these each account for a separate function in ectopic recombination and pairing suppression.

The evolution of stable chromosome segregation in polyploid meiosis is believed to entail both a reduction in COs and a redistribution of chiasmata towards the ends of the chromosome, which may act to minimise chromosomal entanglements and deleterious associations (Bomblies *et al*. 2016). Examination of neo-autotetraploid and evolved natural autotetraploid *Arabidopsis arenosa* demonstrates that the fully evolved polyploid has a lower CO frequency than both the diploid of the species and new autotetraploids, suggesting that there is strong selection to reduce the high levels of recombination that result in multivalents in neo-polyploids (Yant *et al*. 2013; Wu *et al.* 2013). Increased recombination frequencies are frequently observed in newly formed Brassica polyploids, suggesting that the addition of chromosomes can act in *trans* to alter recombination frequency and distribution, before selection pressures reduce CO to maintain genome stability (Leflon *et al*. 2010; Pecinka *et al*. 2011; Yant *et al.* 2013). In addition to decreases in CO number, distal CO localisation in evolved autotetraploid plants is positively correlated with accurate chromosome segregation, which is consistent with the hypothesis that both a reduction of CO number and a relocation of events to the ends of the chromosomes could aid chromosome segregation in polyploid meiosis (Myers 1945; Hazarika and Rees 1967; Bomblies *et al.* 2016).

To understand the evolution of genome stability in polyploids, an analysis of the genomes of recently evolved tetraploids of *A.arenosa* was performed (Hollister *et al*. 2012; Yant *et al*. 2013; Wright *et al*. 2014). This analysis showed clear signatures of selection in conserved components of the meiotic axis and SC, and it was proposed that modification of these components could have an influence on increasing crossover interference distance in autotetraploids, thereby contributing to the skew towards distal COs that can promote accurate chromosome segregation (Hollister *et al*. 2012; Yant *et al*. 2013; Wright *et al*. 2014; Bomblies *et al.* 2016). These factors work in conjunction with elements such as *Ph1* to ensure meiotic stability in polyploids - the high frequency of polyploidization seen in plants makes it likely that several convergent strategies have evolved to stabilise genome inheritance by reducing homeologue interactions. While polyploidy does not necessarily fall into the same category as the other *trans*-acting modifiers described here that could potentially be easily manipulated, it still serves an important evolutionary purpose in regulating recombination frequency and distribution. These effects must be considered when attempting to manipulate recombination in

crop species, many of which exhibit polyploidy, as they could limit the effects of other modifiers, and manipulation of these limitations to facilitate modification of recombination may cause ectopic recombination between homeologues and destabilise chromosome segregation.

1.14 Natural variation in *Arabidopsis thaliana* recombination

Variation in *cis*-acting and *trans*-acting factors influencing recombination between natural accessions produces the extensive variation in recombination frequency and distribution observed between these natural Arabidopsis populations. For example, significant sequence variation between Arabidopsis accessions can have considerable effects on recombination in accession hybrid plants, given the known influence of sequence heterologies on crossovers (Drouaud *et al*. 2013; Alonso-Blanco *et al*. 2016). These heterologies can substantially reduce recombination within specific accession hybrids, for example the chromosome 4 inversion observed in the Sha-0 accession limits recombination between Sha-0 and other accessions on chromosome 4 (Salome *et al*. 2012). Sequence heterologies can also produce variation within natural accession populations – while *A.thaliana* is predominantly self-fertilising, there is still significant evidence of outcrossing. Rapid decay of linkage disequilibrium and direct observation of residual heterozygosity and local outcrossing in natural populations suggests that substantial outcrossing does occur in natural Arabidopsis populations (Cao *et al*. 2011; Horton *et al*. 2012; Choi *et al*. 2013; Bomblies *et al*. 2010). The polymorphisms produced by outcrossing could be maintained in the population by balancing selection, either due to heterozygotes having an advantage over homozygotes, or frequency dependent selection or inconsistent selection pressures resulting in maintenance of different variants (Mitchell-Olds 2001; Yang *et al*. 2015; Alonso-Blanco *et al*. 2016; Ziolkowski and Henderson 2017). While this would act primarily to increase genetic diversity in the population, it could also have a significant effect on recombination patterns by acting as a source of sequence heterology that may affect CO formation or resolution (Mitchell-Olds and Schmitt 2006; Drouaud *et al*. 2013; Ziolkowski *et al*. 2015; Chakraborty and Alani 2016; Ziolkowski and Henderson 2017). Differential rates of outcrossing between Arabidopsis populations (Bomblies *et al*. 2010) could lead to differences in the levels of heterozygosity and, subsequently, different effects on recombination, thereby contributing to variation in recombination between natural populations. Even if the heterozygosity resulting from outcrossing was frequently lost due to inbreeding, or directional or purifying selection, it may still have transient effects on recombination patterns and must therefore be considered as a potentially significant factor influencing recombination in natural *A.thaliana* populations.

Arabidopsis accessions have also been demonstrated to vary for other factors that affect recombination in *cis*. Divergent accessions show variation in centromeric repeats, levels of

heterochromatin and DNA-methylation profiles (Ito *et al*. 2006; Tessadori *et al*. 2009; Pignatta *et al*. 2014; Kawakatsu *et al*. 2016). Analysis of epiRILs in Arabidopsis shows variation in recombination between epigenetically divergent lines (Colomé-Tatché *et al*. 2012), and as a significant amount of heritable natural variation in cytosine methylation has been documented between Arabidopsis accessions (Zhang *et al.* 2008, Vaughn *et al.* 2007), it is possible that this contributes to the differences in recombination observed between accessions.

While variation in factors affecting recombination in *cis* likely contributes to natural variation in recombination, ultimately many of these features are conserved and any observed variations between accessions are small (Colomé-Tatché *et al*. 2012), making them unlikely candidates to underlie the drastic differences in recombination observed between accessions. This suggests that *trans*-acting factors are responsible for much of the observed variation, an observation that is consistent with the identification of the *HEI10* modifier whose accession-specific alleles have a significant effect on recombination frequency across the genome (Ziolkowski *et al*. 2017). However, not all variation present between accessions is explained by these known *trans*-acting modifiers, meaning that additional experiments could identify significant modifiers of meiotic recombination frequency in Arabidopsis. The conserved nature of the recombination process means that modifiers identified in Arabidopsis may have similar functions in other plant species, and potentially other eukaryotes.

1.15 Evolutionary implications of natural variation in recombination

Recombination is a fundamental evolutionary process that facilitates adaptation by generating new allele combinations and promoting the separation of linked genetic variants, thereby allowing natural selection to act more efficiently on individual variants (Marais and Charlesworth 2003; Otto 2009; Webster and Hurst 2012; Stapley *et al*. 2017). However, while recombination can create new favourable allele combinations, it can also break apart advantageous combinations or co-adapted loci, which can be extremely detrimental to the fitness of an organism (Charlesworth and Barton 1996; Rice 2002; Stapley *et al*. 2017). Given the importance of recombination for adaptation, and its requirement in many species for successful gametogenesis, it could be expected that there would be strong selective constraints limiting the amount of recombination to within an optimum range, as extremely high or low recombination rates are likely to be deleterious as is observed in humans (Dapper and Payseur 2017; Stapley *et al.* 2017; Alves *et al.* 2017). Nevertheless, significant variation in recombination has been observed in nature between species, populations and individuals. The reasons behind the evolution of this variation are still largely unknown, as many studies focus on explaining the evolutionary advantage of recombination as a general process rather than recombination variation, although some theoretical evolutionary models posit that selection may act to change recombination in different populations in response to differences in environmental

conditions (Dapper and Payseur 2017; Stapley *et al.* 2017). Variation in recombination could contribute to relative fitness differences between populations by affecting the rate of adaptation to different environmental conditions (Dapper and Payseur 2017; Stapley *et al.* 2017).

Recombination and mutation are two of the principal processes involved in the generation of genetic variation, and both are therefore key factors influencing evolution and adaptation (Stevison *et al.* 2017). Both processes have been shown to be affected by environmental factors, with recombination varying considerably with conditions such as temperature and level of pathogen infection, in addition to varying with organism age in some species (Henderson and Edwards 1968; Hoffman and Parsons 1991; Kovalchuk *et al*. 2003; Andronic 2012; Kerstes *et al*. 2012; Martin *et al.* 2015; Stevison *et al.* 2017; Stapley *et al.* 2017; Morgan *et al.* 2017). While in some cases the changes in recombination observed under differing conditions may be a by-product of the adaptation of another process or feature, in others it is possible that selection may be acting directly on recombination itself as changes in recombination are likely to offer adaptive benefits and affect fertility (Otto and Barton 2001; Smukowski and Noor 2011; Webster and Hurst 2012; Bomblies *et al*. 2015; Wright *et al.* 2015; Morgan *et al.* 2017; Stapley *et al*. 2017; Dapper and Payseur 2017). For example, the increases in recombination sometimes observed with increased age may be selected for to help to combat the chromosome mis-segregation issues observed in some older organisms by increasing the linkages between homologous chromosomes and thereby increasing fertility (Kong *et al.* 2004; Coop *et al*. 2008; Fledel-Alon *et al.* 2011; Baudat *et al*. 2013). Alternatively, increases in recombination observed in organisms infected with pathogens are believed to provide an indirect fitness advantage by increasing genetic diversity in the offspring, which may prevent a pathogen establishing itself in a genotype that it is well-adapted to exploit (Kovalchuk *et al.* 2003; Salathe *et al.* 2009; Andronic 2012; Singh *et al.* 2015; Stapley *et al.* 2017; Dapper and Payseur 2017). Therefore, selection acting specifically to alter recombination rate may be direct, acting to increase recombination to improve gamete viability, or indirect, acting to alter recombination to affect genetic diversity and selection efficiency in the population (Stapley *et al.* 2017; Dapper and Payseur 2017).

Direct advantages to altered recombination patterns, in terms of gamete viability, are not always apparent in populations displaying variation in recombination. Therefore, it is often suggested that the main advantage of recombination is likely to be its indirect advantage of facilitating natural selection and adaptation through the breaking of genetic linkage (Stapley *et al*. 2017). Linkage of variants subject to opposing directions of selection prevents natural selection acting on these variants with full efficiency, often limiting the fixation of beneficial variants or the removal of deleterious mutations (Morrell *et al*. 2004; Marais *et al*. 2004; Comeron 2017). Furthermore, linkage of neutral variants to either deleterious mutations or beneficial variants can lead to a loss of genetic diversity at

neutral sites, as selection acts to remove the deleterious variants and move the beneficial variants to fixation, which in either case results in the loss of linked segregating variants through what is known as background selection or a selective sweep, respectively (Maynard Smith and Haigh 1974; Kim and Stephan 2002; Comeron 2017). The size of the genetic block affected by this loss of diversity is dependent on both the strength of selection and the rate of recombination (Maynard Smith and Haigh 1974; Kim and Stephan 2002; Comeron 2017). Therefore, recombination provides both a short-term advantage by facilitating the rapid removal of deleterious variants, and a long-term advantage in generally increasing selection efficacy and genetic diversity.

However, as recombination can also reduce fitness through breaking beneficial variant combinations, the rate of recombination that would be advantageous in the organism is dependent on how the variants are associated and how well the organism is adapted to its current conditions (Rice 2002; Stapley *et al*. 2017; Dapper and Payseur 2017). Genetic drift, which is particularly important in populations with a small effective population size and high levels of inbreeding as found in many *A.thaliana* accessions, frequently generates associations between beneficial and deleterious variants, thereby resulting in increased selection for recombination to promote their separation (Marais *et al*. 2004; Barton and Otto 2005; Bomblies *et al.* 2015; Dapper and Payseur 2017; Yant and Bomblies 2017; Stapley *et al.* 2017). Conversely, selection for increased fitness in many populations will result in an increase in beneficial variant associations, which will not promote selection for increased recombination as separation of variants would reduce fitness (Dapper and Payseur 2017). Under these conditions only variants associated with a reduction in recombination will increase in frequency due to selection (Feldman and Liberman 1986; Dapper and Payseur 2017). However, this situation is uncommon as most natural populations do not achieve full equilibrium due to mutation and migration introducing further variation (Dapper and Payseur 2017). Therefore, an abundance of beneficial variant associations and subsequent selection for reduced recombination is unlikely to be responsible for much of the observed natural variation in recombination rate.

Interestingly, while selection for increased recombination is expected to occur when traits are experiencing strong directional selection, as this would promote separation of variants and improve selection efficiency, higher recombination rates are also expected in populations experiencing heterogeneity in selection (Charlesworth 1976; Otto and Barton 2001; Stapley *et al*. 2017). This heterogeneity could occur due to fluctuations in environmental conditions, or the continuous evolution of an invading pathogen, causing allele combinations that were advantageous to become disadvantageous, resulting in an increase in deleterious allele associations when conditions change (Otto 2009; Stapley *et al*. 2017; Dapper and Payseur 2017). While there is minimal direct evidence of this in sexually reproducing species, as most recombination experiments are not designed with a view

towards testing selection or adaptation, some studies have found evidence that fluctuating environments can affect recombination rate, with three to four-fold changes observed in some Drosophila populations compared to those under more stable conditions (Charlesworth 1976; Derzhavets *et al*. 1996; Lenormand and Otto 2000; Stapley *et al.* 2017; Dapper and Payseur 2017). Although increased levels of recombination may help to break up maladaptive variant combinations in the short-term, the long-term benefits of this are dependent on the frequency of the environmental fluctuations, and slow or infrequent oscillations are unlikely to generate sufficient selection for increased recombination (Charlesworth 1976; Dapper and Payseur 2017; Stapley *et al*. 2017).

Variation in recombination in response to differential selection generated by changes in environmental conditions, or internal changes such as the level of ploidy, is likely to be key to differential adaptation between populations (Yant *et al.* 2013; Bomblies *et al*. 2016; Stapley *et al*. 2017). Recombination levels in natural populations may be modulated by how well adapted the genotype is to current conditions, with genomes containing deleterious mutations in linkage with other variants generating selection for increases in recombination (Dapper and Payseur 2017; Stapley *et al.* 2017). While theoretical studies of selection and recombination promote understanding of the possible fitness consequences of variation in recombination, more experimental work needs to be completed to determine how changes in recombination actually affect adaptation and to clarify whether natural variation in recombination is predominantly due to selection or factors such as genetic drift (Stapley *et al.* 2017; Dapper and Payseur 2017).

Variation in recombination rate could also be prevalent in natural populations for other evolutionary purposes. There are some indications that reductions in recombination rate could act as a pre-adaptation for the transition to asexuality, as many asexual species perform a modified version of meiosis with low levels of recombination (Lenormand *et al*. 2016; Haag *et al*. 2017). This is because recombination can lead to a loss of heterozygosity in many asexual species which can have negative fitness consequences (Lenormand *et al*. 2016; Haag *et al*. 2017). However, it is currently unclear whether the adjustment to a lower recombination rate generally occurs before or after the transition to asexuality, although observations of low recombination rates have been made in close sexual relatives of asexual species and are suggested to make the transition to asexuality more likely by precluding the possible deleterious effects of crossovers (Rauwolf *et al.* 2011; Haag *et al*. 2017). Therefore, it is possible that selection for reduced recombination could be linked to the transition to asexuality, and natural variation in recombination rate could reflect the probability of a species using that evolutionary strategy. It is also possible that variation in recombination may influence the process of speciation, as differences in recombination between populations can drive genome divergence and hinder stable hybrid formation (Rieseberg 2001; Nachman and Payseur 2011; Stevison *et al.* 2017).

Additionally, altered recombination patterns in some species can cause fertility defects in hybrids which can prevent interbreeding. For example, variation between alleles of the *PRDM9* gene, which is responsible for DSB patterning in mammals, is implicated in hybrid sterility in mice and is a mechanism which could contribute to speciation in several mammalian species (Mihola *et al*. 2009; Davies *et al.* 2016; Alves *et al*. 2017).

1.16 Benefits of improving understanding of recombination

The processes of meiosis and meiotic recombination are vital in sexually reproducing species, not only for the production of haploid gametes to maintain a diploid genome after fertilisation, but also for the generation of genetic diversity and promotion of efficient natural selection and adaptation (Villeneuve and Hillers. 2001; Morrell *et al.* 2004; Roze and Barton 2006; Mercier *et al*. 2015). However, despite the importance of these processes in genome evolution and fertility, a lot is still unknown about the proteins involved and their activities and interactions (Mercier *et al.* 2015). Although there is substantial conservation of meiotic structures and core recombination proteins between eukaryotic species, the considerable variation observed in meiotic processes and recombination patterns suggests that there may also be extensive differences to explore, which may provide an interesting insight into the differential evolution of recombination and meiosis in different species (Zickler 2006; Mercier *et al.* 2015; Lawrence *et al*. 2017). Identification of modifiers responsible for variation in meiotic recombination rates could improve understanding of the recombination process and its regulation, in addition to finding potential interaction partners of known proteins and clarifying the molecular mechanisms involved in recombination.

Meiotic recombination contributes significantly to genetic diversity in sexually reproducing organisms, by shuffling allelic variants to generate new combinations. This prevents the accumulation of deleterious variants, known as Muller's Ratchet, thereby contributing to beneficial evolution (Wijnker *et al* 2013). Hill-Robertson interference dictates that linkage-disequilibrium in the genome slows the process of evolution via natural selection by linking variants that are under the influence of different selection pressures (Hill and Robertson 1966; Roze and Barton 2006). Recombination mitigates the effect of Hill-Robertson interference by reducing genetic-linkage of variants, thereby influencing rates of adaptation and responses to natural selection (Barton and Charlesworth 1998). Outcrossing in natural populations of Arabidopsis generates genetic variation, however recombination is required to integrate variants into new beneficial combinations, and therefore the rate of recombination in a population directly influences the rate of evolution (Hill and Robertson 1966; Barton and Charlesworth 1998). Natural modifiers of crossover frequency and distribution may therefore influence genetic adaptation to diverse environments and conditions. Identification of natural Arabidopsis modifiers

affecting recombination can provide insight into adaptation and natural selection in plant populations, which is a point of general interest.

There are additional benefits to identifying natural modifiers of recombination in plants, if they are confirmed to underlie variation in recombination under natural conditions, as they can provide information to aid understanding of the evolution of recombination in different species and populations (Koornneef *et al.* 2004). It could be interesting to compare modifiers of recombination that are identified in plant populations to modifiers identified in fungi and animals, to determine whether the mechanisms of modulating recombination are conserved between groups, or if they have diverged due to the differences in lifestyle and, consequently, selection pressures. As plants are sessile, they are likely to be subject to different selection pressures compared to animal populations that can move away from adverse conditions (Suzuki *et al*. 2014; Kawakatsu *et al*. 2016). It is possible that plants may therefore require faster adaptation to changing environmental conditions as they cannot be avoided, which could act as a selection pressure on recombination, resulting in changes to recombination in plants that are not found in animal species (Hadany *et al.* 2008; Webster and Hurst 2012; Kawakatsu *et al.* 2016). Increasing the understanding of modifiers of recombination and how they may differ between populations could aid simulations of adaptation in natural populations, as rates of adaptation are believed to vary as a function of recombination variation (Morrell *et al*. 2004; Marais *et al*. 2004).

The conservation of certain features of the recombination process also raises the possibility that some of the information learned from plants could be applicable to other species, including humans. Improving the understanding of recombination in plants could therefore provide information that may have future relevance in humans. However, this is a speculative potential benefit of recombination research in plants, and would largely depend on the modifier protein identified, as many of the modifiers previously identified have not been conserved between plants and mammals (e.g. PRDM9), although some modifiers have (e.g. HEI10) (Chelysheva *et al.* 2012; Kong *et al.* 2013; Mercier *et al*. 2015). If modifiers identified in plants were demonstrated to also affect recombination in humans, it is possible that they could provide a function in human infertility research, as there is evidence of a correlation between recombination rate and reproductive success in women (Kong *et al.* 2004; Fledel-Alon *et al.* 2011). As increased recombination is believed to contribute to fertility in older women, factors responsible for variation in recombination rates in humans could be used as predictors of declining fertility with age for fertility screening procedures (Kong *et al.* 2004; Coop *et al*. 2008). While several modifiers are already known in human populations, it is possible that identification of modifiers in other systems could suggest further potential avenues of investigation in humans which could be of use for medical purposes.

Identification and characterisation of modifiers of meiotic recombination frequency in natural Arabidopsis accessions could also offer more applied benefits in plants. Genetic re-assortment of variants is exploited during crop breeding to create favourable allele combinations and introgress individual traits into optimised lines (Bauer *et al*. 2013). Identification of modifiers upregulating recombination could be utilised to reduce linkage drag of unwanted variants and reduce the number of generations required to generate the desired lines (Bauer *et al*. 2013). Identification of global modifiers could be used to upregulate recombination across the entire genome, and characterisation of region-specific modifiers could be used to upregulate crossovers in specific areas of the genome, for example in the gene-rich euchromatic arms, without compromising the stability of centromeric repeat arrays. There is some evidence that selection acts in natural populations to specifically reduce recombination in areas where it would not be advantageous, for example in regions of the genome containing co-adapted loci, therefore the use of region-specific modifiers to manipulate recombination patterns could mimic natural strategies and prevent the breaking up of beneficial variant combinations in crops (Lenormand and Otto 2000; Stapley *et al.* 2017). Alteration of recombination could also be used in crop breeding to stabilise neo-polyploid lines, as selection is believed to cause a reduction and redistribution of COs in fully evolved autotetraploids to minimise multivalent formation and promote stable chromosome segregation, but high recombination rates and multivalents are frequently observed in newly-formed autotetraploids (Yant *et al*. 2013; Wu *et al.* 2013; Yant and Bomblies 2015; Bomblies *et al*. 2016).

Natural modifiers found in Arabidopsis could be used directly to transgenically modify crop species, which may affect their recombination (Crismani *et al*. 2013). Alternatively, high levels of conservation in the recombination process among plant species means that this information could also be used to identify a species-specific homolog of the modifier in the crop species, potentially in a wild relative. If an allele was identified that associated with higher recombination, this could then be bred into the crop population to create a non-transgenic hyper-recombinogenic line that would not be subject to the same legal regulations as a genetically modified crop (Jones 2015; Huang *et al*. 2016). However, it must be noted that the effect of these modifiers is dependent on genome architecture and population structure, therefore modifiers identified in the relatively low-complexity Arabidopsis genome may not have the same effect in repeat-rich crop species that frequently exhibit polyploidy.

Upregulation of recombination also offers benefits within laboratory plant populations used in genetic research. Populations generated to identify genetic factors underlying a phenotype via linkage or association-based mapping are often limited by low levels of recombination leading to genetic-linkage of candidate variants (Bauer *et al*. 2013). Increases in recombination, or alteration of crossover distribution, could reduce linkage and improve the precision of genetic mapping.

1.17 Aims and objectives

The primary aim of this project was to identify natural *trans*-acting modifiers of meiotic recombination frequency that vary between *Arabidopsis thaliana* accessions. Analysis of recombination variation between accessions suggests that there are likely to be large-effect modifiers differing between populations, and these may correspond to genes whose influence over recombination is currently unknown (Ziolkowski *et al*. 2015). After modifiers were identified in mapping populations, mapping of the underlying loci and confirmation of their effects on recombination would provide more detailed information. While identification of a novel recombination gene is the desired outcome, finding a role in the modulation of recombination frequency for a gene already known to be involved in the recombination pathway, although perhaps in a mechanism that is not believed to influence crossover resolution, could also be of interest for improving understanding of the recombination process.

Identifying natural variants associated with differences in recombination in a gene known to have a role in recombination is particularly of interest in terms of the implications of these findings in the evolution of natural variation in recombination (Koornneef *et al.* 2004; Mitchell-Olds and Schmitt 2006). The identification of genes that are likely to be responsible for some proportion of natural variation in recombination in *A.thaliana* could provide some insight into the mechanisms that are used to alter recombination during evolution, for example whether these genes are regulatory or structural, or even part of the recombination machinery itself. While this information would primarily aid understanding of recombination, it would also aid understanding of factors influencing adaptation and selection efficiency in natural populations, as these are affected by changes in recombination, and would therefore be of general interest to many biologists (Stapley *et al.* 2017). Identification of these genes, and preliminary characterisation of their effect on recombination, could not only potentially improve understanding of parts of the recombination pathway that are still not fully understood, but could also provide potential targets for use in the experimental manipulation of recombination to facilitate crop improvement (Bauer *et al*. 2013).

Secondary aims for this project included characterising variation in recombination between accession-crosses across different intervals of the genome, to determine whether recombination patterns differed between accessions, and creating several mapping populations utilising different accessions and genome intervals to determine whether modifiers have a consistent effect across the genome and whether they are conserved between accessions. This characterisation of the variation in recombination, both in frequency and distribution, between different accessions serves several purposes beyond improving understanding of recombination. Firstly, it may serve a practical purpose in the experimental use of different *A.thaliana* accessions, as understanding the recombination

patterns of individual *A.thaliana* accessions and cross hybrids can inform us about genetic linkage patterns in those lines, which could be of interest for use of these lines in mapping experiments, or for calculating the probability of linkage between genes underlying specific phenotypes, which may affect experimental planning. More importantly, characterisation of recombination differences between natural accessions provides insight into the evolution of variation in recombination, potentially facilitating understanding of strategies used by plants in natural populations to adapt to changes in environmental or genomic conditions (Stapley *et al.* 2017; Dapper and Payseur 2017). For example, characterisation of recombination differences between neo-polyploids and fully evolved polyploids has established that evolved polyploid lines often show both a comparative reduction in CO number and a redistribution of events towards the ends of chromosomes (Yant *et al*. 2013; Wu *et al.* 2013; Bomblies *et al.* 2016). The characterisation of these recombination patterns determined a strategy used to stabilise chromosome segregation in the presence of homeologous chromosomes and improved the understanding of mechanisms used in the adaptation to polyploidy (Bomblies *et al.* 2016). Characterisation of patterns of recombination found in accessions that are under different environmental selection pressures could aid understanding of how recombination phenotypes may change under different conditions, which could provide an insight into mechanisms of adaptation (Stapley *et al.* 2017; Dapper and Payseur 2017).

Objectives:

- Measure recombination across several genomic intervals in *A.thaliana* accession-cross F_1 hybrids to confirm observations made by Ziolkowski *et al.* (2015) regarding variation between Col-0, Cvi-0 and Can-0 accessions and identify promising crosses to take forward into mapping populations (See Chapter 3).
- Create segregating mapping populations from accession-crosses to identify modifiers of recombination using QTL mapping (See Chapter 3) (Ziolkowski *et al*. 2017).
- Generate fine mapping populations from initial mapping lines through backcrossing into a common genetic background to reduce the effects of other segregating loci and increase mapping resolution to identify candidate genes for modifiers (See Chapter 4) (Alonso-Blanco and Koornneef 2000).
- Introgress QTLs individually into a common genetic background to characterise their separate effects on recombination (See Chapters 4 and 5).
- Confirm the effects of candidate genes on recombination through measurement of recombination in mutant lines containing a T-DNA insertion in the gene of interest or through transformation of accession alleles into a common genetic background and comparison of

recombination rates (See Chapters 4 and 5) (Alonso-Blanco and Koornneef 2000; Mitchell-Olds and Schmitt 2006).

1.18 Experimental considerations

The decision to use QTL-mapping to identify these unknown modifiers of recombination stems from examination of a multitude of other studies mapping genes underlying traits that vary between *A.thaliana* accessions (Koornneef *et al.* 2004). QTL-mapping in segregating accession cross populations had previously been used to identify modifiers of meiotic recombination (Esch *et al*. 2007) and had resulted in the successful identification of the HEI10 modifier (Ziolkowski *et al*. 2017). However, while HEI10 was found to be responsible for a significant proportion of the variation in recombination within a Col-0/L*er*-0 F_2 mapping population measuring recombination in the sub-telomere of chromosome 3, there was still additional variation within the population that could not be attributed to the segregation of *HEI10* alleles (Ziolkowski *et al*. 2017). Additionally, HEI10 cannot provide an explanation for some of the variation in recombination observed between different accession/Col-0 F_1 crosses, as some of these accessions share a *HEI10* allele with Col-0 and would therefore match the homozygous Col-0 recombination rate if HEI10 was the only modifier responsible for variation in recombination between accessions (Ziolkowski *et al*. 2015; Ziolkowski *et al*. 2017). Furthermore, while HEI10 affects recombination across the whole genome, it does not appear to be responsible for large changes in recombination around the centromeres (Ziolkowski *et al.* 2017; Serra *et al*. bioRxiv). As highly variable recombination rates were observed between different accession/Col-0 F_1 lines when recombination was measured across an interval spanning the centromere and pericentromeric regions of chromosome 3, this suggests that there may be additional significant modifiers of recombination that vary between *A.thaliana* accessions that have yet to be identified (Ziolkowski *et al*. 2015). Modifiers affecting recombination around the centromere may even be of particular interest to plant breeders as they could be manipulated to promote recombination in regions that are frequently crossover-suppressed in crops, allowing shuffling of variants that are usually found in haplotype blocks that persist for many generations (Cheng *et al*. 2001; Tenaillon *et al.* 2002; Nachman 2002; Yao and Schnable 2005; Saintenac *et al.* 2009; Rodgers-Melnick *et al*. 2015). Therefore, I decided to generate segregating mapping populations from crosses between accessions shown to vary in recombination frequency (Ziolkowski *et al*. 2015), utilising multiple accession-crosses to maximise the probability of identifying a significant modifier.

Analysis of segregating populations, created from parents that differ for the trait of interest, is a powerful method to identify and map unknown genetic modifiers of phenotypes (Koornneef *et al*. 2004). *Arabidopsis thaliana*, with its large range of genetically variable accessions (The 1001 Genomes

Consortium 2016), and considerable available sequence and expression information (Alonso-Blanco *et al*. 2016; Krishnakumar *et al*. 2015), offers an excellent system for genetic mapping. A number of modifiers of Arabidopsis meiotic recombination frequency have been discovered in recent years by genetic screens, including *FANCM, FIGL1, RECQ4A* and *RECQ4B* (Crismani *et al*. 2012; Girard *et al*. 2015; Seguela-Arnaud *et al*. 2015). However, while mutagenic screens have proven to be a valuable approach, natural variation as a complementary tool for mapping recombination modifiers in Arabidopsis remains relatively underexploited.

Screening natural variation for modifiers of interest offers some advantages over chemical mutagenesis screens (Alonso-Blanco and Koornneef 2000). Mutagenic screens are highly efficient in achieving saturation of genes and avoid the chromosomal re-arrangements sometimes caused by T-DNA insertions, but they require an initial mutagenesis step and then laborious screening of a large population of descendent plants (Kim *et al*. 2006). While mutagenic screens have proven capable of identifying genes involved in complex traits (Crismani *et al*. 2012), QTL mapping in segregating natural accession cross populations can allow rapid identification of multiple modifiers in a single mapping population, often utilising only a few hundred plants compared to thousands required for gene saturation in a mutant screen (Alonso-Blanco and Koornneef 2000). Additionally, the frequency of non-synonymous mutations achieved in Arabidopsis, by ethyl methanesulfonate (EMS) mutagenesis for example, is between 5 and 65%, meaning that often mutations do not disrupt gene function, thereby reducing the efficacy of the screen (McCallum *et al*. 2000). Even if mutagenesis does cause a disruption of gene function, this may be compensated for by another gene if there is functional redundancy or if the gene is part of a large multigene family, thereby preventing detection of a mutant phenotype (Bouché and Bouchez 2001; Mitchell-Olds and Schmitt 2006). Moreover, mutant screens can be limited by the genetic background they are performed in - if the wild-type line contains a weak or functionally null allele for the gene of interest, the mutant phenotype may not be detected (Alonso-Blanco and Koornneef 2000). Similarly, epistatic effects relating to functional gene redundancy affecting the phenotype may only be apparent in specific genetic backgrounds, placing considerable significance on the choice of background for the screen (Alonso-Blanco and Koornneef 2000).

While weak alleles and epistatic effects may also be present in natural populations, analysis of multiple different accessions and genetic backgrounds increases the probability of detecting alternative alleles. This can be used to detect novel genes affecting the phenotype, or new functional alleles of genes that were identified by mutant screens, thereby providing additional information about the molecular mechanism underlying the gene function (Alonso-Blanco and Koornneef 2000). New alleles that are identified by natural variation screens often have an effect on the phenotype without disrupting gene function, unlike many alleles identified by mutagenic screens (McCallum *et al.* 2000). Functional alleles

that have evolved in natural populations are less likely to have deleterious off-target effects than mutagenic gene disruptions, meaning that they are frequently better suited to investigations of gene function (Bouché and Bouchez 2001; Mitchell-Olds and Schmitt 2006). Additionally, identification of functional alleles allows an analysis of how altered gene expression or protein activity may affect the phenotype, which can aid understanding of how the protein contributes to molecular processes related to the phenotype (Alonso-Blanco and Koornneef 2000; Bouché and Bouchez 2001; Page and Grossniklaus 2002).

However, mapping using natural variation does have some disadvantages, primarily that mapping resolution is often much lower, and with additional polymorphisms between parental accessions in the region of interest, it can be difficult to determine the causal mutation. Furthermore, segregation of multiple loci and possible interactions with environmental factors can make identification of genetic modifiers in natural populations more difficult. However, advances in statistical and computational methods for QTL mapping and data integration have improved natural variation mapping efforts (Alonso-Blanco and Koornneef 2000), and the advantages over mutagenesis make it a promising alternative method for increasing understanding of meiotic recombination modifiers.

Natural variation screens in Arabidopsis exploit information that is directly accessible without the need for an initial mutagenesis screen. Information gained about wild-type populations can therefore be linked to additional evidence about accessions in the literature, such as accession phenotypes, expression patterns, life history and habitat, providing a wider context for results (Koornneef *et al*. 2004; 1001 Genomes Consortium 2016). As substantial variation in the recombination phenotype has already been observed between accessions (Lopez *et al*. 2012; Ziolkowski *et al*. 2015), large-effect modifiers could be assumed to be present and potentially mappable, removing much of the uncertainty present in screening. The use of natural variation also provides additional information not obtainable from mutagenic screening systems, such as providing insight into the evolution and adaptation of the modifiers, as natural variants may be biologically relevant to the phenotype in extant natural populations (Mitchell-Olds and Schmitt 2006).

Mapping in Arabidopsis accession cross populations has confirmed the presence of genetic modifiers of recombination (Esch *et al*. 2007), and attempts have been made to map these natural variants in segregating populations using QTL mapping. Once two accessions suspected to differ for recombination frequency have been identified, they can be crossed to generate an F_1 and then self-fertilised to generate an F_2 QTL mapping population where each individual is phenotyped for recombination and genotyped across the genome. The recombination phenotype is then associated with sections of the genome where the allele distribution in the population is linked to the variation

in the phenotype. Genetic linkage means that these regions contain multiple linked genes, although subsequent recombination events in back-cross fine-mapping populations can reduce linkage and narrow the search area (Mitchell-Olds and Schmitt. 2006). However, extremely large mapping populations are required to accumulate sufficient crossovers to provide a precise estimate of the QTL location (Salome *et al*. 2012).

While evidence of the presence of genetic modifiers can be found relatively easily through QTL mapping in accession-cross populations, confirmation of the gene or sequence feature responsible for the phenotype can be challenging (Esch *et al*. 2007). The presence of multiple linked polymorphisms that vary between the two accessions can prevent easy identification of the causative difference. To confirm the effect, analysis of mutants from insertional mutagenesis line collections, and reciprocal accession allele transformations can be performed for candidate genes. Analysis of allele distributions among other accessions using sequence data can also aid confirmation and provide information as to the relevance of the variant in different populations and environments. While multiple recombination QTL have been identified in Arabidopsis accession cross populations, including a large effect locus on chromosome 1 of a L*er*xCvi cross (Log of Odds [LOD] 4.95), only one successful gene identification has been made – that of the *HEI10* gene (Esch *et al*. 2007; Ziolkowski *et al*. 2017).

To perform QTL mapping in the large populations required for effective locus resolution, a high-throughput method of phenotyping is required. As recombination is a molecular phenotype and therefore not directly observable, and cytological observation is time consuming and can only analyse a small number of meiotic products, alternative methods have been explored. Esch *et al* (2007) utilised recombination breakpoints found between genotyping markers as a measure of crossover frequency in recombinant-inbred individuals, but this only allows measurement of recombination in a small number of meioses. Recently, use of fluorescent reporter systems, based around the segregation of fluorescence in the products of meiosis, allows high-throughput measurement of recombination across defined intervals of the genome in Arabidopsis (Melamed-Bessudo *et al*. 2005; Berchowitz and Copenhaver. 2008, Yelina *et al*. 2013).

Fluorescent transgenes expressed in either pollen or seed are inserted into the genome, with two different colours defining either end of an interval (Figure 10). Lines have been generated containing T-DNAs expressing eCFP, dsRed or eYFP fluorescent proteins in mature pollen from the post-meiotic *LAT52* promoter and dsRED and GFP fluorescent proteins under the control of the seed specific *NapA* promoter, allowing measurement of marker segregation in both seed and pollen (Melamed-Bessudo *et al*. 2005; Berchowitz and Copenhaver 2008; Yelina *et al*. 2013). When these transgenes are present in an individual in a hemizygous *cis*-configuration, recombination between them produces pollen or

seed with differing fluorescence patterns. The relative levels of recombinant and parental products, defined by fluorescence combination, are used to perform a calculation of the recombination frequency between these markers. This method utilises information from thousands of meioses, providing a more robust measurement of recombination. One disadvantage is that the use of the Fluorescent-Transgenic Line (FTL) system requires fluorescent transgenes to be maintained in a hemizygous state, preventing the use of homozygous accessions (with the exception of the Col-0 accession which is the genetic background of the FTL system), Recombinant Inbred Lines (RILs) or multiparent mapping populations such as Multiparent Advanced Generation Inter-Cross (MAGIC) (Kover *et al*. 2009), though it is effective for F_2 mapping populations.

The FTL method, in combination with genotyping across the genome using Simple Sequence Length Polymorphism (SSLP) and Cleaved Amplified Polymorphic Sequence (CAPS) markers, permits efficient screening of large populations, and subsequent QTL mapping. While low-coverage sequencing could be used for genotyping F_2 mapping populations, this method becomes less efficient after progression to fine-mapping, where the number of samples can be much larger. Unlike in an F_2 accession-cross mapping population where the whole genome is segregating, not all individuals in a back-crossed fine-mapping population are useful. Individuals without a recombination event within the QTL region of interest to break-up haplotype blocks do not provide information. As the region of interest decreases in size through further fine-mapping, the proportion of the population containing a beneficial recombination event decreases. Production of sequencing libraries for each individual in the population, regardless of mapping utility, would be inefficient and expensive. Use of accession-specific PCR-based markers flanking the region of interest allows identification of individuals with a valuable recombination event for further genotyping and measurement of recombination phenotype. For ease and consistency, this system can also be used in the F_2 mapping population, where additional genotyping markers can subsequently be added to improve mapping resolution in areas of the genome containing putative QTL.

Identification of natural modifiers of recombination through QTL mapping would provide an opportunity for future work characterising their effects on recombination, and how they integrate into the complex process of recombination. It could also provide additional information about the evolution of recombination modifiers in natural populations, and how this may relate to adaptation and natural selection. This project provides information about modifiers of an essential molecular process in sexually-reproducing organisms, and ultimately identifies potential candidates for manipulation in crop species.

Chapter 2-Materials and Methods

2.1 Plant material

The *Arabidopsis thaliana* accession lines used for analysis of natural variation were Columbia (Col-0), Landsberg *erecta* (L*er*-0), Canary Islands (Can-0) and Cape Verde Islands (Cvi-0), which were obtained from the Nottingham Arabidopsis Stock Centre (NASC) and were provided by Dr Piotr Ziolkowski (Ziolkowski *et al*. 2015).

Fluorescent transgenic lines (FTLs) with fluorescent proteins transcribed from the *LAT52* pollen-specific promoter in the Col-0 background were used for measurement of recombination in the *CEN3* and *I2f* intervals (Francis *et al*. 2007). *CEN3* with fluorescent transgene insertions on chromosome 3 (generated from FTL3332 and FTL2536 insertion lines, see Yelina *et al*. 2013) and *I2f* with fluorescence-gene insertions on chromosome 2 (generated from FTL800 and FTL3411 insertion lines, see Lambing *et al*. 2015), were used to measure recombination in intervals on chromosomes 3 and 2 respectively. *420* FTL seed, with fluorescent transgenes under the control of the seed-specific *NapA* promoter on chromosome 3, was obtained from Prof. Avraham Levy (The Weizmann Institute, Israel) and provided by Dr Piotr Ziolkowski. Additional seed-based traffic lines using *eGFP* and *dsRED* genes fused to the *NapA* promoter in the Col-0 background were used to measure recombination in the *5.11* (generated from CR1082 and CG445 insertion lines), *3.9* (generated from CG17 and CR55 insertion lines) and *1.19* (generated from CR871 and CG294 insertion lines) intervals on chromosomes 5, 2 and 1 respectively (Wu *et al*. 2015). Table 1 contains additional information on fluorescent lines used to measure crossovers.

Arabidopsis thaliana T-DNA insertion lines in the Col-0 background were used for analysis of candidate genes. Salk lines (Alonso *et al*. 2003), Syngenta Arabidopsis Insertion Library (SAIL) lines (Sessions *et al*. 2002) and GABI-Kolner Arabidopsis T-DNA (GABI-Kat) lines (Kleinboelting *et al*. 2012) were obtained from NASC. See Tables 2, 3 and 4 for full list of lines and stock IDs.

The *HEI10* overexpressor line C2 was generated by agrobacterium-mediated transformation with additional transgenic *HEI10* copies in the Col-0 background and provided by Drs Piotr Ziolkowski and Charles Underwood (Ziolkowski *et al*. 2017).

2.2 Plant propagation

Unless otherwise indicated, all plants were propagated by self-fertilisation. Additional lines were generated from the above strains through cross-fertilisation, performed by manual emasculation and fertilisation.

2.3 Plant growth conditions

Plants were cultivated on commercial F2 compost with fine vermiculite and grown in controlled environment chambers at 20˚C, 60% humidity in a long day photoperiod (16 hours light), light intensity 150 µmols following 4 days stratification at 4˚C (as per Yelina *et al*. 2012). Plants were grown at a density of six plants per 9 cm^2 pot and were evenly spaced.

T-DNA insertion line seeds were sterilised using a seed cleaning solution (0.05% Triton x-100 v/v), followed by 100% ethanol, and sown onto selective media plates containing 0.5 strength Murashige and Skoog media with 0.8% Agar (w/v), 0.5% sucrose (w/v) and the relevant antibiotic. Kanamycin was used at a concentration of 50 µg/ml, Basta/PPT at 10µg/ml. After 4 days stratification in the dark at 4˚C, plates were moved to growth cabinets set to 20˚C, 60% humidity in a long day photoperiod (16 hours light) with a light intensity 150 µmols. Seedlings were then transferred to soil in long day photoperiod growth chambers.

2.4 DNA extraction

Genomic DNA for plant genotyping was extracted from leaf tissue using the protocol outlined in Edwards *et al*. (1991), adapted for 96-well plate extractions: Plant tissue was ground in 200µl extraction buffer (200mM Tris pH7.5, 250mM NaCl, 25mM EDTA) without SDS detergent using a QIAGEN Tissue Lyser with 3mm borosilicate beads, before addition of 200µl of buffer containing SDS (1% (v/v)) to minimise manual grinding effort. The isopropanol precipitation step was extended from 2 to 20 minutes at room temperature to maximise yield. Centrifugation steps were performed at 3800rpm, with times extended accordingly, due to the capacity of the larger plate-spinning centrifuge and rotor.

DNA for PCR amplification and sequencing was extracted from leaf tissue using the CTAB (cetyltrimethyl ammonium bromide) method (adapted from Clarke 2002). Frozen tissue samples were ground using the Qiagen Tissue Lyser with borosilicate beads, before incubation at 65°C in CTAB buffer (150mM sorbitol, 220mM Tris pH8, 22mM EDTA, 800mM NaCl, 0.1% (v/v) sarcosine, 0.8% (w/v) CTAB) for 30 minutes with gentle agitation. An equal volume of pure chloroform was added before vortexing and centrifugation at room temperature. Liquid from the upper layer was added to an equal volume of isopropanol before vortexing and incubation at room temperature for 5 minutes. Samples were centrifuged at 13000rpm for 15 minutes at 4°C, the supernatant was removed and the pellet was washed with 70% EtOH before air drying. The dry pellet was resuspended in water containing RNase A at a concentration of 100ng/ml, before incubation at 37°C for 30 minutes. DNA was then precipitated by addition of 1/10 volume of 3M NaAc, followed by 2.5 volumes of 100% EtOH. Samples

were frozen at -20°C for 30 minutes before centrifugation at 13000rpm for 15 minutes at 4°C. Supernatant was discarded and the pellet was washed with 70% EtOH before air drying and resuspension in water.

2.5 Genotyping

Genotyping of Arabidopsis populations was performed by PCR-based methods, using a mixture of Simple Sequence Length Polymorphism (SSLP) and Cleaved Amplified Polymorphic Sequence (CAPS) markers (Supplementary table 1). Markers were designed using polymorphism data from the Salk 1,001 Genomes Project browser (Alonso-Blanco *et al*. 2016), to create SSLP and CAPS markers between 35 bp and 300 bp in length. PCR products were separated by gel electrophoresis and visualized by ethidium bromide-staining under UV light (3% agarose gel in 1xTBE, 1/10000 EtBr, 240V) in order to genotype individuals.

Genotyping of T-DNA insertion lines was performed using the primers in Supplementary Table 2, in conjunction with the Salk LBb1.3 T-DNA left border primer for Salk lines (recommended by the Salk T-DNA primer design tool (http://signal.salk.edu/tdnaprimers.2.html)) , GKo8474 left border primer for GABI-Kat lines (https://www.gabi-kat.de/duplofaq/confirmation-strategy.html) or SAIL LB1 left border primer for SAIL lines (Salk T-DNA primer design tool). Primers were designed using the Salk T-DNA primer design tool. Visualisation of PCR products was performed as above using gel electrophoresis and ethidium bromide-staining.

2.6 FTL pollen fluorescence measurement and calculation of crossover frequency

Crossover frequency was measured using fluorescent reporter genes under the control of the pollen-specific *LAT52* promoter (Francis *et al*. 2007). Red and yellow/green fluorescent transgenes were inserted either side of the interval to be measured (Francis *et al*. 2007; Berchowitz and Copenhaver 2008). Individuals with the transgenes in a hemizygous state in *cis*-configuration produced pollen segregating for different fluorescent colours. The numbers of recombinant (single colour) and parental (double or non-colour) pollen are used to calculate the frequency of crossover between the transgenes (Figure 10). As the *qrt1* mutation, which prevents separation of the pollen tetrad into individual grains, is also present in the populations derived from pollen-based FTLs (Francis *et al*. 2007), pollen must also be phenotyped under a microscope to ensure pollen is from a *QRT1* background, which is necessary for flow-cytometry analysis (Yelina *et al*. 2013).

Inflorescences were collected in sterile 50ml polypropylene tubes from individual plants with the fluorescent FTL transgenes in a hemizygous cis-configuration (RY/--) and a wild-type *QRT1* phenotype. Pollen was extracted from inflorescences by vigorous shaking in pollen sorting buffer (PSB)(10 mM

$CaCl_2$, 1 mM KCl, 2 mM MES, 5% sucrose (w/v), 0.01% Triton X-100 (v/v), pH 6.5). The pollen suspension was filtered through a 40µm cell strainer into a fresh tube and centrifuged at 450g for 5 minutes. The supernatant was then removed and the pollen pellet washed with 10ml of PSB without Triton. The suspension was centrifuged again at 450g for 5 minutes, and the supernatant was discarded. The pollen pellet was resuspended in 500µl of PSB without Triton and used for flow cytometry as described by Ziolkowski *et al*. (2015). For two-colour flow cytometry, data was collected using a BD Accuri C6 flow cytometer with a 488nm laser and 530/30nm and 570/20nm band-pass filters and the fast flow setting (66 µl/min). Events were separated by side (90 degrees) and forward (0 degrees) scatter which was used as a proxy for pollen size, and gated to enrich for hydrated pollen (Gate R_5) and to remove debris (Yelina *et al*. 2013). Pollen events were then gated by fluorescence emission into red (Q2UL), yellow (Q2LR), double colour (Q2UR) and non-colour (Q2LL) categories (Figure 10). Calculation of crossover frequency between reporters was performed by the equation cM= 100 x (2 x Q2LR)/(R_5 - (Q2LL – Q2UR)). This is an estimation of the recombinant pollen as a proportion of the total pollen. Double colour and non-colour pollen classes should be reciprocal and therefore equal in size, so the difference in counts between gates Q2LL and Q2UR can be accounted for by dead pollen or debris of a similar size. These events are manually subtracted from the total R_5 count to give an adjusted count of total pollen analysed. Similarly, gates Q2UL and Q2LR should have equal counts, as both represent reciprocal products of recombination between the reporters, and should be summed to calculate the number of recombinant pollen. However, a skew into the Q2UL gate is often observed causing the count to exceed Q2LR. This is believed to be attributable to an artefact caused by non-hydrated pollen in the sample having altered fluorescence and so, alternatively the more reliable Q2LR count is doubled to give the reciprocal recombinant events in the calculation (Yelina *et al*. 2013, Figure 10).

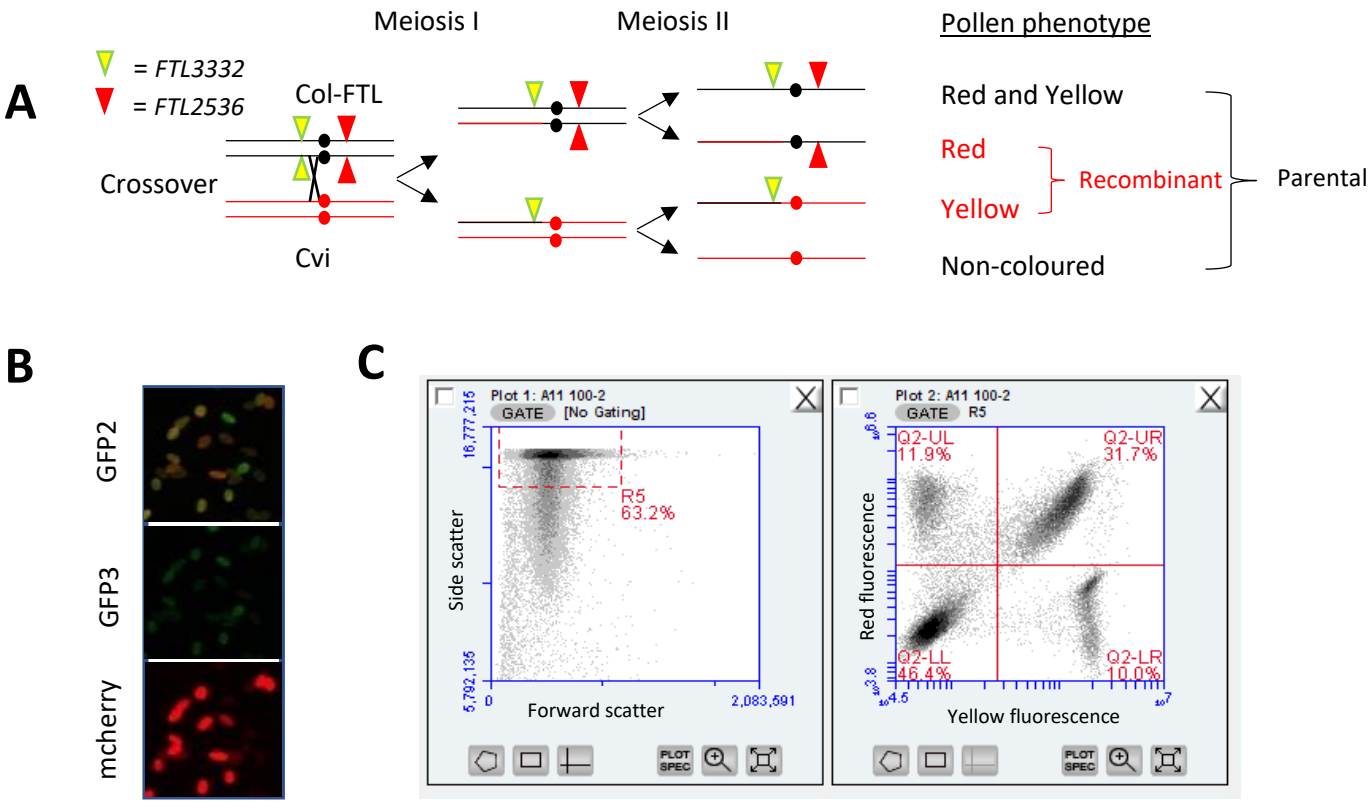
Three-colour flow cytometry of the *I3bc* interval was performed on a Beckman Coulter CyAn ADP Analyser, with 405nm and 488nm lasers and 530/40nm, 575/25nm and 450/50nm band-pass filters. Samples were prepared and run as above, with gating performed as described by Ziolkowski *et al*. (2015)(Figure 11). Gating and data analysis was performed using FlowJo single-cell flow cytometry analysis software (FlowJo v10.0.7, copyright Tree Star, Inc. Available at: https://www.flowjo.com/solutions/flowjo/downloads). Eight fluorescence classes were identified: Non-colour (NC), Triple colour (BYR), Yellow (Y), Blue and Red (B_R), Yellow and Red (YR), Blue (B), Blue and Yellow (BY), and Red (R). Total pollen is the sum of these eight classes (T). *I3b* genetic distance was calculated cM= (Y+B_R+YR+B)/T. *I3c* genetic distance was calculated cM= (Y+B_R+BY+R)/T. Crossover interference was then calculated from the difference between observed and expected levels of double-crossover events (DCO) given the recombination frequency of the intervals. Observed DCO = (Y+B_R), Expected DCO = (*I3b* cM/100)*(*I3c* cM/100)*T, Coefficient of Coincidence (CoC) =

Observed DCO/Expected DCO, Interference = 1-CoC (Ziolkowski *et al*. 2015). To compare interference estimates for two categories of plants, a two sample t-Test assuming unequal variances was used to test for significant differences.

2.7 FTL seed fluorescence scoring and calculation of crossover frequency

Recombination frequency was measured by analysing the segregation of fluorescent reporter transgenes under the control of the seed-specific *NapA* promoter in individuals hemizygous for the reporter genes (Melamed-Bessudo *et al*. 2005; Ziolkowski *et al*. 2015). Dried seed was collected from individual plants (RG/--) and filtered using a metal sieve to remove plant debris, before being imaged in a single-layer spread, achieved by gently pressing down with a microscope slide, under a dissecting epifluorescence microscope (Leica M165 FC, Leica Microsystems; Nikon Eclipse E1000 camera fitted with mCherry, GFP and CFP fluorescence filters). Three images were acquired per sample, using a charge coupled device camera; (i) brightfield, (ii) UV through a GFP3 filter and (iii) UV through an mCherry filter. These images were then analysed using an adapted CellProfiler image analysis software pipeline (Carpenter *et al*. 2006; Ziolkowski *et al*. 2015, Figure 10) that identifies individual seed objects and assigns an RFP and GFP fluorescence intensity to each object. As seed are diploid, they can be non-colour, single-copy or double copy for each fluorescent transgene, resulting in nine different fluorescence phenotypes in self-fertilised seed from a hemizygous plant (Ziolkowski *et al*. 2015). While individual classes can usually be distinguished with the software, single and double fluorescence categories often overlap. Therefore, fluorescence vs non-fluorescence was used for recombination measurement as previously reported (Ziolkowski *et al*. 2015). Histograms of seed fluorescence were used to manually set fluorescence intensity thresholds between fluorescent and non-fluorescent seeds for each sample (Figure 10). Once a fluorescence value had been assigned to each seed, genetic distance between the reporter genes was calculated using the equation $cM = 100*(1-(1-2(G+R)/T)^{1/2})$, where G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al.* 2015).

To compare categories of plants for statistically significant differences in FTL-based recombination estimates, the aggregate number of recombinant and non-recombinant pollen/seed were used to construct 2x2 contingency tables and perform a two-tailed χ^2 test to test for significant differences (Zibran 2015).



Fluorescent seed images and graphs demonstrating seed fluorescence intensity removed for copyright reasons.

Figure 10. Measurement of recombination using fluorescent-transgenic lines (FTL). D,E and F adapted from Ziolkowski *et al*. 2015. A. Fluorescent marker segregation as seen in a Cvi-0/*CEN3* F_1 hybrid undergoing crossover. **B**. Images of pollen from an individual hemizygous for RFP and eYFP transgenes, taken on a dissecting epifluorescence microscope under GFP2/GFP3/mcherry filters. **C**. Flow cytometry analysis of fluorescent pollen in a Cvi-0/*CEN3* F_1 individual. Total events separated by side (90 degrees) and forward (0 degrees) scatter, used as an approximation of size. Gate R5 contains pollen. Gate R5 events are then separated by red and yellow fluorescence emission into non-fluorescent (Gate Q2LL), red (Q2UL), yellow (Q2LR) and double fluorescent pollen (Q2UR). The equation cM= 100 x (2 x Q2LR)/(R_5 - (Q2LL – Q2UR)) is used to calculate the genetic distance (cM) of the interval (Yelina *et al*. 2013). **D**. Images of seed from an individual hemizygous for *RFP* and *GFP* transgenes, taken on a dissecting epifluorescence microscope under GFP2/GFP3/mcherry filters. **E**. CellProfiler output showing histograms of seed fluorescence intensities, with coloured and non-coloured seed divided by vertical dotted lines. **F**. Plot of seed red vs green fluorescence intensities demonstrating fluorescence categories. Non-coloured and coloured seed separated by dotted lines. The formula used for calculation of genetic distance in seed based intervals is cM = 100*(1-(1-2(Green+Red)/Total)$^{1/2}$).

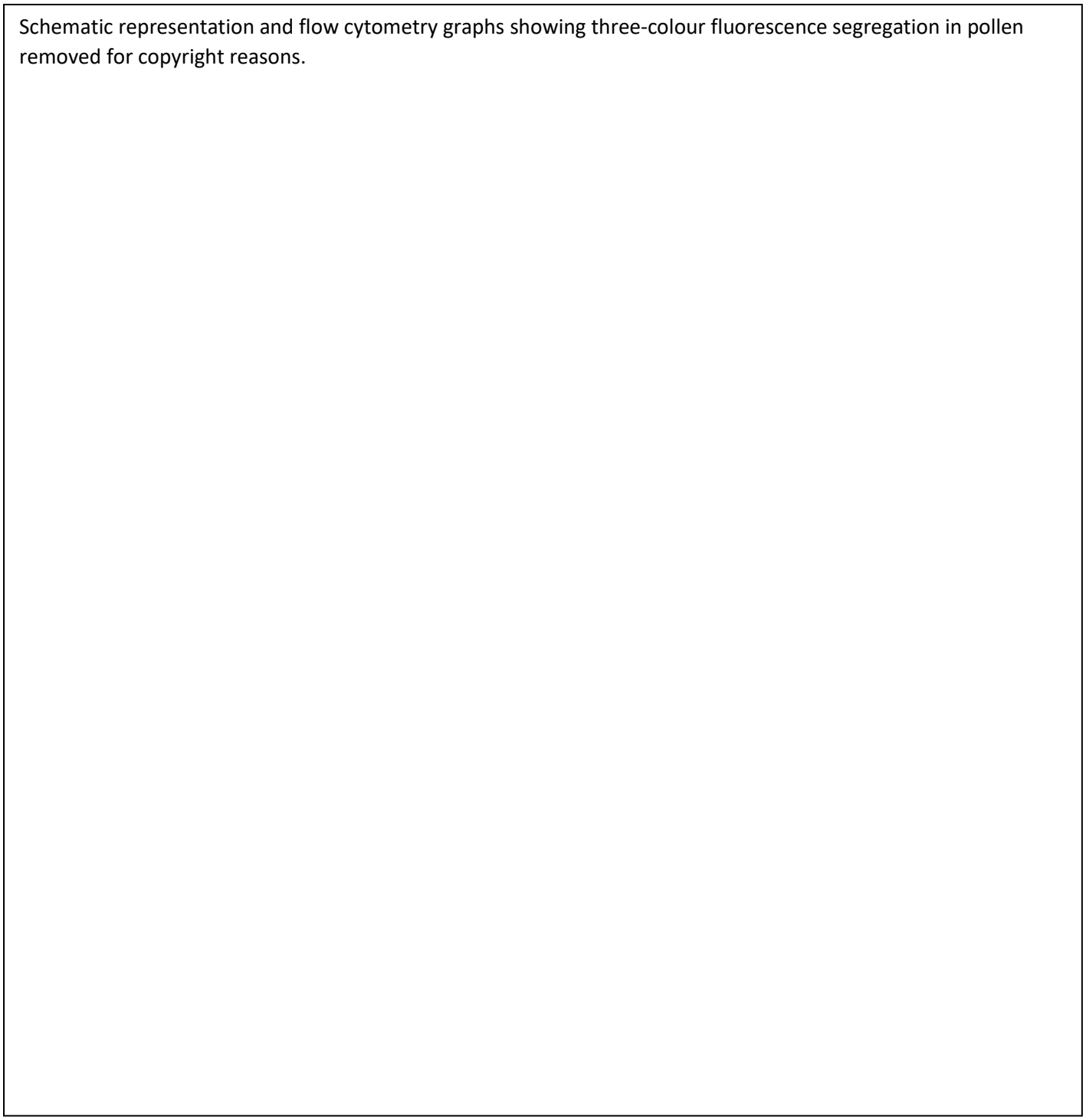


Figure 11. Analysis of fluorescence segregation in three-colour pollen FTL flow cytometry. Adapted from Yelina *et al*. (2013). A. Schematic showing segregation of linked hemizygous fluorescence transgenes encoding eYFP, dsRed and eCFP through two meiotic divisions indicated by arrows after a single crossover event. **B.** Segregation after a double-crossover event. **C-K**. Flow cytometry acquisition plots for three-colour pollen analysis. **C.** Gating for single pollen events based on an approximation of size. **D**. Separation of hydrated pollen. **E**. Identification of pollen events. **F-K**. Pollen separation along fluorescence axes demonstrating fluorescence classes. Pollen fluorescence count data collected on a CyAn ADP analyser (Beckman Coulter) and analysed using FlowJo software.

2.8 Quantitative Trait Locus (QTL) mapping

Individuals from accession/FTL cross mapping populations were phenotyped for crossover frequency using seed or pollen fluorescence scoring and genotyped across the genome with SSLP and CAPS markers (Supplementary Table 1). This information was combined and analysed using the R statistical package rQTL (Broman *et al.* 2003, Arends *et al.* 2010), to perform one- and two-dimensional QTL mapping, using Haley-Knott regression (Ziolkowski *et al*. 2017; Broman *et al.* 2003, Arends *et al.* 2010; Zeng 1993; Haley and Knott 1992). The *fitqtl* function was used to fit multiple-QTL models and identify QTL:QTL interactions (Broman and Sen 2009).

1,000 permutations were used for each mapping population to calculate genome-wide log of odds (LOD) score significance thresholds (Ungerer *et al*. 2003; Esch *et al.* 2007; Broman and Sen 2009). Observed LOD scores were compared to the distribution of the genome-wide maximum LOD score, under the assumption that there were no QTL. The 95th percentile of this distribution was used as a genome-wide LOD threshold. For individual QTL peaks, the p-value of significance was calculated as a genome-scan-adjusted p-value, which is the chance under the no-QTL null hypothesis of obtaining a LOD score that large or larger in the genome (Broman *et al*. 2009).

2.9 *HEI10* cloning

HEI10 was amplified from CTAB-extracted genomic DNA using primers *HEI10-XbaI* and *HEI10-BamHI* (Supplementary Table 3 (Ziolkowski *et al*. 2017)). Amplification products were digested using *Xba*I and *Bam*HI restriction enzymes, and ligated into the pGREEN0029 binary vector (Hellens *et al*. 2000). Primers from Dr Ziolkowski were used to match the size of the amplification product from Ziolkowski *et al*. (2017), to include any potential *HEI10* regulatory elements. Polymorphisms were confirmed by Sanger sequencing PCR products (see Supplementary Table 4 for primers).

2.10 *HEI10* transformation

Vectors containing *HEI10* transgenic constructs were transformed into the *Agrobacterium tumefaciens* strain GV3101 by electroporation, and then transformed into Col-0/*420* plants containing the fluorescent transgenes in a hemizygous *cis*-configuration via the floral dip method (Zhang *et al*. 2006). Primary inflorescences were cut back six days before dipping to provide multiple secondary inflorescences for transformation. After dipping, plants were kept in the dark for 24 hours, in sealed bags to maintain high humidity, before being returned to long-day photoperiod conditioned growth chambers.

Chapter 3 – Identification of modifiers of meiotic crossover frequency in *Arabidopsis thaliana* accessions.

3.1 Summary

Arabidopsis thaliana displays a wide range of genetically distinct natural accessions showing huge variation in phenotypes (Koornneef *et al*. 2004). Variation in meiotic recombination frequency has been observed between accessions using a range of methods, including scoring meiotic chiasmata, genotyping marker segregation in progeny lines and analysis of FTL segregation in the products of meiosis (Barth *et al*. 2001; Esch *et al*. 2007; Lopez *et al*. 2012; Salome *et al*. 2012; Ziolkowski *et al*. 2015; Sanchez-Moran *et al*. 2002). However, identification of the underlying genetic modifiers involved in this variation has been limited. This project aimed to characterise variation in recombination between Arabidopsis accessions across multiple intervals in the genome and identify accessions that differ from the Col-0 reference to use for QTL mapping experiments to identify potentially novel modifiers of meiotic recombination. Analysis of multiple populations measuring recombination across different regions of the genome could provide insight into variation in modifiers regulating recombination in different sections of the chromosome. Additionally, comparison of different accession populations could clarify whether modifiers are consistent between different natural populations. In this chapter, I describe how evidence of modifier loci was identified using seed and pollen FTL segregation to measure meiotic recombination across defined intervals of the genome in crosses between different accessions. Analysis of accession-FTL crosses identified significant variation between the Col-0 accession and relict accessions from the island populations, Cvi-0 and Can-0, and allowed subsequent mapping of several genetic modifiers in segregating F_2 populations.

3.2 Introduction: Finding modifiers of meiotic recombination in *Arabidopsis thaliana*.

Although one crossover per chromosome is obligatory for correct chromosome segregation during meiosis (Youds and Boulton, 2011), additional crossovers are known to vary considerably between *Arabidopsis thaliana* accessions, both in number and distribution across the genome (Esch *et al*. 2007; Lopez *et al*. 2012; Salome *et al*. 2012; Ziolkowski *et al*. 2015). While some of this variation can be attributed to differences in factors that are known to influence recombination (see Chapter 1), a large proportion of the variation observed is caused by unknown variables. Efforts to identify natural genetic factors affecting recombination that vary between *Arabidopsis thaliana* accessions have revealed the presence of numerous modifiers (Esch *et al*. 2007; Ziolkowski *et al*. 2015), however many of the underlying genes remain unidentified, indicating a promising direction for future mapping experiments. The identification of the HEI10 E3-ubiquitin ligase, as a dosage-dependent factor

affecting recombination frequency, in an accession-cross QTL mapping population measuring recombination with the high-throughput FTL-system, demonstrates the potential yield of utilising bi-parental accession-cross mapping populations (Ziolkowski *et al*. 2017). With a wide range of potential accessions available to cross and generate mapping populations, this strategy is promising for future identification of genetic recombination modifiers. However, selection of the parental accessions determines the success of the mapping endeavour, and therefore careful identification of appropriate lines had to be performed before QTL mapping could begin (Alonso-Blanco and Koornneef 2000).

3.3 Identification of parental accessions for mapping populations using the FTL system

A number of fluorescent-transgenic lines (FTLs) in the Col-0 accession background have previously been exploited to measure recombination in a range of experiments (Melamed-Bessudo *et al*. 2005; Berchowitz and Copenhaver 2008, Yelina *et al.* 2012; Crismani *et al.* 2012; Ziolkowski *et al*. 2015). Crossing these FTLs to other Arabidopsis accessions provides a way to map natural recombination QTL (rQTL) in bi-parental populations, providing that the parental accessions have differing recombination phenotypes. Appropriate parental lines for mapping populations could be identified by a systematic analysis and comparison of recombination in different accession-cross hybrids using FTLs. Several studies have been performed comparing recombination between different *A.thaliana* accessions, one of which used several different FTLs to measure recombination in accession-cross hybrids – information from this study was used to select accessions, and Col-0 FTL lines, to be used as parents for QTL mapping populations in this project (Sanchez-Moran *et al*. 2002; Lopez *et al*. 2012; Ziolkowski *et al*. 2015).

Ziolkowski *et al.* (2015) crossed five FTLs covering different regions of the genome, including centromeric, sub-telomeric and interstitial regions, to 25 different accessions which originated from across four continents (Figure 12A). Recombination measurements were made in F_1 plants and compared to establish the level of variation between accessions (Ziolkowski *et al.* 2015, Figure 12C-F). Lower variation (<5cM range) was observed between the mean recombination estimates of accessions in the *I1b* and *I1fg* chromosome interstitial intervals on chromosome 1, but greater variation was observed between accessions in the *I2f* and *420* sub-telomeric intervals on chromosomes 2 and 3 respectively (16cM and 21cM ranges respectively), and in the *CEN3* centromeric interval on chromosome 3 (23cM range). This suggested that large *trans*-effect modifiers that differed between the accessions were acting on recombination in these intervals, although it is also likely that some of the variation can be attributed to variation in *cis*-acting factors, possibly including juxtaposition of heterozygous and homozygous regions of the genome that depend on the pattern of polymorphism between parental accessions (Ziolkowski *et al*. 2015). It also suggested that these modifiers may be

Figure 12. Measurement and variation of recombination in F_1 accession-FTL crosses, adapted from Ziolkowski *et al*.2015. A. Map showing the geographical origin of the Arabidopsis accessions studied, indicated by red points. **B**. Historical crossover frequency shown in red (cM/Mb), sequence diversity in blue (π) along the physical length of the *Arabidopsis thaliana* chromosomes (Mb) with means shown by dotted lines (Cao *et al*. 2011; Choi *et al*. 2013). Centromeres are indicated by vertical dotted lines and FTL intervals by solid vertical lines and coloured triangles. **C-E**. Genetic distance (cM) measurements for FTL intervals *I2f, 420* and *CEN3* showing individual replicates in black and mean values for the accession cross F_1 in red (Source data Ziolkowski *et al*. 2015). **F**. Heatmap summarising crossover frequency data for F_1 crosses with data from all five intervals. Accessions are listed as rows and fluorescent intervals listed as columns. The heatmap is ordered according to ascending 'Total' cM (red = highest, blue = lowest), which is the sum of the individual interval genetic distances. Col/Col homozygous data, Cvi/Col data and Can/Col data are labelled and highlighted with an arrow in each plot.

region-specific, as they were not detected affecting the intervals on chromosome 1. This is corroborated by the analysis of the individual accession crosses, where the effect on recombination as compared to the Col/Col control varied between intervals e.g. Mt-0/Col F_1 individuals showed an increase in recombination compared to Col/Col in the *I2f* and *420* intervals, but a decrease in *CEN3* (Ziolkowski *et al.* 2015). Another possible source of variation between these accessions is the effect of differing gene-environment interactions on recombination – different accession genomes may interact differently with the same environmental growth conditions, contributing to the observed differences in CO frequency (Mitchell-Olds and Schmitt, 2006).

To map QTL in segregating populations, loci with a large effect on the phenotype are required as their effect will be clearly observable and less likely to be masked by background variation or confounding environmental variation. Interval mapping and introgression methods have low power to detect small-effect loci (Mitchell-Olds and Schmitt 2006; Broman and Sen 2009). Two accessions in particular that stood out from this analysis of accession-cross hybrids were the Cvi-0 and Can-0 accessions, which had the highest recombination overall, when all five intervals were summed (Ziolkowski *et al*. 2015). This suggested potential global modifiers of recombination frequency acting across the genome in these backgrounds (Figure 12F) (Ziolkowski *et al*. 2015). They also had the highest levels of recombination in the *CEN3* interval by a substantial margin (Figure 12E) and were both considerably colder than Col/Col lines in the *420* interval (Figure 12D), suggesting that there may be multiple modifiers with different effects, or individual modifiers with region-specific effects, that are appropriate for mapping.

The choice of accessions to use as parents for a mapping population also depends on the level of genetic variation present between populations. Recent analysis of 1135 Arabidopsis accessions by the 1001 Genomes Consortium (Alonso-Blanco *et al*. 2016) and 78 African accessions by Durvasula *et al*. (2017) has revealed that while most European accessions appear to have resulted from relatively recent expansion of a single clade after the end of the last glacial period, African accessions show more ancestral variation. These accessions contain more private Single Nucleotide Polymorphisms (SNPs), and are more dissimilar to each other than European accessions are. It is likely that Arabidopsis was ancestrally native to Africa and spread into Eurasia under favourable climatic conditions (Durvasala *et al*. 2017; Brennan *et al.* 2014). However, expansion of the species into a new range and characterisation of accession genome differences by an Isolation by Distance (IBD) model does not fully explain the pairwise differences between accession sequences that are observed (Alonso-Blanco *et al*. 2016). Peaks of pairwise differences between accessions do not reflect geographical distance, and extreme pairwise divergences that do not correspond to distance are also observed in 26 accessions. These divergent accessions, named relicts, show divergence from the rest of the European accessions (non-relicts) and from each other. 22 of these accessions are found on the Iberian

Peninsula, and show similarities to Moroccan clades, suggesting that they may be the result of an earlier expansion into Europe across the Strait of Gibraltar that survived subsequent changes in climate in glacial refugia (Durvasula *et al*. 2017; Brennan *et al*. 2014).

Other relict populations were identified, including populations from the Cape Verde Islands and the Canary Islands (Cvi-0 and Can-0) that had divergent genomes from all other accessions analysed (Alonso-Blanco *et al*. 2016). Relicts exist in locations where the climate has changed less since the last glacial period than in most of Europe, which supports the theory that these are from earlier expansions that survived in glacial refugia (Alonso-Blanco *et al*. 2016). Relict populations were found to contain the most genes with potentially deleterious mutations, which is consistent with the smaller effective population size resulting in reduced selection efficiency (Alonso-Blanco *et al*. 2016). This effect could result in rare alleles not seen in non-relict accessions, making relicts a promising avenue for identification of natural variants affecting phenotype, through crosses with non-relict lines and analysis of segregating mapping populations. Cvi-0 had previously been identified as an outlier accession (Nordborg *et al*. 2005), and it is likely that the Cvi-0 and Can-0 divergent genomes, which are from isolated island populations with limited gene flow and arid climates that have contributed to adaptation, could contain rare alleles that differ from the Col-0 accession, the genetic background for the FTL system. Any variants affecting meiotic recombination frequency could be revealed by mapping in accession-cross populations.

As both Cvi-0 and Can-0 accessions show extreme variation in recombination frequency in F_1 crosses with Col-0 when compared to the Col homozygous line (Ziolkowski *et al* 2015), and they are known relicts with divergent genome sequences and phenotypes adapted to arid island climates, they were chosen as promising candidates for mapping genetic recombination modifiers. While using accessions which are extremely divergent from the Col-0 reference genome could reduce genetic mapping power due to the increase in genetic heterogeneity and the number of variants genetically linked to any potential modifier of recombination, it also captures more variation, thereby increasing the probability of identifying a genetic modifier that varies between the accessions (Alonso-Blanco *et al*. 2016). As Can-0 is a late flowering accession, and Cvi-0 is early flowering, Cvi-0 crosses were initially taken forward for QTL mapping.

3.4 Validation of recombination phenotypes in Cvi-0 and Can-0 FTL crosses.

Both Cvi-0 and Can-0 accessions previously showed similar crossover patterns in the F_1 hybrid with Col-0 (Ziolkowski *et al*. 2015), as both showed increases in recombination in the centromeric *CEN3* interval and decreases in sub-telomeric *420* relative to the Col/Col homozygous F_1. While it is plausible that Cvi-0 and Can-0 could have similar modifiers of recombination when compared to a

geographically distinct accession such as Col-0, this is not definitively evident. To clarify this further, the cross to the *420* FTL performed by Ziolkowski *et al.* (2015) was repeated, and a cross to a seed-based FTL interval named *5.11* (Wu *et al*. 2015) spanning the centromere and pericentromere (the region containing predominantly heterochromatin surrounding the centromeric repeats) of chromosome 5, was also conducted (Figure 13 and Figure 14). While analysis had been previously performed on the *CEN3* centromeric interval (Ziolkowski *et al*. 2015), this interval is pollen-based and therefore measures only male recombination (Yelina *et al*. 2013). To achieve a comparable sex-averaged estimate of a centromeric region, a seed-based interval (*5.11*) was required.

While Cvi-0/*420* F_1 and Can-0/*420* F_1 lines behave similarly, having means of 11.8 cM and 11.1 cM respectively compared to the Col-0/*420* F_1 mean of 18 cM (corresponding to a 34% and 38% reduction relative to Col-0/*420* F_1), the two accessions behave very differently within the *5.11* interval. Cvi-0/*5.11* F_1 is significantly hotter than Col-0/*5.11* F_1 (26.1 cM > 19.7 cM, $\chi^2(1) = 138.9628$, $p=8.98 \times 10^{-32}$, increase of 32%) in agreement with *CEN3* F_1 data, whereas Can-0/*5.11* is closer to Col-0, with a mean of 19.9 cM, which is not significantly different ($\chi^2(1) = 0.198319$, p=0.6561).

The Cvi-0 data suggests that modifiers of recombination present in this background do not operate evenly across the genome, as differences are evident between intervals in different positions across the chromosome. This could be due to a property of a *trans*-acting modifier and how it operates, or it could be related to the different *cis*-contexts at centromeric and sub-telomeric intervals discussed previously, broadly relating to chromatin accessibility (see Chapter 1). There is also the possibility of crossover interference contributing to the skew, as an increase in recombination in the pericentromere (*CEN3*) could result in an inhibition of additional recombination events in the surrounding area, including in the subtelomere (the region adjacent to telomeric repeats containing predominantly heterochromatin) (*420*) of the same chromosome arm (Giraut *et al*. 2011; Drouaud *et al.* 2007; Berchowitz and Copenhaver 2010).

The difference in recombination rate between the Can-0/*5.11* F_1 and Cvi-0/*5.11* F_1 lines suggests that Cvi-0 and Can-0 may not have the same *trans*-acting modifiers of meiotic crossover frequency, although it is possible that the same modifiers could be having different effects in *5.11* due to differences in the *cis*-context. For example, DNA methylation is known to affect recombination around the centromere (Yelina *et al*. 2015), and Cvi-0 is notably hypomethylated in comparison to other accessions (Kawakatsu *et al*. 2016), which could potentially contribute to the observed difference in *5.11* recombination. However, when the *5.11* data is compared to another centromeric interval, *CEN3*, which only measures male recombination, the observed difference suggests that this disparity may not be due to a centromere-specific effect, but rather a sex-specific effect - the modifier causing an

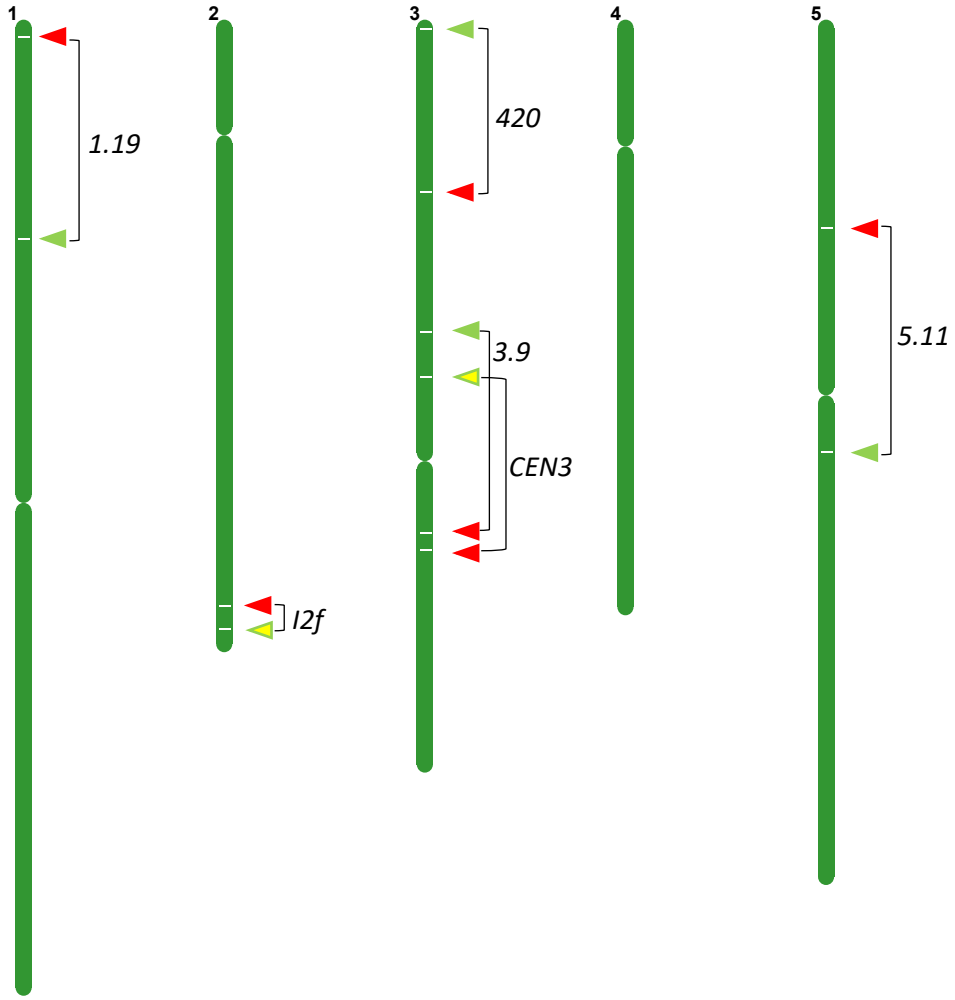


Figure 13. Chromosomal position of FTL intervals used for measurement of recombination. FTL intervals are indicated by black lines. Red triangles denote *dsRed* transgene insertions, green triangles denote *GFP* transgene insertions and yellow triangles denote *eYFP* transgene insertions. Generated using TAIR Chromosome map tool (The Arabidopsis Information Resource (TAIR), https://www.arabidopsis.org/jsp/ChromosomeMap/tool.jsp, on www.arabidopsis.org, 1st August 2017)

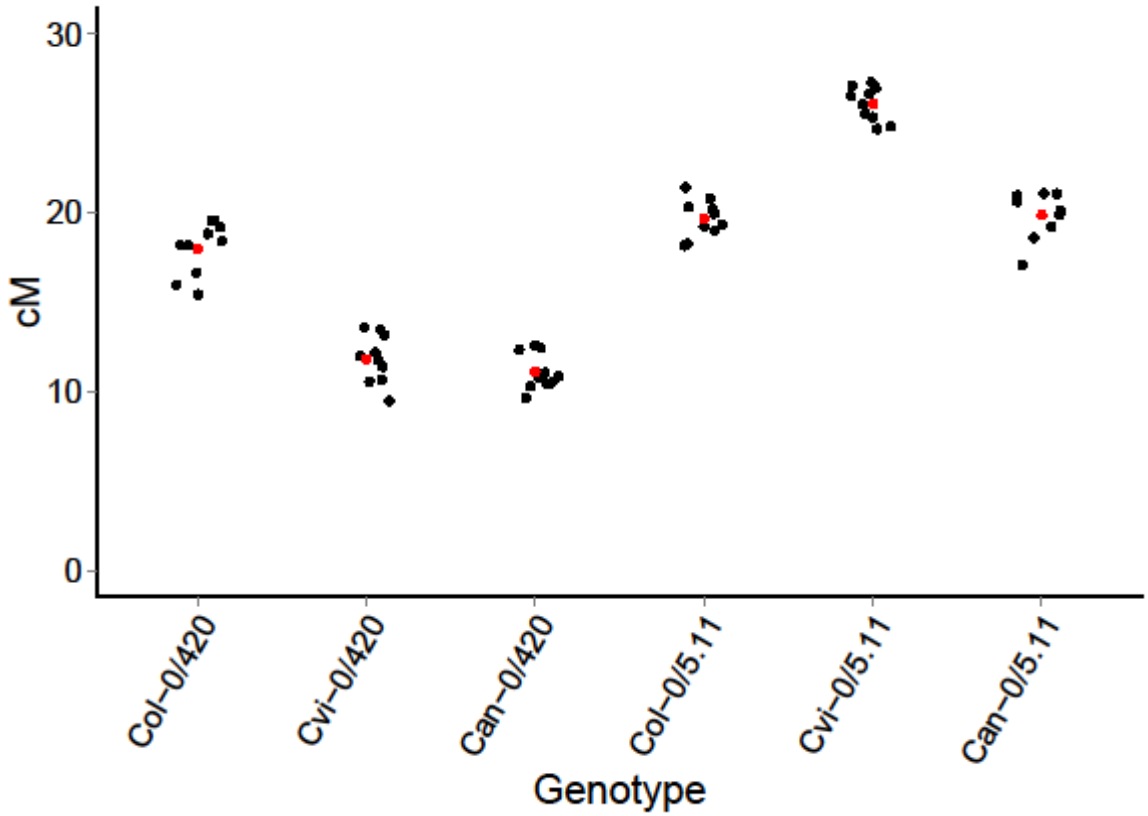


Figure 14: Genetic distance (cM) as measured in *420* and *5.11* seed-based FTL-intervals. Measurements were made in Col-0/*FTL*, Cvi-0/*FTL* and Can-0/*FTL* F_1 individuals (Tables 5 and 6: Total 43693 and 51311 seeds respectively). Replicates are in black, with the mean of each genotype highlighted in red. Differences between Cvi-0/*420* and Col-0/*420* F_1 lines are significant ($\chi^2(1) = 180.3919$, $p=7.96 \times 10^{-41}$), as are differences between Can-0/*420* and Col-0/*420* F_1 lines ($\chi^2(1) = 211.4977$, $p=1.29 \times 10^{-47}$). Differences between Cvi-0/*5.11* and Col-0/*5.11* F_1 lines are significant ($\chi^2(1) = 138.9628$, $p=8.98 \times 10^{-32}$), but differences between Can-0/*5.11* and Col-0/*5.11* F_1 lines are not ($\chi^2(1) = 0.198319$, $p=0.6561$).

increase in Can-0/*CEN3* lines may act only in male meiosis. Alternatively, the modifiers may only act on chromosome 3, as the *5.11* rate doesn't significantly increase from the Col-0 level, despite being sex averaged, and therefore containing male recombination data. From this F_1 data, it seems probable that there are modifiers acting on recombination that could be mapped in Cvi-0/*420*, Can-0/*420*, Cvi-0/*CEN3*, Can-0/*CEN3* and Cvi-0/*5.11* F_2 segregating populations.

3.5 Region-specific effects on recombination

To determine whether the observed *420* and *5.11* recombination measurements were atypical for the regions they represent, analysis was also performed on two more seed-FTL intervals. Interval *3.9* spans the centromere and pericentromere of chromosome 3, and *1.19* covers the subtelomere and part of the euchromatic arm of chromosome 1 (See Table 1 for fluorescent-transgene position)(Figure 15) (Wu *et al*. 2015). The significant increase in centromeric/pericentromeric recombination in Cvi-0 crosses is observed again in interval *3.9* (mean 26.3 cM > 16 cM Col-0/*3.9,* 64% increase relative to Col-0/*3.9* F_1, $\chi^2(1) = 338.760478$, $p=1.19 \times 10^{-75}$). The relative decrease in recombination at the subtelomere is also repeated in interval *1.19* to a lesser, though still significant, degree than in *420* (Cvi-0/*1.19* F_1 mean 17.3 cM < 19.1 cM Col-0/*1.19,* 9% decrease relative to Col-0/*1.19* F_1, $\chi^2(1) = 11.23291$, $p=8.04 \times 10^{-4}$). This suggests that the effect is not interval specific, and is likely to be consistent across chromosomes, although measurement in additional intervals is required to corroborate this more substantially.

In the Col-0/FTL F_1 lines, recombination measurements are consistent between replicates. The range between replicates didn't exceed 6.1 cM (in *1.19,* standard deviation 1.97), and the smaller intervals had much lower variability (*I2f* 1.4 cM, standard deviation 0.47). However, larger standard deviations were observed within Cvi-0/FTL F_1 lines, where variability between replicates was generally much higher. Oddly, while the variability in seed-based intervals (*1.19, 3.9, 420, 5.11*) was comparable with Col lines (e.g. Col-0/*3.9* F_1 lines standard deviation 1.32, Cvi-0/*420* F_1 lines standard deviation 1.55), variability in pollen-based lines was considerably larger (Figure 16). Cvi-0/*CEN3* F_1 lines had a range of 9.3 cM (Standard deviation 2.66) between replicates, compared to 3 cM in Col-0/*CEN3* F_1 lines (Standard deviation 0.96)(Brown-Forsythe Test $P=7.22 \times 10^{-7}$). Similarly, Cvi-0/*I2f* F_1 lines varied by 8.3 cM (Standard deviation 2.35) compared to 1.4 cM in Col-0/*I2f* F_1 lines (Standard deviation 0.47) (Brown-Forsythe Test $P=2.37 \times 10^{-5}$). The reason for this disparity is unclear, as it is unlikely to be a technical issue with pollen preparation or flow cytometry, due to the lower variation between the Col-0/FTL homozygous pollen crossover replicates. Pollen was collected from plants at the same growth stage, and all F_1 plants were grown in the same conditions in parallel, therefore it is also unlikely to be due to variation in environmental factors. The minimal variation observed between replicates for Cvi-

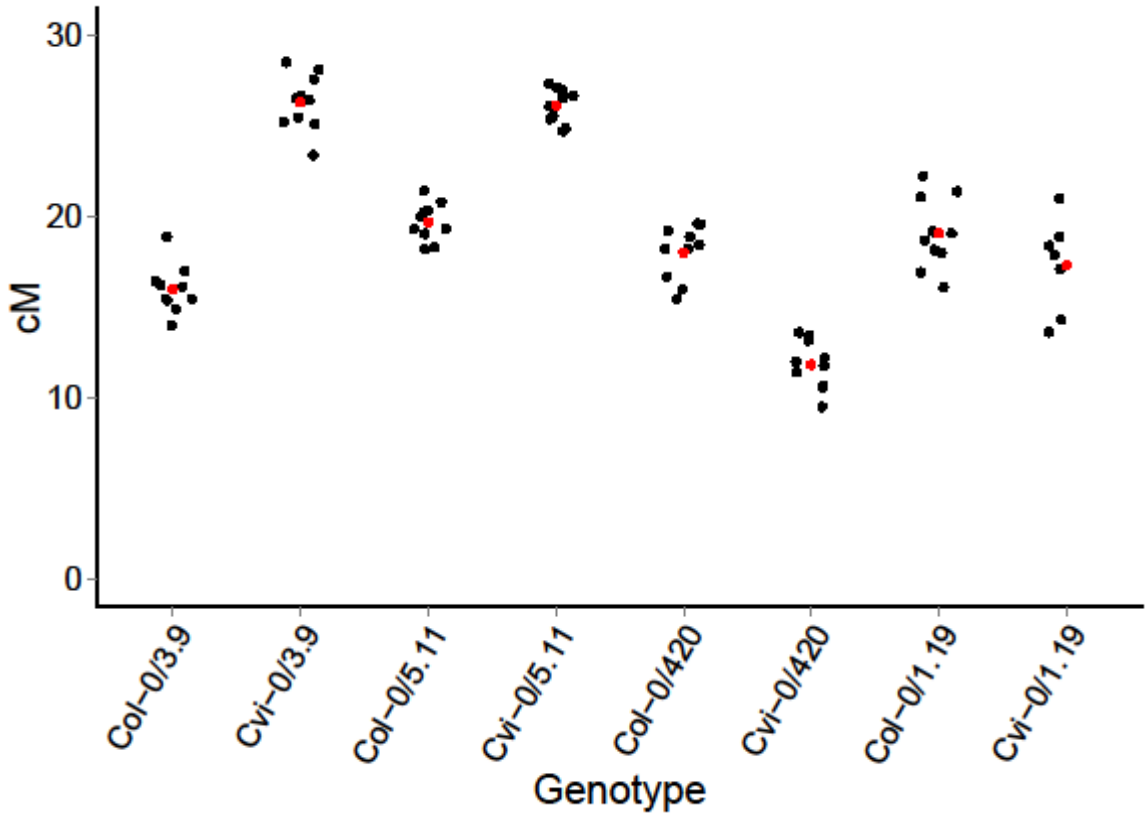


Figure 15: Genetic distance (cM) as measured in seed-based FTL-intervals. Measurements were made in Col-0/*FTL* and Cvi-0/*FTL* F_1 individuals (Tables 5,6,7 and 8: Total 30060, 36049, 31282 and 24606 seeds respectively). Replicates are in black, with the mean of each genotype highlighted in red. *3.9* and *5.11* are intervals spanning centromeric regions, *420* and *1.19* are intervals spanning sub-telomeric and interstitial regions of the chromosome. Differences between Col/Col and Col/Cvi F_1 lines are significant for intervals *3.9, 5.11, 420* and *1.19* ($\chi^2(1) = 338.760478$, $p=1.19 \times 10^{-75}$, $\chi^2(1) = 138.9628$, $p=8.98 \times 10^{-32}$, $\chi^2(1) = 180.3919$, $p=7.96 \times 10^{-41}$, $\chi^2(1) = 11.23291$, $p=0.000804$ respectively).

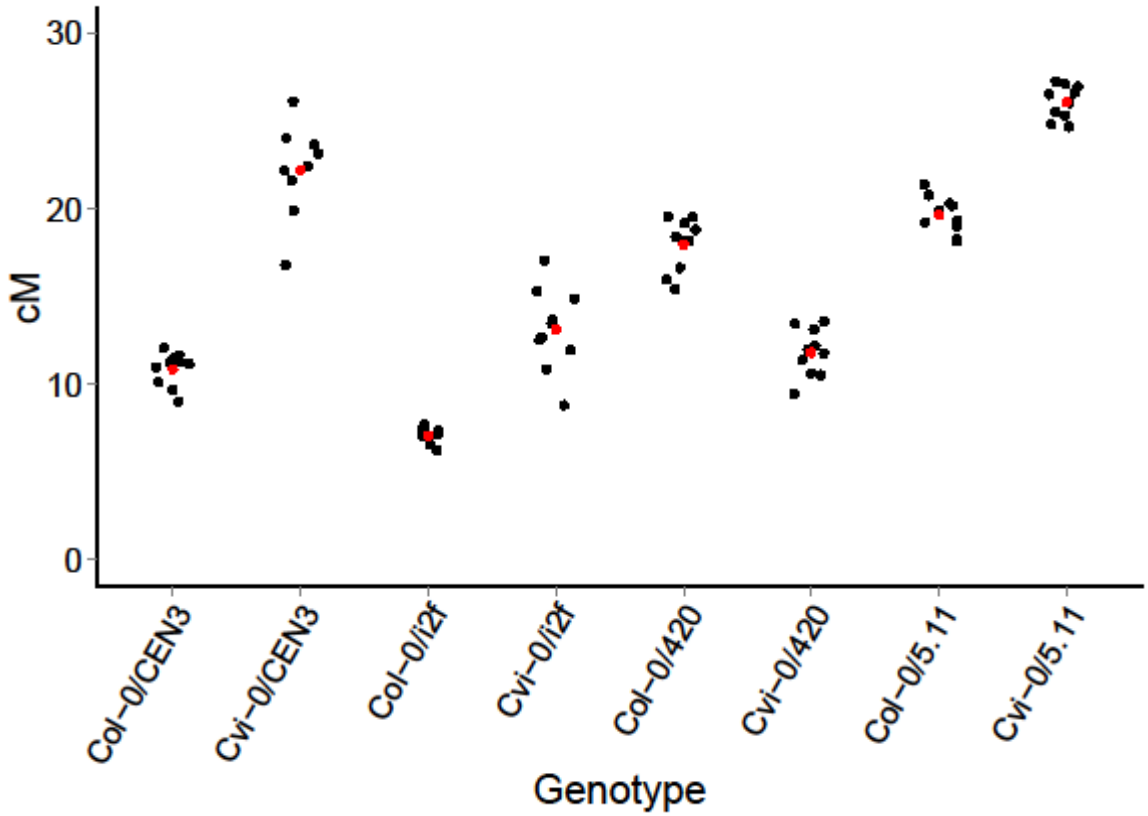


Figure 16: Genetic distance (cM) as measured in *CEN3* and *I2f* pollen-based FTL-intervals and *420* and *5.11* seed-based FTL-intervals. Measurements were made in Col-0/*FTL* and Cvi-0/*FTL* F_1 individuals (Tables 5, 6, 9 and 10: Total 30060 and 36049 seeds, and 430961 and 186326 pollen grains respectively). Replicates are in black, with the mean of each genotype highlighted in red. Differences between Col/Col and Col/Cvi F_1 lines are significant for intervals *CEN3, I2f, 420* and *5.11* ($\chi^2(1)$ = 3097.34, p=0.00, $\chi^2(1)$ = 661.68, p=6.44 x 10^{-146}, $\chi^2(1)$ = 180.3919, p=7.96 x 10^{-41}, $\chi^2(1)$ = 138.9628, p=8.98 x 10^{-32} respectively).

0/FTL F_1 seed-based intervals, which have the same genetic composition with the exception of the FTL transgenes, suggests that the effect is also unlikely to be due to a difference in gene-environment interactions, although it is possible that such an interaction could specifically affect male meiosis. Irrespective of the cause, this inherent variability in the pollen-based systems, specifically in Cvi crosses, could cause difficulties in QTL-mapping populations, with the background fluctuations making it harder to detect true associations. As the Cvi-0/*CEN3* line differs so drastically from the Col line, it is still possible that modifiers could be detected over the background interference. As such, initial mapping populations were constructed for Cvi-0/*CEN3* and Cvi-0/*420*.

It is clear from the intervals analysed that recombination at the centromere is controlled differently to recombination in the sub-telomere, therefore I decided to develop mapping populations for both types of interval. As Cvi-0/*420* F_1 and Cvi-0/*CEN3* F_1 lines demonstrated the largest recombination differences when compared with Col-0 homozygous recombination, they were chosen to be taken forward to generate F_2 mapping populations.

3.6 Seed vs pollen-based FTL systems for measuring crossover frequency

Seed-based FTL systems offer technical benefits over the pollen-FTLs in that individuals can be pre-selected for fluorescence. To score recombination in an FTL background, the transgenes need to be in a hemizygous state to permit segregation, and while recombination can be measured in plants with the transgenes in either a *cis*- (RY/--) or *trans*-configuration (R-/-Y), for consistency with F_1 measurements they are usually measured in *cis*-configuration (RY/--) (Yelina *et al*. 2013). With fluorescence expression in seeds it is possible to select seeds with single-copy red and green transgenes for scoring under a dissecting fluorescence microscope, and while it isn't possible to ensure *cis*-configuration (RY/--) as only the copy number can be determined by fluorescence, most seeds with single-copy fluorescence from an F_1 plant with transgenes in *cis* (RY/--) will contain the transgenes in the parental configuration. This means almost the entire population of plants will be scorable for recombination following pre-selection, increasing the efficiency of the experiment. The seed can be collected and stored, and as they maintain bright fluorescence for long periods of time (several months), they can then be scored at a later date, which is beneficial when scoring large populations of plants.

Pollen-based FTL systems offer their own advantages. Specifically, they allow the measurement of male-specific recombination in segregating populations, and so provide additional information about the individuals. Three-colour pollen-FTLs are also available, measuring recombination in two adjacent intervals, which allows calculation of crossover interference acting across the intervals (Francis *et al*. 2007; Yelina *et al*. 2013; Ziolkowski *et al*. 2015). While this is theoretically possible in seed-FTLs, lines

containing blue fluorescent-transgenes as the third colour are not currently available. Pollen-FTLs can also be scored earlier when plants are flowering, without having to wait for dried seed. However, they cannot be pre-selected for fluorescence before sowing, so a larger proportion of the population that is sown is unscorable due to a non-hemizygous configuration of one or both fluorescence-transgenes. Therefore, plants also need to be phenotyped for fluorescence before flow cytometry. Additionally, the preparation of pollen for flow cytometry takes more time than the preparation of seed for scoring under a fluorescence microscope. Therefore, these factors and practicalities need to be taken into consideration when selecting lines for mapping.

3.7 Mapping recombination QTLs in a Cvi-0/*CEN3* F_2 population

A Cvi-0/*CEN3* F_1 line was self-fertilised to produce seeds that were sown to generate an F_2 mapping population. Plants hemizygous for the FTL transgenes were scored for recombination, where the rest of the genome was mosaic between the two genotypes. 804 plants were sown, and after phenotyping for fluorescence to confirm single hemizygous copies of *Lat52::RFP* and *Lat52::eYFP*, 243 plants were measured for crossover frequency in the *CEN3* interval using flow cytometry. A further 81 samples had to be removed for technical considerations. Specifically, due to; (i) low pollen counts (double-colour parental fluorescence class count below 2500 events) that would yield unreliable estimates of crossover frequency, (ii) distortions in fluorescence classes that indicate silencing of fluorescent transgenes (e.g. where the eYFP single-fluorescence class exceeded the reciprocal RFP single-fluorescence class by more than 100 events, or the double-colour fluorescence class exceeded the reciprocal non-colour fluorescence class) or (iii) fluorescence counts that indicate the transgenes are in a *trans*-configuration (where single-colour fluorescence class counts exceed double-colour fluorescence-class counts). The remaining 162 samples were used for genotyping and rQTL mapping (Table 11).

The variation observed between these samples substantially exceeded the variation observed between F_1 samples (Table 9, Brown-Forsythe Test $P=2.88 \times 10^{-16}$). The Cvi-0/*CEN3* F_2 population had a mean recombination rate of 17.63 cM, which was substantially higher than the Col-0/*CEN3* F_1 mean of 10.89 cM (62% increase relative to Col-0/*CEN3* F_1, Two-sample t-Test assuming unequal variances $p=4.42 \times 10^{-19}$). In addition to significantly higher variation (range= 33.1 cM, standard deviation 4.71, Brown-Forsythe test $P=6.69 \times 10^{-19}$), this suggested that large effect *trans*-acting modifiers could be present within the population (Figure 17A).

To map loci related to the recombination phenotype, DNA was extracted from these individuals and genotyped using SSLP and CAPS markers in an approximately even distribution across the genome (Figure 17B). Some areas of the genome had low marker coverage where appropriate markers could

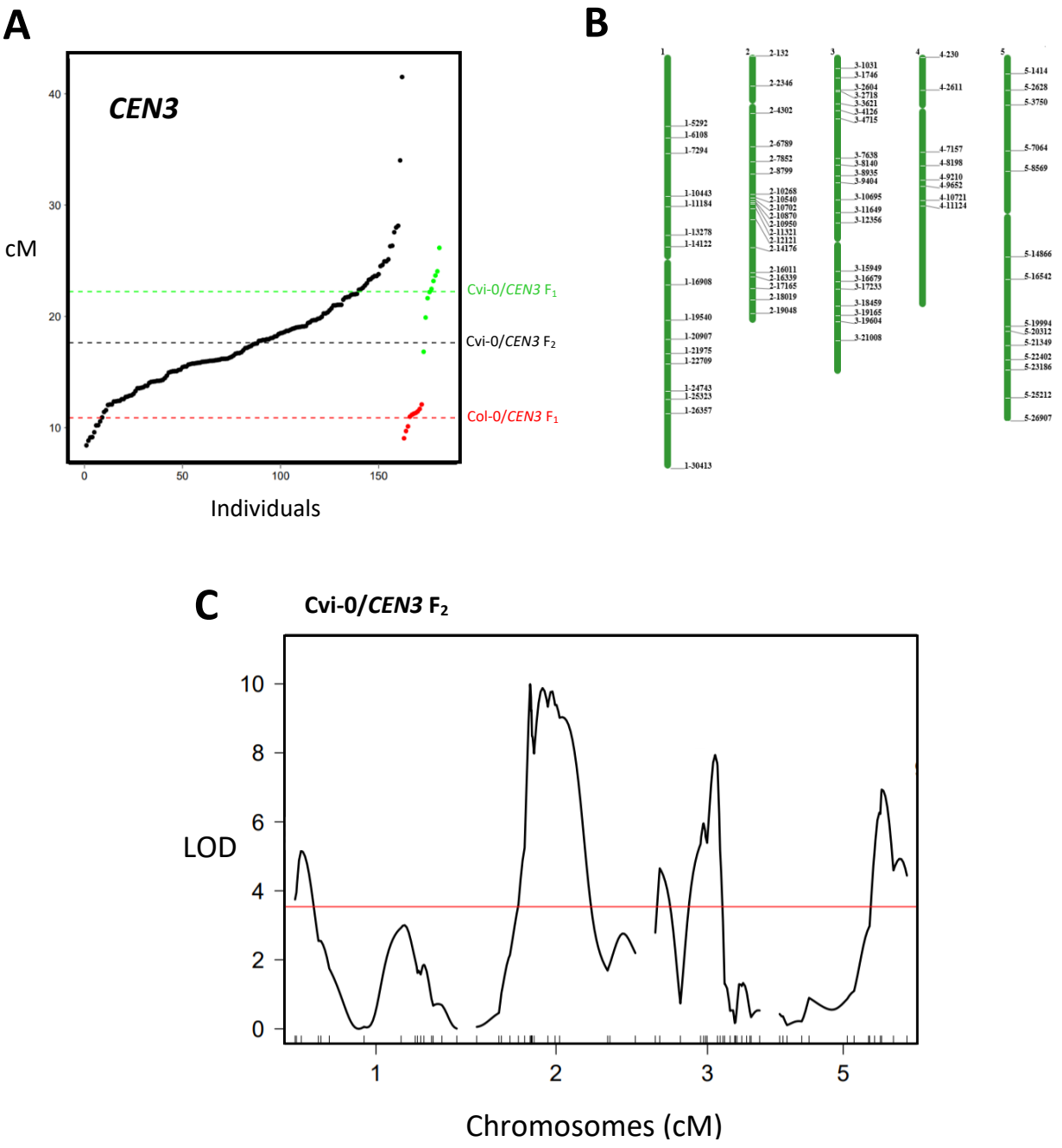


Figure 17. Association of recombination phenotype with marker genotype in the Cvi-0/*CEN3* F_2 mapping population. **A.** Measurements of *CEN3* recombination (cM) in Cvi-0/*CEN3* F_2 (black), Col-0/*CEN3* F_1 (red) and Cvi-0/*CEN3* F_1 (green) individuals. Mean values for each genotype are represented by horizontal dotted lines (Tables 9 and 11: Total 430961 and 2356575 pollen grains respectively). **B**. Distribution of genotyping markers used in the Cvi-0/*CEN3* F_2 population, generated using TAIR Chromosome map tool (The Arabidopsis Information Resource (TAIR), https://www.arabidopsis.org/jsp/ChromosomeMap/tool.jsp, on www.arabidopsis.org, 2nd August 2017). **C**. Logarithm of odds (LOD) scores for association between *CEN3* recombination frequency and genotyping markers in Cvi-0/*CEN3* F_2 population generated by Haley-Knott regression QTL mapping (Table 13). Marker positions indicated by tick marks on the genetic map of chromosomes (cM) on the x-axis. 95% significance threshold indicated by horizontal red line.

not be designed, possibly due to minor differences between the accessions used and the reference sequence from the 1001 genomes project (Alonso-Blanco *et al*. 2016) utilised to design the markers. Additionally, the long arm of chromosome 4 demonstrated a potential segregation distortion in the F_2 population, as markers that had been verified in F_1 and parental lines all returned Col-0 genotypes. This was also detected in the Cvi-0/*420* F_2 population. Nevertheless, sufficient markers (n=78) were verified to give coverage on every chromosome arm to detect any possible linkage with phenotype (Figure 17B).

Genotype and phenotype information for the Cvi-0/*CEN3* F_2 population (n=162) was used to construct a QTL map using the R/qtl package in the statistical program R (Broman *et al.* 2003, Arends *et al*. 2010). This identified QTL on chromosomes 1, 2, 3 and 5 with LOD values of 8.65 (F-test, $p=2.752 \times 10^{-3}$), 14.59 (F-test, $p=6.14 \times 10^{-7}$), 10.17 (F-test, $p=3.67 \times 10^{-4}$) and 9.4 (F-test, $p=1.036 \times 10^{-3}$), respectively (Figure 17C). All four rQTL were above the genome-wide significance threshold (LOD 3.54, Alpha level 0.05, 1000 permutations). The Cvi-0 genotype at $rQTL1^{CEN3}$ and $rQTL2^{CEN3}$ was associated with high recombination rates, and heterozygotes exhibited an intermediate phenotype (Figure 18). This indicates semi-dominance, as previously observed for the *HEI10* rQTL in L*er*/Col-0 crosses (Ziolkowski *et al*. 2017). In contrast, while $rQTL5^{CEN3}$ also exhibits semi-dominance, higher recombination frequency is associated with the Col-0 genotype (Figure 18C).

The LOD peak on chromosome 3 appears to indicate a recessive effect, with individuals Col-0 for the markers around the peak showing much higher recombination than heterozygotes or Cvi-0 individuals (Figure 18D). However, this could be an artefact caused by scoring few individuals with homozygous genotypes, as heterozygotes are the dominant genotype on chromosome 3 due to selection of individuals with proximal FTL transgenes in a hemizygous state. Alternatively, the peak on chromosome 3 could be attributed to the juxtaposition heterozygosity effect observed in previous populations, where homozygosity proximal to the heterozygous FTL interval increased recombination within the interval (Ziolkowski *et al*. 2015).

Consistent with this, marker homozygosity (Col/Col or Cvi/Cvi) either side of the *CEN3* interval on chromosome 3 was associated with high *CEN3* recombination (Table 12). To test for this association, homozygous and heterozygous individuals were counted at each genotyping marker for the population quartiles with highest and lowest recombination (hottest 25% and coldest 25% of the population respectively), and used to construct 2 x 2 contingency tables and perform chi-square tests with FDR correction for multiple testing (Table 12)(Ziolkowski *et al.* 2015). The quartile with the highest *CEN3* recombination levels (hot quartile) showed increased marker homozygosity outside of the *CEN3* interval when compared with intermediate and low recombination quartiles (median and

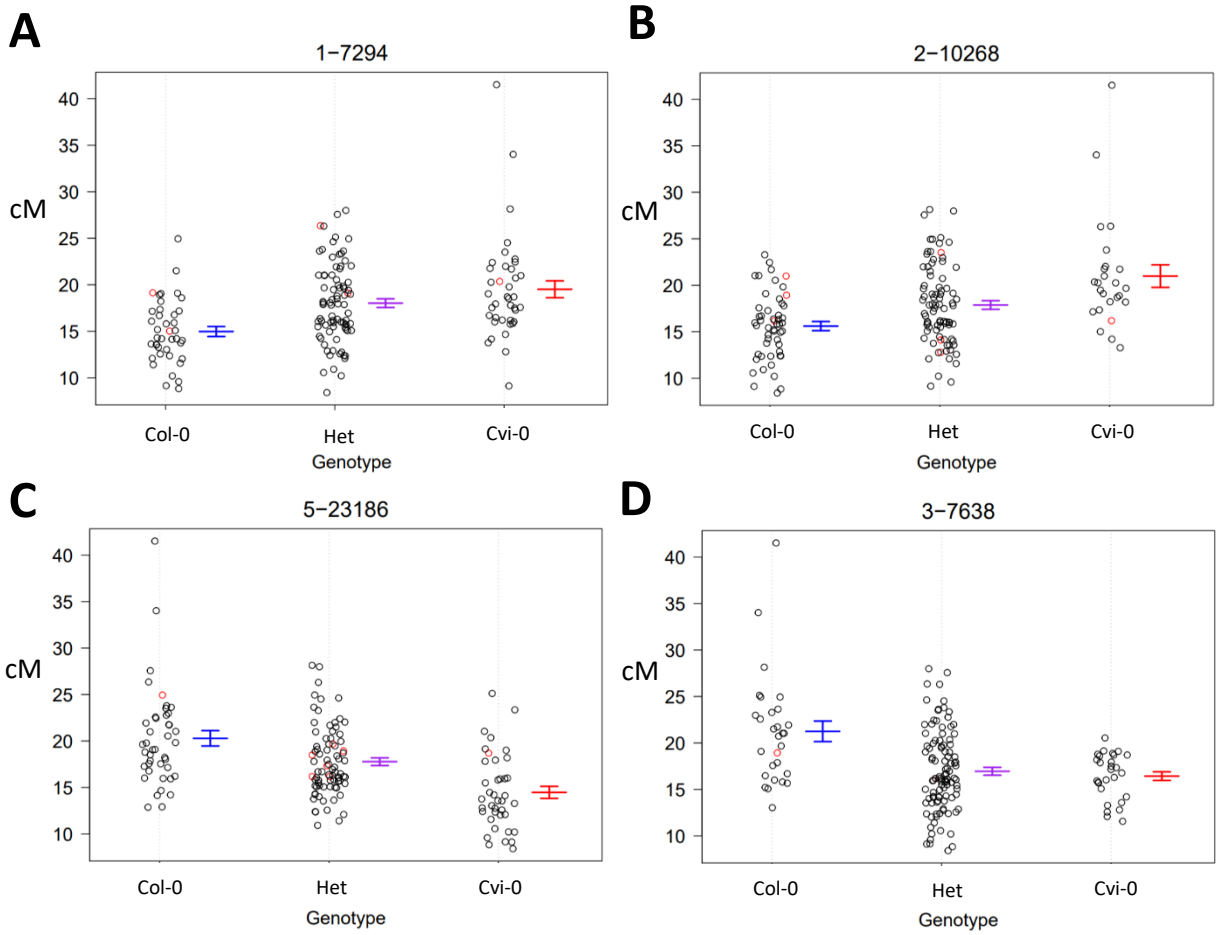


Figure 18. Cvi-0/*CEN3* F_2 individual QTL allele effect dotplots. Effect plots showing recombination frequency (cM) of individuals with Col/Col, Col/Cvi or Cvi/Cvi genotypes at markers most strongly associated with recombination, for each QTL identified in the Cvi-0/*CEN3* F_2 population. Mean (±SEM) given for each genotype at each marker. **A**. rQTL1^{CEN3} marker (7,294,957bp). Col-0 14.9cM (± 0.54); Het 18.1cM (± 0.46); Cvi-0 19.5cM (± 0.93). Differences between Col-0 and Het genotypes, and Col-0 and Cvi-0 genotypes were significant (Two sample t-Test assuming unequal variance, $p=1.91 \times 10^{-5}$ and $p=6.35 \times 10^{-5}$ respectively). The difference between Het and Cvi-0 genotypes was not significant (p=0.17). **B**. rQTL2^{CEN3} marker (10,268,940bp). Col-0 15.6cM (± 0.49); Het 17.9cM (± 0.47); Cvi-0 21cM (± 1.21). Differences between Col-0 and Het genotypes, Col-0 and Cvi-0 genotypes, and Het and Cvi-0 genotypes were significant (Two sample t-Test assuming unequal variance, $p=5.59 \times 10^{-4}$, $p=2.26 \times 10^{-4}$ and p=0.03 respectively). **C**. rQTL5^{CEN3} marker (23,186,865bp). Col-0 20.2cM (± 0.84); Het 17.9cM (± 0.41); Cvi-0 14.5cM (± 0.65). Differences between Col-0 and Het genotypes, Col-0 and Cvi-0 genotypes, and Het and Cvi-0 genotypes were significant (Two sample t-Test assuming unequal variance, p=0.016, $p=8.31 \times 10^{-7}$ and $p=3.69 \times 10^{-5}$ respectively). **D**. rQTL3^{CEN3} marker (7,638,911bp). Col-0 21.3cM (± 1.1); Het 17cM (± 0.42); Cvi-0 16.4cM (± 0.47). Differences between Col-0 and Het genotypes, and Col-0 and Cvi-0 genotypes were significant (Two sample t-Test assuming unequal variance, $p=8.43 \times 10^{-4}$ and $p=2.68 \times 10^{-4}$ respectively). The difference between Het and Cvi-0 genotypes was not significant (p=0.41). Missing genotypes (shown in red) filled in by a random imputation that is conditional on the flanking marker genotypes.

cold quartiles), but this effect was not significant after correction for multiple testing. The effect also did not extend far beyond the *CEN3* interval, as positive associations were only detected up to distances of 3.5 Mb and 0.7 Mb on either side of the interval. This distance is in agreement with the effect observed in the Ct-1/*CEN3* F_2 population and is consistent with adjacent heterozygosity patterns affecting local recombination frequency (Ziolkowski *et al*. 2015). While the observed effect was not statistically significant, it is possible that it contributes to the variation observed within the *CEN3* interval, as the pattern of heterozygosity juxtaposition follows the same trend as that observed previously (Ziolkowski *et al*. 2015).

QTL interactions were tested for using Hayley-Knott regression (Haley and Knott 1992; Broman and Sen 2009; Broman 2012), according to the full model y= $rQTL1^{CEN3}$ + $rQTL2^{CEN3}$ + $rQTL3^{CEN3}$ + $rQTL5^{CEN3}$ + $rQTL1^{CEN3}$: $rQTL2^{CEN3}$ + $rQTL1^{CEN3}$: $rQTL3^{CEN3}$ + $rQTL1^{CEN3}$: $rQTL5^{CEN3}$ + $rQTL2^{CEN3}$: $rQTL3^{CEN3}$+ $rQTL2^{CEN3}$: $rQTL5^{CEN3}$ + $rQTL3^{CEN3}$: $rQTL5^{CEN3}$, where : designates an interaction. The full model's total LOD is 32.94, which explains 60.8% of the variance in recombination frequency. $rQTL1^{CEN3}$, $rQTL2^{CEN3}$, $rQTL3^{CEN3}$ and $rQTL5^{CEN3}$ explain 10.9%, 20.1%, 13.1% and 12% of the variance individually, respectively. There were no significant interactions detected between any of the rQTL (Table 13).

While a significant proportion of the variation is explained by the model with all of the rQTL in combination (60.8%, F-test $p=2.35x10^{-14}$), none of the individual peaks had LOD values high enough to make subsequent mapping likely to be successful. By comparison the previous L*er*/*420* F_2 population, used to map *HEI10*, contained rQTL with LOD values over 40 (Ziolkowski *et al*. 2017). Given the inherent variability that was present in Cvi/*CEN3* F_1 measurements and the potential heterozygosity *cis*-effect present in the F_2 (although this could be minimised by back-crossing to reduce heterozygosity) it was possible that further mapping would be unable to resolve these low-effect loci. As alternative mapping populations were also being explored that offered more significant QTL, this population was not taken further.

As Can-0/*CEN3* F_1 and Cvi-0/*CEN3* F_1 lines showed similar divergence from the Col-0/*CEN3* F_1 reference recombination phenotype (Ziolkowski *et al*. 2015), it was possible that rQTL could be mapped in a segregating F_2 population from this cross as an alternative, as this population may not contain the same level of background variability. However, attempts to score this population resulted in only 30 samples (Table 14), as loss of a large proportion of the samples due to prolific silencing of the *RFP* and *eYFP* transgenes in this background prevented a sufficiently large population being generated for mapping. Comparable silencing was not observed in Can-0/*420* lines, suggesting that this effect is not due to an alteration in the general silencing mechanism in Can-0. It is possible that expression from the pollen-specific *Lat52* promoter was increased in the Can-0 background due to differences in

transcription factors, and that this over-expression triggered silencing, however this is speculative. The phenomenon was also frequently observed in Can-0/*CEN3* F_1 lines, preventing replication of the measurements made previously (Ziolkowski *et al*. 2015). The 30 F_2 samples obtained had a mean recombination rate of 15.45 cM compared to Col/*CEN3* F_1 10.89 cM (42% increase relative to Col-0/*CEN3* F_1, Two-sample t-Test assuming unequal variances $p=5.13 \times 10^{-4}$), and their variation was significantly larger than the F_1 (Table 9, Brown-Forsythe Test $P=5.32 \times 10^{-4}$), suggesting that modifiers were present (Figure 19).

3.8 Cvi-0/*I2f* F_2 data

The Cvi-0/*I2f* F_1 data also offered mapping potential as the recombination frequency diverged considerably from the Col-0 homozygous estimate, indicating the presence of genetic modifiers (mean 13.1 cM, Col-0/*I2f* F_1 mean 7.1 cM, 85% increase relative to Col-0/*I2f* F_1, $\chi^2(1) = 661.68$, $p=6.44 \times 10^{-146}$). An F_1 line was self-fertilised to generate a small preliminary F_2 population, consisting of 47 individuals (Table 15), to look for variation in recombination indicative of segregating modifiers. This population had a mean of 8.9 cM, which was similar to the Col-0/*I2f* F_1 mean (25% increase relative to Col-0/*I2f* F_1), although the variation between lines was significantly larger (Figure 19 and Table 10, Brown-Forsythe Test P= 0.016). This variation suggests the presence of recombination modifiers varying between Col-0 and Cvi-0 backgrounds. However, the Cvi-0/*I2f* F_1 lines exhibited substantial amounts of background variation between replicates (8.26cM range, standard deviation 2.35). As this background variation was also observed in Cvi-0/*CEN3* lines which, despite showing even more variation in the F_2, were not ultimately considered suitable for further mapping, this population was believed to be unsuitable for QTL analysis and was therefore not extended for mapping.

3.9 Cvi-0/*420* F_2 mapping population

While the rQTL mapping populations using pollen-FTL systems had not yielded any large effect QTL, the crossover frequency variation in these populations, which significantly exceeded F_1 variation, indicated the presence of genetic modifiers of recombination frequency varying between accessions. As seed-based FTL systems were more consistent between F_1 replicates (Figure 16), and successful mapping had previously been performed in a L*er*/*420* F_2 population (Ziolkowski *et al*. 2017), I decided to attempt to map Col/Cvi rQTL using the *420* interval. Notably, Cvi-0/*420* F_1 lines differed significantly from Col-0/*420* F_1 lines ($\chi^2(1) = 180.3919$, $p=7.96 \times 10^{-41}$), although in this case crossover frequency was lower than Col/*420* (Figure 14, Table 5).

A Cvi-0/*420* F_1 line was self-fertilised to generate an F_2 mapping population (n=220). These were scored for recombination frequency in the interval between the *RFP* and *GFP* seed-expressed transgenes by

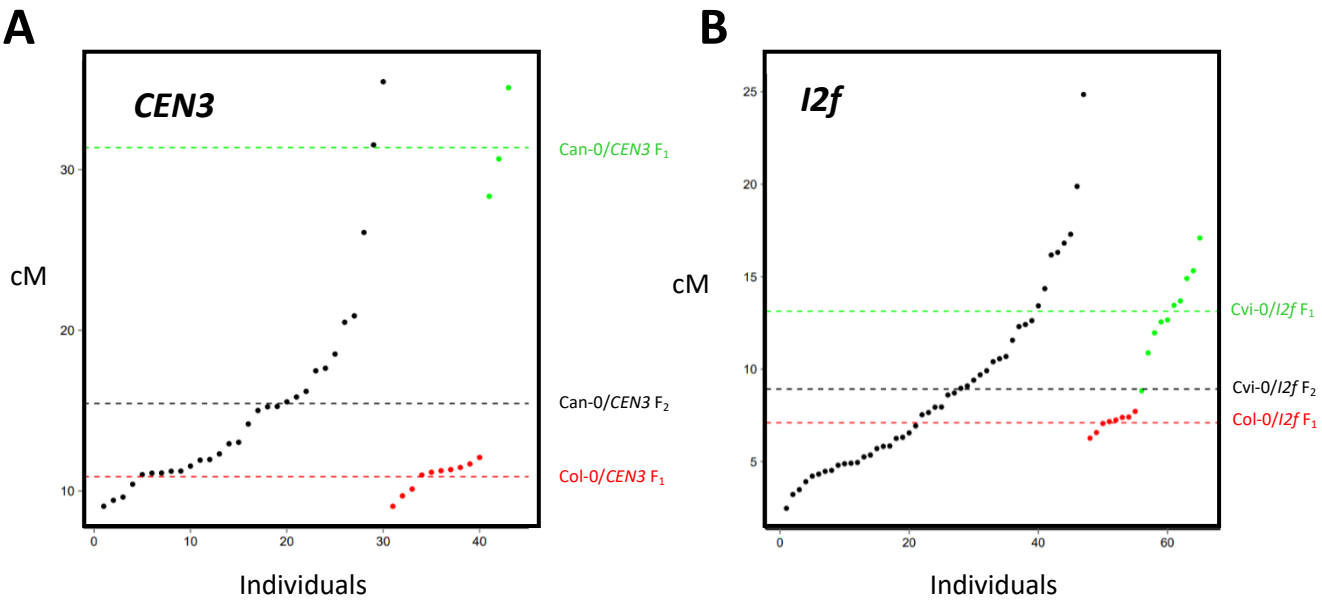


Figure 19. Observation of variation in recombination frequency in Can-0/*CEN3* F_2 and Cvi-0/*I2f* F_2 populations. **A.** Measurements of *CEN3* recombination (cM) in Can-0/*CEN3* F_2 (black), Col-0/*CEN3* F_1 (red) and Can-0/*CEN3* F_1 (green) individuals. Mean values for each genotype are represented by horizontal dotted lines (Tables 9 and 14: Total 152458 and 344193 pollen grains respectively. Can-0/*CEN3* F_1 data reproduced from Ziolkowski *et al*. 2015). **B**. Measurements of *I2f* recombination (cM) in Cvi-0/*I2f* F_2 (black), Col-0/*I2f* F_1 (red) and Cvi-0/*I2f* F_1 (green) individuals. Mean values for each genotype are represented by horizontal dotted lines (Tables 10 and 15: 186326 and 495415 pollen grains respectively).

measurement of seed fluorescence under a dissecting epifluorescence microscope (Table 16, Figure 10) and analysed using a CellProfiler image analysis pipeline (Carpenter *et al*. 2006; Ziolkowski *et al*. 2015). The Cvi-0/*420* F_2 population had a mean recombination frequency of 13.45 cM, compared to the Col-0/*420* F_1 mean of 18 cM (25% decrease relative to Col-0/*420* F_1, Figure 20A). The variation in the F_2 population also significantly exceeded the F_1 variation (Brown Forsythe Test $P=1.14x10^{-7}$), with a range of 31.1 cM (standard deviation 5.11, Figure 20A). This indicated the presence of large effect modifiers of recombination frequency segregating in the population.

F_2 individuals (n=220) were genotyped using SSLP and CAPS markers across the genome (n=64), again giving an approximately even coverage of chromosomes (Figure 20B). The markers used differed slightly from the Cvi-0/*CEN3* F_2 mapping population, as additional markers were designed and subsequent marker choices were adjusted to provide evenly distributed coverage. This information was utilised to create an F_2 QTL map of the recombination modifiers using R/qtl.

QTL were identified on chromosomes 1, 2 and 5 with LOD values of 44.02 (F-test, $p<2 \times 10^{-16}$), 16.5 (F-test, $p=3.18 \times 10^{-11}$) and 19.72 ($p=6.33 \times 10^{-14}$) respectively (Figure 20C). All three QTL had a significant effect on recombination (Threshold LOD 3.03, Alpha level 0.05, 1000 permutations). The Col-0 genotype at $rQTL1^{420}$ and $rQTL5^{420}$ was associated with higher recombination rate, whereas at $rQTL2^{420}$ the Cvi-0 genotype was associated with higher recombination (Figure 21). For all three QTL, heterozygous individuals exhibited an intermediate phenotype, suggesting semi-dominance. QTL interactions were tested for using Hayley-Knott regression, according to the full model $y = rQTL1^{420} + rQTL2^{420} + rQTL5^{420} + rQTL1^{420}: rQTL2^{420} + rQTL1^{420}: rQTL5^{420} + rQTL2^{420}: rQTL5^{420}$, where : designates an interaction. The full model's total LOD is 58.5, and this explains 70.6% of the variance in recombination frequency. $rQTL1^{420}$, $rQTL2^{420}$ and $rQTL5^{420}$ respectively explain 44.4%, 12.1% and 15% of the variance individually. No significant interactions were detected between rQTL (Table 17). No heterozygosity *cis*-effect was detected on chromosome 3 either, suggesting that the local genotype had no effect on *420* recombination.

Together, these three rQTL peaks explain most of the variation in crossover frequency observed in the population, with the rest attributable to multiple small effect loci below the significance threshold (Mitchell-Olds and Schmitt, 2006). Interestingly, some parallels can be drawn between the Cvi-0/*CEN3* and Cvi-0/*420* F_2 populations. While the rQTL on chromosome 1 are on opposite arms in each population, the rQTL on chromosome 2 are on the same arm. The peak markers for each are 4 Mb apart, so it is unlikely that the rQTL can be attributed to the same locus (Figures 18B and 21B), although peaks can shift when mapping is refined by additional recombination events and genotyping markers (Broman and Sen 2009; Zeng 1994). However, both rQTL behave semi-dominantly with the Cvi-0 allele

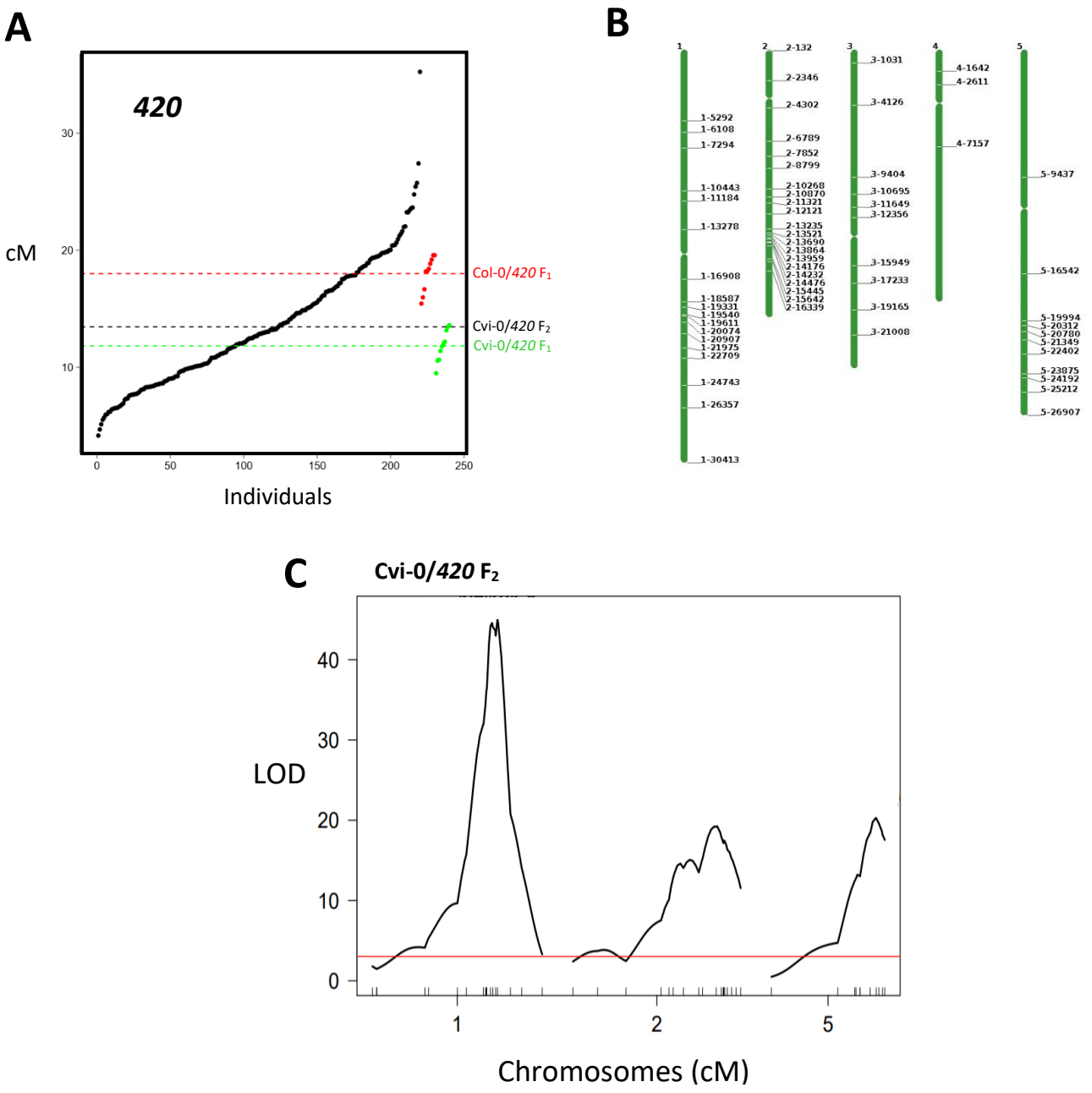


Figure 20. Association of recombination phenotype with marker genotype in the Cvi-0/*420* F_2 mapping population. **A.** Measurements of *420* recombination (cM) in Cvi-0/*420* F_2 (black), Col-0/*420* F_1 (red) and Cvi-0/*420* F_1 (green) individuals. Mean values for each genotype are represented by horizontal dotted lines (Tables 5 and 16: Total 30060 and 393897 seeds respectively). **B**. Distribution of genotyping markers used in the Cvi-0/*420* F_2 population, generated using TAIR Chromosome map tool. **C**. Logarithm of odds (LOD) scores for association between *420* recombination frequency and genotyping markers in Cvi-0/*420* F_2 population generated by Haley-Knott regression QTL mapping (Table 17). Marker positions indicated by tick marks on the genetic map of chromosomes (cM) on the x-axis. 95% significance threshold indicated by horizontal red line.

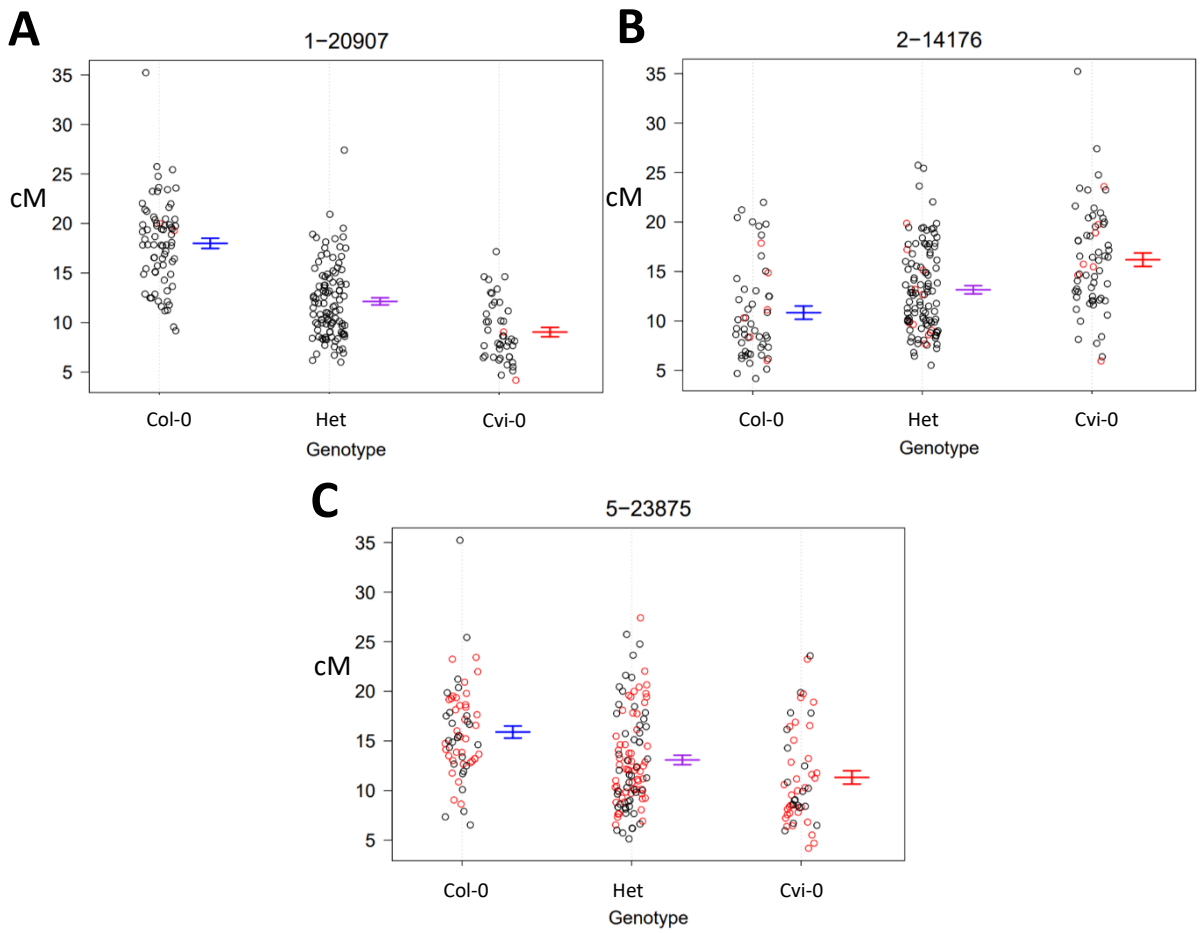


Figure 21. Cvi-0/*420* F_2 individual QTL allele effect dotplots. Effect plots showing recombination frequency (cM) from individuals with Col/Col, Col/Cvi or Cvi/Cvi genotypes at markers most strongly associated with recombination, for each QTL identified in the Cvi-0/*420* F_2 population. Mean (±SEM) given for each genotype at each marker. **A**. $rQTL1^{420}$ marker (20,907,282bp). Col-0 18cM (± 0.52); Het 12.1cM (± 0.36); Cvi-0 9cM (± 0.47). Differences between Col-0 and Het genotypes, Col-0 and Cvi-0 genotypes, and Het and Cvi-0 genotypes were significant (Two sample t-Test assuming unequal variance, $p=3.37 \times 10^{-16}$, $p=1.29 \times 10^{-23}$ and 1.08×10^{-6} respectively). **B**. $rQTL2^{420}$ marker (14,176,271bp). Col-0 10.8cM (± 0.67); Het 13.1cM (± 0.42); Cvi-0 16.1cM (± 0.66). Differences between Col-0 and Het genotypes, Col-0 and Cvi-0 genotypes, and Het and Cvi-0 genotypes were significant (Two sample t-Test assuming unequal variance, $p=4.29 \times 10^{-3}$, $p=1.48 \times 10^{-7}$ and 2.41×10^{-4} respectively). **C**. $rQTL5^{420}$ marker (23,875,653bp). Col-0 16cM (± 0.64); Het 13cM (± 0.46); Cvi-0 11.6cM (± 0.67). Differences between Col-0 and Het genotypes, and Col-0 and Cvi-0 genotypes were significant (Two sample t-Test assuming unequal variance, $p=1.88 \times 10^{-4}$ and $p=4.08 \times 10^{-6}$ respectively). The difference between Het and Cvi-0 genotypes was not significant (p=0.07). Missing genotypes (shown in red) filled in by a random imputation that is conditional on the flanking marker genotypes.

associated with the higher recombination phenotype, so it is possible that they are caused by the same locus (Figures 18B and 21B). rQTL5^{CEN3} and rQTL5^{420} have peak markers that are less than 1 Mb apart on chromosome 5, and they also behave similarly with Col-0 genotypes being associated with higher recombination and heterozygotes having intermediate crossover frequency (Figures 18C and 21C). These rQTL are likely to be related to the same locus, suggesting that some rQTL are acting across the genome in different intervals, while others act in a region-specific manner. To test whether these are the same rQTL acting in different intervals, Near Isogenic Lines (NILs) containing the Cvi-0 allele of the rQTL in a predominantly Col-0 background could be constructed and crossed to different FTLs. Measurement of recombination in different intervals could demonstrate whether the rQTL is region-specific, or capable of affecting recombination in multiple intervals (see Chapter 4).

rQTL1^{420} is the largest rQTL in the population, and therefore the most obvious choice for further mapping to identify a candidate gene. However, the rQTL peak is located in close proximity to a recombination modifier found in a previous L*er*/*420* F_2 mapping population, the *HEI10* E3 ubiquitin ligase gene (Ziolkowski *et al*. 2017). rQTL1^{420} behaves semi-dominantly in the same way as *HEI10* in the L*er*/*420* F_2 population, with the Col-0 genotype associated with higher recombination (Figure 21A). Information on polymorphisms in the *HEI10* gene from the 1,001 Genomes Project (Alonso-Blanco *et al*. 2016) revealed that there are a number of differences between the Col-0 and Cvi-0 alleles, including some that are shared between Cvi-0 and L*er*-0. In particular, there is a nonsynonymous polymorphism in the HEI10 C-terminal region that is shared between the L*er*-0, Bur-0 and Cvi-0 accessions (R264G), all of which show an rQTL peak in a similar position on chromosome 1 in *420* cross F_2 populations (Ziolkowski *et al*. 2017). While this has not been definitively proven as the causal variant, it is consistent with current data (Ziolkowski *et al*. 2017). It seemed probable therefore, that the gene underlying rQTL1^{420} was *HEI10*, and while further work was required to confirm this effect, rQTL1^{420} was not considered a strong candidate for mapping a novel modifier locus. In light of this, rQTL2^{420} and rQTL5^{420} were taken forward for fine-mapping to identify candidate loci.

3.10 Can-0/*420* F_2 mapping population

It was established by the F_1 accession data (Figures 12 and 14) that while Cvi-0 and Can-0 seemed to have similar recombination patterns in many intervals, they differed in some respects, such as in the *5.11* interval. Therefore, it is likely that they may differ for underlying recombination modifiers and generating additional mapping populations utilising both accessions could identify further recombination QTL.

A Can-0/*420* F_2 population (n=114) was generated from a self-fertilised F_1 plant and scored for recombination frequency using seed fluorescence segregation (Table 18). The population had a mean

recombination frequency of 14.64 cM, below the Col-0/*420* F_1 mean of 18 cM (19% decrease relative to Col-0/*420* F_1, Figure 22A). The variation in recombination phenotype was significantly higher than that observed between the Col-0/*420* F_1 lines (Brown Forsythe Test $P=1.16 \times 10^{-5}$), with a range of 15.8 cM (standard deviation 3.1). Although considerably smaller than the recombination range of the Cvi-0/*420* F_2 population, this still suggested the presence of segregating recombination modifiers in the population (Figure 22A).

57 SSLP and CAPS genotyping markers were developed to differentiate between the Col-0 and Can-0 genotypes across the genome, and then used to genotype the population (n=114) (Figure 22B). This information was used in conjunction with the recombination frequency data to create an F_2 QTL map of recombination modifiers using R/qtl. Significant QTL were identified on chromosomes 3 and 4 with LOD values of 8.98 (F-test, $p=7.86 \times 10^{-7}$) and 13.6 (F-test, $p=8.1 \times 10^{-11}$), respectively (Figure 22C). Homozygosity was associated with higher recombination at the rQTL on chromosome 3, with heterozygous individuals having lower recombination than either homozygous genotype (Figure 23).

At CanQTL4 (the QTL identified on chromosome 4), the Can-0 genotype was associated with higher recombination (Figure 23B). The peak marker on chromosome 4 appears to behave recessively, although semi-dominance is also possible, as the mean heterozygous crossover phenotype is slightly higher than in the Col-0 genotype. Both rQTL had a significant effect on recombination (Threshold LOD 2.9, Alpha level 0.05, 1000 permutations).

QTL interactions were tested for using Hayley-Knott regression, according to the full model y = CanQTL3 + CanQTL4 + CanQTL3:CanQTL4, where : denotes an interaction. The total LOD is 20.5 for the full model, explaining 56.3% of the variation in *420* recombination in the population. CanQTL3 and CanQTL4 individually explain 19.1% and 32% of the variance, respectively. No significant interaction was detected between the two rQTL (Table 19). The homozygosity effect observed at CanQTL3 mirrors a similar effect observed in a Ct-1/*CEN3* F_2 population (Ziolkowski *et al*. 2015), which was accounted for by a *cis*-acting juxtaposition effect related to local patterns of heterozygosity around the FTL interval being used for crossover measurement. To clarify whether this was the cause of the effect, markers on chromosome 3 were analysed in the Can-0/*420* F_2 population.

Homozygosity (Col/Col or Can/Can) outside of the *420* interval on chromosome 3 was associated with high recombination within *420* (Figure 24 and Table 20). Homozygous and heterozygous individuals were counted at each genotyping marker for the population quartiles with highest and lowest recombination (hot and cold quartiles respectively) and used to construct 2 x 2 contingency tables and perform chi-square tests with FDR correction for multiple testing (Table 20). The hot recombination quartile showed significantly greater marker homozygosity outside *420*, compared to the intermediate

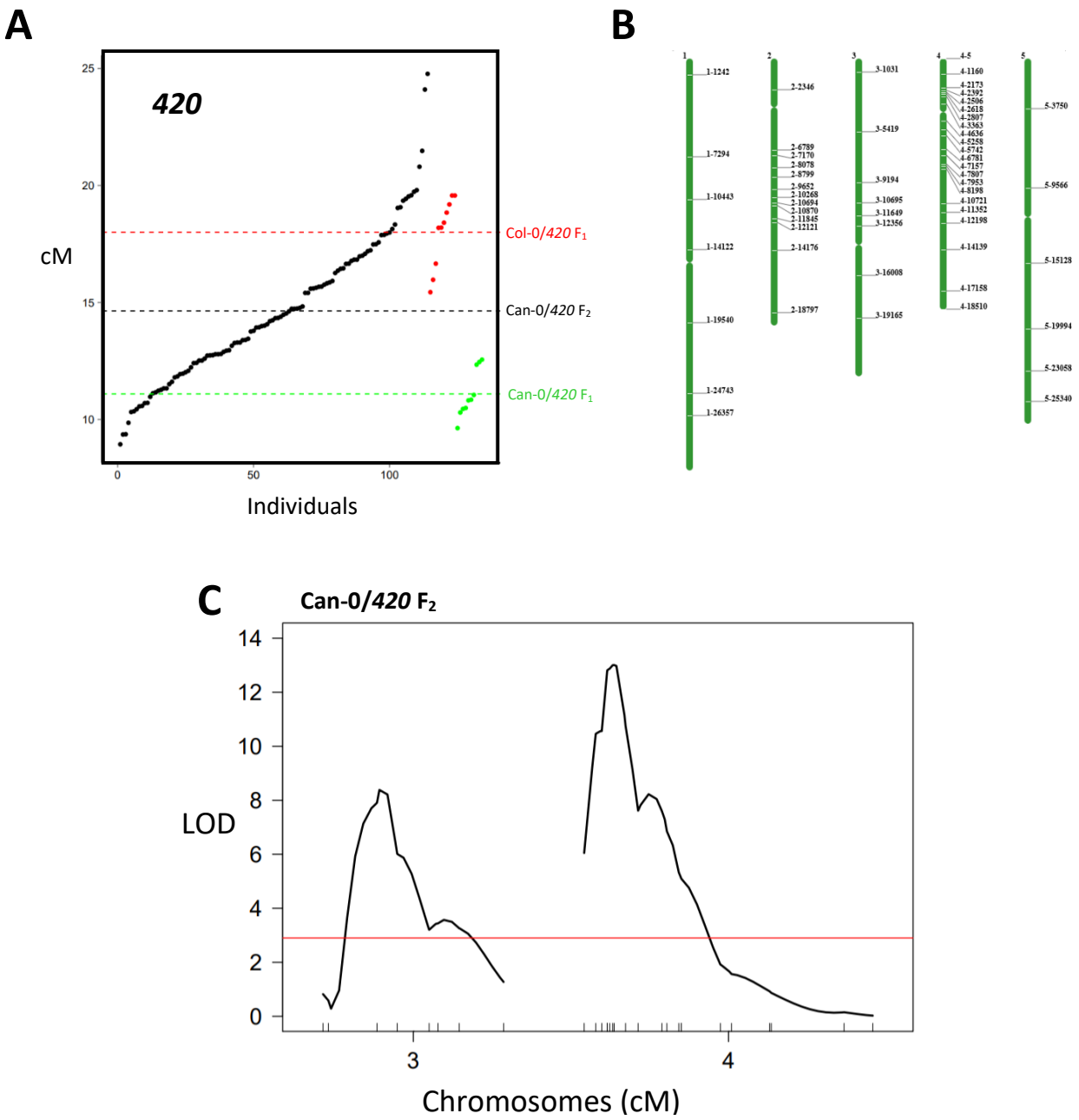


Figure 22. Association of recombination phenotype with marker genotype in the Can-0/*420* F_2 mapping population. **A.** Measurements of *420* recombination (cM) in Can-0/*420* F_2 (black), Col-0/*420* F_1 (red) and Can-0/*420* F_1 (green) individuals. Mean values for each genotype are represented by horizontal dotted lines (Tables 5 and 18: Total 29886 and 174789 seeds respectively). **B**. Distribution of genotyping markers used in the Can-0/*420* F_2 population, generated using TAIR Chromosome map tool. **C**. Logarithm of odds (LOD) scores for association between *420* recombination frequency and genotyping markers in Can-0/*420* F_2 population generated by Haley-Knott regression QTL mapping (Table 19). Marker positions indicated by tick marks on the genetic map of chromosomes (cM) on the x-axis. 95% significance threshold indicated by horizontal red line.

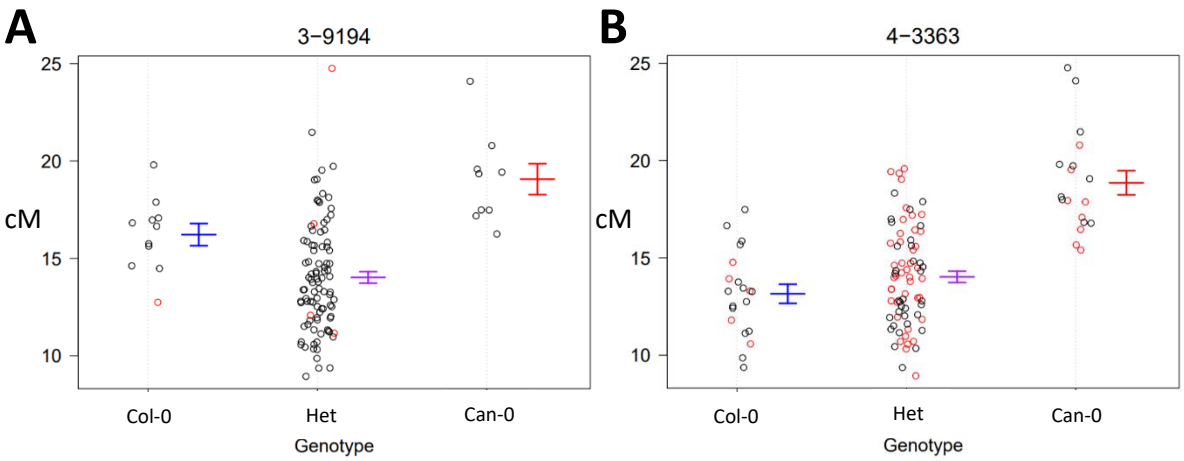


Figure 23. Can-0/*420* F_2 individual QTL allele effect dotplots. Effect plots showing recombination frequency (cM) from individuals with Col/Col, Col/Can or Can/Can genotypes at markers most strongly associated with recombination, for each QTL identified in the Can-0/*420* F_2 population. Mean (±SEM) given for each genotype at each marker. **A**. CanQTL3 marker (9,194,020bp). Col-0 16.2cM (± 0.57); Het 14cM (± 0.3); Can-0 19.1cM (± 0.79). Differences between Col-0 and Het genotypes, Col-0 and Can-0 genotypes, and Het and Can-0 genotypes were significant (Two sample t-Test assuming unequal variance, $p=3.49 \times 10^{-3}$, p=0.0105 and 1.38×10^{-4} respectively). **B**. CanQTL4 marker (3,363,996bp). Col-0 12.9cM (± 0.46); Het 13.9cM (± 0.28); Can-0 18.9cM (± 0.62). Differences between Col-0 and Can-0 genotypes, and Het and Can-0 genotypes were significant (Two sample t-Test assuming unequal variance, $p=8.86 \times 10^{-9}$ and $p=1.77 \times 10^{-7}$ respectively). The difference between Col-0 and Het genotypes was not significant (p=0.0697). Missing genotypes (shown in red) filled in by a random imputation that is conditional on the flanking marker genotypes.

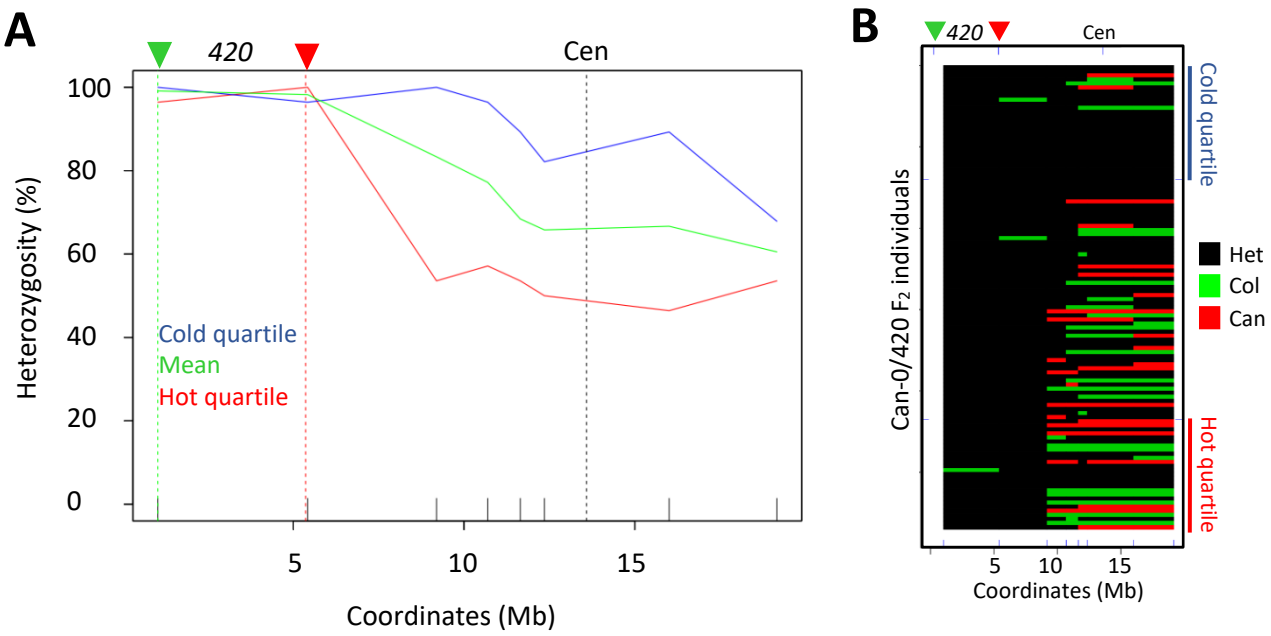


Figure 24. Modification of *420* recombination frequency by effect of genotype in flanking regions in Can-0/*420* F_2 population. **A**. Percentage of heterozygosity along chromosome 3 in the highest (red), and lowest (blue) F_2 recombination quartiles (Table 20). Mean shown in green. Location of *420* fluorescent T-DNAs and the centromere are indicated by vertical dashed lines. **B**. Chromosome 3 genotypes for Can-0/*420* F_2 individuals ranked by recombination frequency. Each horizontal row corresponds to one F_2 individual. X-axis ticks indicate genotyping marker position. *420* fluorescent T-DNAs and centromere indicated at top. Hot and cold quartiles (the 25% of the population with the highest and lowest recombination frequencies respectively) marked on the right.

and cold quartiles (Figure 24). This effect extended the length of chromosome 3, with the difference between quartiles decreasing with distance from the *420* interval, although the difference was only significant for the 5 markers in closest proximity to *420* (Figure 24A and Table 20). This further confirms that the juxtaposition of the heterozygous *420* interval with a homozygous region can modify local crossover frequency (Ziolkowski *et al.* 2015).

Despite parallels in recombination patterns at the F_1 level (Figure 14), Cvi-0 and Can-0 *420* F_2 populations revealed entirely different underlying modifiers. This indicates that recombination is controlled differently in these accessions, and therefore may not be attributed to shared modifiers. It is also interesting that a *cis*-effect is observed in Can-0, but not in Cvi-0, although the reason for this is unclear. CanQTL4 explains a substantial proportion of the variance in the Can-0/*420* F_2 population, and its effect is not masked by other QTL, potentially making it easier to map, as multiple backcross generations are not required to remove other segregating QTL. Therefore, CanQTL4 was taken forward for further mapping to identify potential candidate genes underlying the recombination phenotype.

3.11 Cvi-0/*5.11* F_2 mapping population

While rQTL that were promising for fine-mapping had been identified in two separate populations, both of these populations measured recombination in a sub-telomeric interval (*420*). As such, these modifiers had not been demonstrated to have a global effect on recombination, and therefore could be region-specific. As previous attempts to map rQTL in the centromeric pollen-FTL *CEN3* had not provided suitable modifiers for fine-mapping due in part to inherent variability in Cvi-0/pollen-FTL lines, the *5.11* interval was used in an attempt to map modifiers affecting recombination in the centromere and pericentromere using a seed-FTL crossover reporter. F_1 data had shown that the Cvi-0/*5.11* cross gave less variable estimates of recombination than Cvi-0/*CEN3* lines (range 2.61 cM < 9.34 cM). As background variability was predicted to be lower in the *5.11* F_2 population than in *CEN3*, it was possible that mappable rQTL could be identified.

A Cvi-0/*5.11* F_2 population (n=111) was generated from a self-fertilised F_1 line and recombination frequency was measured using seed fluorescence segregation patterns (Table 21). The mean recombination frequency of the F_2 population was 27 cM, significantly higher than the corresponding mean of Col-0/*5.11* F_1 lines (19.67 cM, 37% increase relative to Col-0/*5.11* F_1, Two-sample t-Test assuming unequal variances $p=8.66 \times 10^{-14}$). The variation in recombination frequency in the F_2 population was significantly higher than the variation between Col-0/*5.11* F_1 lines (Brown Forsythe Test $P=8.66 \times 10^{-14}$), with a range of 14.1 cM (standard deviation 2.51). This was comparable with the

Can-0/*420* F_2 population variation, indicating that modifiers of recombination frequency affecting the *5.11* interval may be segregating in the population (Figure 25A).

A preliminary genome scan was performed using 27 SSLP and CAPS genotyping markers across the genome to ascertain whether any significant rQTL could be detected (Figure 25B). An F_2 rQTL map was created using R/qtl (Figure 25C), but no significant rQTL or interactions were detected (Threshold LOD 2.88, Alpha level 0.05, 1000 permutations). No *cis*-effect was detected on chromosome 5, suggesting that local heterozygosity patterns did not affect recombination within the *5.11* interval. This indicates that while considerable variation is present in recombination rates, it cannot be attributed to large-effect loci. Therefore, variation must be caused by a combination of small-effect loci below the significance threshold and potential allele-environment interactions (Mitchell-Olds and Schmitt, 2006). As the Cvi-0/*CEN3* F_2 population yielded similarly small-effect loci, it is possible that recombination at the centromere is controlled by minor variations in multiple processes, rather than by large changes in a few modifiers as seems to be the case in the sub-telomeric *420* interval.

3.12 Conclusions

Significant variation in recombination frequency is present in Arabidopsis accession-FTL crosses. Dissecting the relative causes of this variation in different accessions and different FTL-intervals revealed that while significant amounts of variation can be attributed to minor loci, allele-environment effects, technical variance and *cis*-acting heterozygosity juxtaposition effects, large-effect loci can be identified in mapping populations. Four significant *trans*-acting modifiers were identified that could be refined to identify candidate genes: $rQTL1^{420}$, $rQTL2^{420}$ and $rQTL5^{420}$ in a Cvi-0/Col-0 cross, and CanQTL4 in a Can-0/Col-0 cross. As a putative gene for $rQTL1^{420}$ was already known, further work focused on confirming the *HEI10* locus as the underlying gene, and fine-mapping candidate genes for the remaining three QTL.

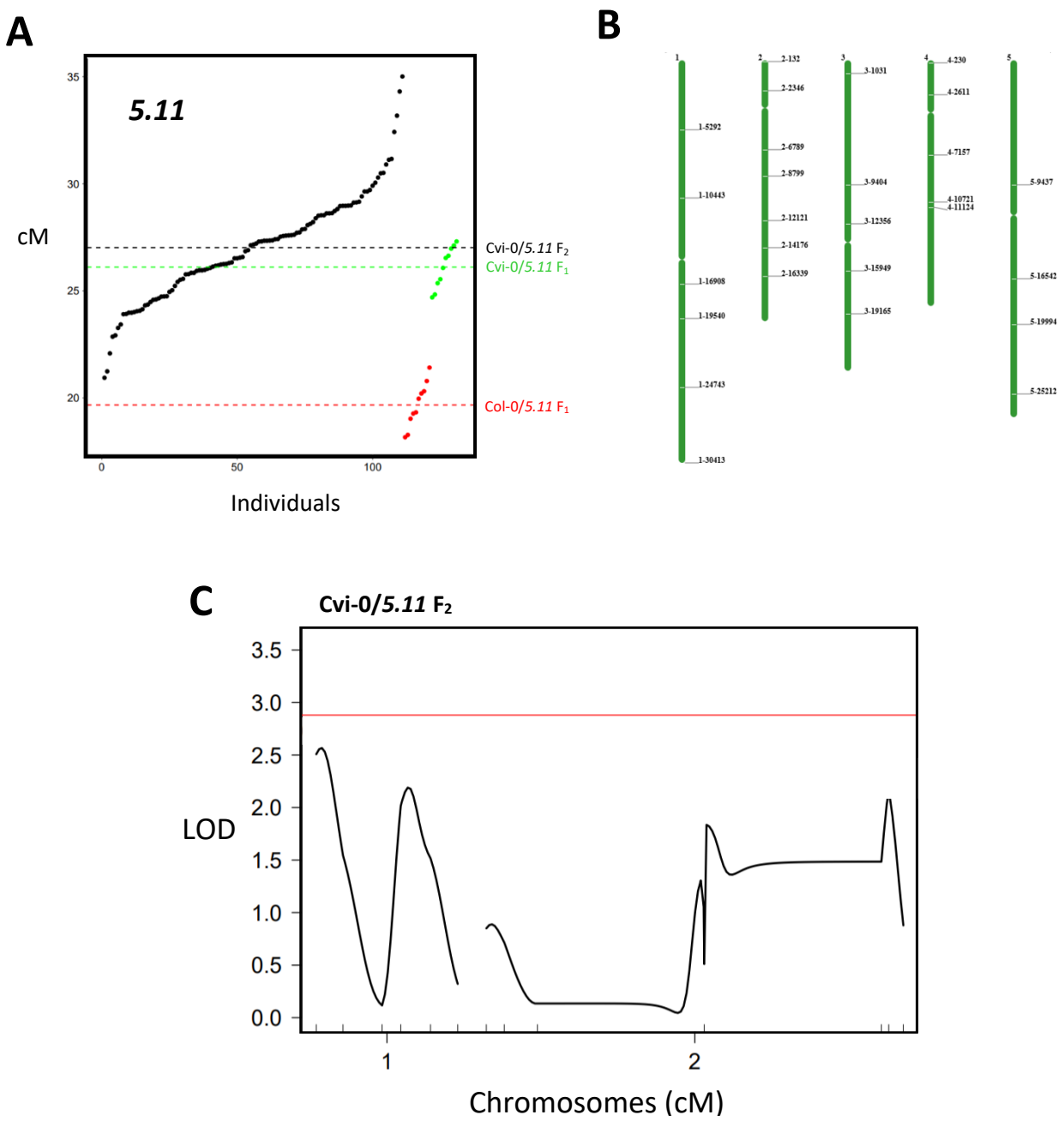


Figure 25. Association of recombination phenotype with marker genotype in the Cvi-0/*5.11* F_2 mapping population. **A.** Measurements of *5.11* recombination (cM) in Cvi-0/*5.11* F_2 (black), Col-0/*5.11* F_1 (red) and Cvi-0/*5.11* F_1 (green) individuals. Mean values for each genotype are represented by horizontal dotted lines (Tables 6 and 21: Total 36049 and 179180 seeds respectively). **B**. Distribution of genotyping markers used in the Cvi-0/*5.11* F_2 population, generated using TAIR Chromosome map tool. **C**. Logarithm of odds (LOD) scores for association between *5.11* recombination frequency and genotyping markers in Cvi-0/*5.11* F_2 population generated by Haley-Knott regression QTL mapping. Marker positions indicated by tick marks on the genetic map of chromosomes (cM) on the x-axis. 95% significance threshold indicated by horizontal red line.

Chapter 4 – Fine-mapping of genetic modifiers of meiotic recombination frequency and identification of candidate genes

4.1 Summary

As described in the previous chapter, several large-effect, potentially novel, modifiers of meiotic recombination frequency were detected in *Arabidopsis thaliana* accession-cross segregating F_2 QTL mapping populations. In this chapter I describe how individual QTLs from the Cvi-0/*420* F_2 and Can-0/*420* F_2 populations were introgressed into the Col-0 background through backcrossing and marker-assisted selection (MAS), allowing quantification of their individual effects. Generation of additional recombinant lines narrowed QTL intervals and allowed generation of candidate gene lists for $rQTL2^{420}$, $rQTL5^{420}$ and CanQTL4. Analysis of T-DNA insertion mutants allowed functional analysis of candidate genes for each QTL via testing for significant effects on recombination in the *420* subtelomeric FTL-interval.

4.2 Introduction: QTL fine-mapping; resolving candidate loci and confirming effects on phenotype

Identification of QTL affecting meiotic recombination frequency has been performed previously in Arabidopsis via genetic fine-mapping (Esch *et al*. 2007; Ziolkowski *et al*. 2017), which can narrow the credible interval sufficiently to identify candidate loci. *Arabidopsis thaliana* has a relatively small genome of only 125 megabases (Mb), with a gene density of approximately 200 per Mb (Swarbreck *et al*. 2008). This means that any credible interval determined to contain the QTL being mapped will still most likely contain multiple genes. As 1-2 crossovers form per chromosome per Arabidopsis meiosis (Giraut *et al*. 2011; Mercier *et al*. 2015), when fine-mapping, the number of recombinant plants dividing the QTL interval into smaller intervals is low in any population. Therefore, large recombinant populations and multiple backcross generations may be required to refine the interval sufficiently to identify a small number of candidate genes. In cases where the QTL in question is in an area of the genome with low levels of recombination, for example near the centromere (Nachman 2002; Kauppi, Jeffreys and Keeney 2004; Ellermeier *et al.* 2010), mapping can require substantially larger populations to identify recombinant lines.

One strategy for identification of genes underlying QTLs is to introgress the QTL into a common genetic background, such as the Col-0 reference accession, to generate a near isogenic line (NIL) (Alonso-Blanco and Koornneef 2000). The quality and efficacy of QTL mapping depends on the overall heritability of the trait in the population, and how much of this can be attributed to an individual QTL. Therefore, removal of other segregating QTL and potential small effect loci and *cis*-acting effects, such as heterozygosity (Ziolkowski *et al*. 2015), through back-crossing into the Col-0 genetic background,

should improve QTL mapping resolution (Alonso-Blanco and Koornneef 2000). Isolation of individual QTL in NILs also allows comparison and characterisation of QTL as would be performed for newly-isolated mutants, such as through analysis of gene expression or of QTL interactions with environmental factors (Alonso-Blanco and Koornneef 2000). QTL NILs can also be used to test complementation between QTL alleles and candidate mutants to confirm underlying genes. Repeated back-crosses utilising MAS to identify the most isogenic lines can generate a NIL in just a few generations (5-7), and substantially refine the QTL interval, thereby reducing the number of potential candidate genes (Alonso-Blanco and Koornneef 2000).

Once candidate genes have been identified by fine-mapping refinement of the QTL interval, they need to be analysed to determine precisely which locus is causing the accession-specific effect on the phenotype, and to quantify the effect of the causative locus. This can be achieved by isolating independent mutant alleles in candidate genes. In Arabidopsis, large collections of insertional mutagenesis lines are available, with known insertions disrupting most genes in the genome (Alonso *et al*. 2003; Sessions *et al*. 2002; Kleinboelting *et al*. 2012). Analysis of appropriate lines disrupting the genes of interest can reveal a particular locus as having an effect on the phenotype, and this locus can then be taken forward for further characterisation and confirmation of phenotype (Alonso-Blanco and Koornneef 2000). A drawback of this strategy is that insertional mutant alleles may differ from natural accession alleles, for example which may be hypomorphs, and therefore have a different effect. An alternative strategy, which can be used in conjunction with mutant analysis to confirm candidate genes, or as a stand-alone approach to assess candidates, is cloning and complementation of candidates using transformation (Mitchell-Olds and Schmitt 2006). Cloning of accession-specific alleles and construction of transgenes can be used for reciprocal transformations. Transformation of one allele into the genetic background of the other accession parent can then be used to prove an effect, although this is dependent on dominance effects of alleles (Alonso-Blanco and Koorneef 2000). Interestingly, this approach lead to observation of a dosage-dependent effect on recombination rate following transformation of the *HEI10* E3 ubiquitin ligase gene (Ziolkowski *et al*. 2017).

In some cases, QTL peak intervals may contain genes already known to influence the phenotype being studied. In these instances, analysis of candidate genes to confirm allele-specific effects may be directed to these known genes. For example, when mapping a large-effect QTL of recombination frequency in a L*er*-0/Col-0 cross QTL-mapping population by Ziolkowski *et al*. (2017), the *HEI10* gene, which was known to be required for Class I crossovers in Arabidopsis (Chelysheva *et al*. 2012), mapped in the vicinity of the peak associated region. Further genetic analysis, including *HEI10* transformation of L*er*-0 and Col-0 alleles into the Col-0/*420* reporter line, revealed that the two alleles had different

effects on recombination, confirming *HEI10* as the gene underlying the identified QTL (Ziolkowski *et al*. 2017).

If no obvious candidate is revealed within the QTL credible interval, then systematic elimination of the candidate genes must be performed to identify causative loci, and multiple candidate genes may need to be assessed through mutant screening and/or transgenic complementation experiments. Evidence used to identify likely candidate genes can include expression pattern, protein localisation, polymorphisms between parental accessions and predicted or known gene functions (Koornneef *et al*. 2004). However, in complex traits such as recombination, genes that do not appear to have any functional relevance may indirectly affect the phenotype. Loci affecting gene expression, protein translation, post-translational modifications, stress responses and cellular conditions, to name but a few, can have pleiotropic effects on the cell which may include effects on recombination. Therefore, while these genes may not be considered true regulators of recombination, and they offer limited real-world applications through genetic manipulation due to their pleiotropic effects, they may appear as QTLs during mapping experiments, and must therefore still be appraised as potential candidates. It is also possible that the QTL peak is not due to a single locus, but that it is caused by multiple tightly linked loci (Alonso-Blanco and Koornneef 2000), suggesting that analysis of multiple candidates can be a beneficial strategy.

Identification of loci and the causal polymorphisms underlying the QTL can be complicated further by genetically linked polymorphisms (Mitchell-Olds and Schmitt. 2006; Koornneef *et al*. 2004). In accession-cross populations generated for mapping QTL, considerable genetic differences may exist between parental accessions (Alonso-Blanco and Koornneef 2000). The use of two parental lines that have extreme pairwise sequence divergence, such as using relict lines Cvi-0 or Can-0 as one parent (Alonso-Blanco *et al*. 2016), can be beneficial in increasing the probability of observing differences in phenotype, but it may limit the ability to resolve the underlying polymorphisms, due to their density. However, transformation experiments such as those discussed earlier, can be performed using modified transgenes in order to identify the causal polymorphism (Ziolkowski *et al.* 2017; Filiault *et al.* 2008; Glazier, Nadeau and Aitman 2002). Typically, multiple genetic differences will exist between the parental accessions in candidate genes. Therefore, even if a QTL is mapped to a narrow region of the genome, the causal variant may not be identified, as was the case for a recombination QTL identified by Esch *et al*. (2007).

In light of these considerations, the strategy chosen for identification of the loci underlying the peaks I had identified in the F_2 mapping populations was to backcross QTLs into the Col-0 background and utilise MAS to identify recombinants within the broad QTL intervals. Once the intervals were

narrowed, candidate gene lists were generated and then queried using information concerning cellular location, expression pattern, polymorphisms and gene function. This generated a priority list of candidates for each of the three QTL (rQTL2^{420}, rQTL5^{420} and CanQTL4) that were then assessed by analysis of T-DNA insertion mutants crossed to the *420* FTL reporter line.

4.3 Fine-mapping rQTL2^{420}

QTL mapping in the Cvi-0/*420* F_2 population revealed a significant large-effect QTL on chromosome 2 (LOD 16.5, F-test p=3.18 x 10^{-11}), affecting recombination in the *420* interval and explaining 12.1% of the variance in the population (Figure 20C). The Cvi-0 allele at the peak QTL marker showed an association with higher recombination compared to the Col-0 allele, with heterozygous individuals having an intermediate crossover frequency (Figure 21B). This QTL peak did not match the position of any known modifiers of meiotic recombination frequency (Lawrence *et al*. 2017), so it was assumed to represent a novel recombination modifier of interest for fine mapping.

While the F_2 population had indicated a broad interval of the genome that contained the QTL, further recombination events were required to narrow this interval and allow identification of potential candidate genes. Additionally, the effects of rQTL1^{420} and rQTL5^{420}, which were also significant modifiers segregating in the F_2 population (Figure 20C), had to be removed. Segregating smaller effect loci and potential *cis*-acting factors in the F_2 population also have minor effects on recombination, and therefore it is beneficial to remove these to facilitate genetic mapping. To achieve this, an F_2 line (individual 10-6) was selected that was fixed in the Col-0 genotype for the majority of the genome but had chromosome 2 fixed as Cvi-0. This individual was self-fertilised and an F_3 line was backcrossed to Col-0, then MAS was used to identify a progeny line that was Col-0 genotype for all tested markers across chromosomes 1, 4 and 5 and for all markers across chromosome 3 except two at the end of the chromosome farthest from the *420* interval, which were heterozygous (markers '17233' position 17,233,598 bp, and '19165' position 19,165,521 bp) (line Cvi*420*F_3(F_2 10-6)/Col 1-3, Figure 26). This meant that (i) the other two major QTL in the population were fixed for the Col-0 genotype, (ii) any heterozygosity effects acting in *cis* on chromosome 3 were removed, and (iii) most small effect loci would also be fixed in the Col-0 genotype. Chromosome 2 markers were all heterozygous (Figure 27).

Cvi*420*F_3(F_2 10-6)/Col 1-3 (a BC_1F_1 line) was then self-fertilised to generate a large fine-mapping population (BP, a BC_1F_2 population) to identify recombinants on chromosome 2 near the QTL peak that could narrow the credible interval. 512 BC_1F_2 lines were sown and genotyped, and after removal of lines that did not contain a recombination event in the region surrounding the peak marker, 164 individuals were scored for *420* recombination (Table 22). Three lines from this population with promising recombination events breaking up the QTL region were then taken forward to generate

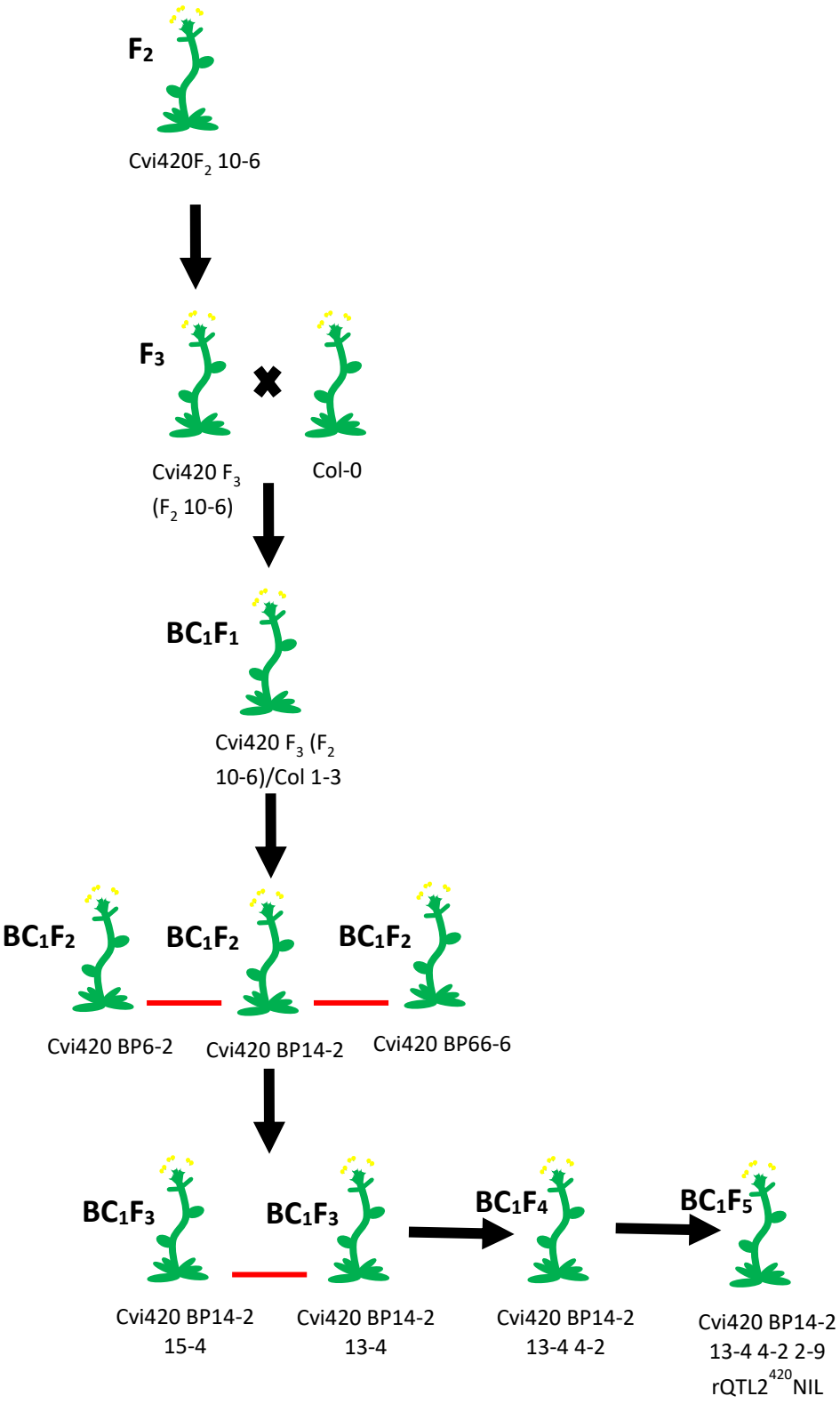


Figure 26: Schematic of rQTL2^{420} fine mapping plant pedigree. Plants were genotyped using accession-specific SSLP and CAPS markers (see Appendix) and measured for recombination in the *420* FTL interval. For genotyping see Figure 27. For scoring see Table 22, Table 23 and Table 24. Black arrows represent generation of new seedlings. Black crosses represent a cross-fertilisation. Sibling relationships are represented by red lines between plants.

Figure 27. SSLP and CAPS genotyping data for lines generated for fine-mapping of rQTL2^{420}. Parent line denotes the line that was self-fertilised to generate the individual, the Generation column denotes the generation that the individual line belongs to, and the cM column denotes the recombination measurement of the *420* FTL interval in the individual. Genotyping markers are indicated by a number representing their location in the genome (chromosome-kilobase) e.g. marker 1-6108 is on chromosome 1 at position 6108789bp. Col-0 genotype is indicated by the letter C, Cvi-0 genotype by the letter V and segregating markers by the letter H for each individual at each genotyping marker.

				Genotyping markers																	
Parent line	Individual	Generation	cM	1-6108	1-16908	1-19540	1-30413	2-132	2-12121	2-16339	3-1031	3-4126	3-15949	3-19165	4-230	4-2611	4-7157	5-9437	5-19994	5-25212
Cvi-0/420 F_1	10-6	F_2	23.43	H	H	H	C	V	V	V	H	H	H	V	C	C	C	H	C	C
Cvi420F_3 (F_2 10-6)/Col-0	1-3	BC_1F_1	N/A	C	C	C	C	H	H	H	C	C	C	H	C	C	C	C	C	C

Parent line	Individual	Generation	cM	2-132	2-6789	2-8799	2-10540	2-10870	2-12121	2-12399	2-12618	2-12768	2-13157	2-13188	2-13195	2-13235	2-13427	2-13521	2-14176	2-16339
Cvi420F_3 (F_2 10-6)/Col 1-3	BP 6-2	BC_1F_2	24.82	V	V	V	V	H	H	H	H	H	H	H	H	H	H	H	C	C
Cvi420F_3 (F_2 10-6)/Col 1-3	BP 14-2	BC_1F_2	24.48	H	C	C	C	C	C	C	C	C	C	C	H	H	H	H	H	H
Cvi420F_3 (F_2 10-6)/Col 1-3	BP 66-6	BC_1F_2	31.22	C	C	C	C	C	C	H	H	H	H	H	H	H	H	H	H	H
BP 14-2	13-4	BC_1F_3	28.64	C	C	C	C	C	C	C	C	C	C	C	H	H	H	H	H	H
BP 14-2	15-4	BC_1F_3	26.66	C	C	C	C	C	C	C	C	C	C	C	H	H	H	C	C	C
BP 14-2 13-4	4-2	BC_1F_4	28.33	C	C	C	C	C	C	C	C	C	C	C	H	H	C	C	C	C

smaller BC_1F_3 mapping populations (Figures 26 and 27, Lines BP6-2, BP14-2 and BP66-6). After removal of lines without beneficial recombination events, an additional 70 BC_1F_3 lines were scored (Table 23). Again, additional informative lines were identified (Figure 27) and two were taken forward to generate final BC_1F_4 mapping populations (BC_1F_3 lines Cvi*420*BP 14-2 13-4 and Cvi*420*BP 14-2 15-4, Figure 26) and an additional 67 useful BC_1F_4 recombinant lines were scored (Table 24).

Genotyping and recombination data from these additional 301 fine-mapping lines were combined with Cvi-0/*420* F_2 population data to generate a population with additional power to resolve the QTL. This information was utilised to create a QTL map to refine the $rQTL2^{420}$ peak on chromosome 2 using R/qtl (Broman *et al*. 2003; Arends *et al*. 2010. Figure 28A). The peak marker from this population demonstrated that after the addition of fine-mapping individuals, the QTL still behaved semi-dominantly with the Cvi-0 genotype associating with higher recombination (Figure 28B). However, significant LOD scores for chromosome 2 markers from this population still indicate a broad peak of approximately 1.4 Mb (Threshold LOD 2.78, Alpha level 0.05, 1000 permutations) (Table 25), as sufficient recombination events were not generated within this region to refine the interval further. While this interval size is substantially reduced from that indicated in the F_2 population of over 5 Mb, it still contains hundreds of genes, preventing immediate identification of candidate genes.

Analysis of genotype patterns across chromosome 2 from specific fine-mapping individuals revealed additional information (Figure 27). Line Cvi*420*BP14-2 13-4 (a BC_1F_3 line) is fixed in the Col-0 genotype along chromosome 2 up to marker '13188' (position 13,188,597 bp), yet shows a high recombination phenotype of 28.64 cM, above the Col-0/*420* F_1 average of 18 cM (59% increase relative to Col-0/*420* F_1), suggesting that $rQTL2^{420}$ is located distal to marker '13188'. Similarly, line Cvi*420*BP14-2 15-4 (a BC_1F_3 line) is fixed for the Col-0 genotype along chromosome 2 up to marker '13188' and from marker '13521' (position 13,521,040 bp), yet also has a high recombination phenotype of 26.66 cM (48% increase relative to Col-0/*420* F_1), suggesting that $rQTL2^{420}$ is between markers '13188' and '13521'. Line Cvi*420*BP14-2 13-4 4-2 (BC_1F_4, Figure 26) has a high *420* recombination frequency of 28.3 cM (57% increase relative to Col-0/*420* F_1), and is fixed in the Col-0 genotype along chromosome 2 up to marker '13188' and from marker '13427' (position 13,427,779 bp), suggesting that $rQTL2^{420}$ is between markers '13188' and '13427'. As this interval contained 87 loci (Salk 1001 genomes browser, Alonso-Blanco *et al*. 2016; TAIR seqviewer found at: https://seqviewer.arabidopsis.org/), a number that could subsequently be reduced further through analysis of gene expression, polymorphisms between parental accessions and gene function, thereby producing a short-list of potential candidates, no further fine-mapping populations for $rQTL2^{420}$ were generated.

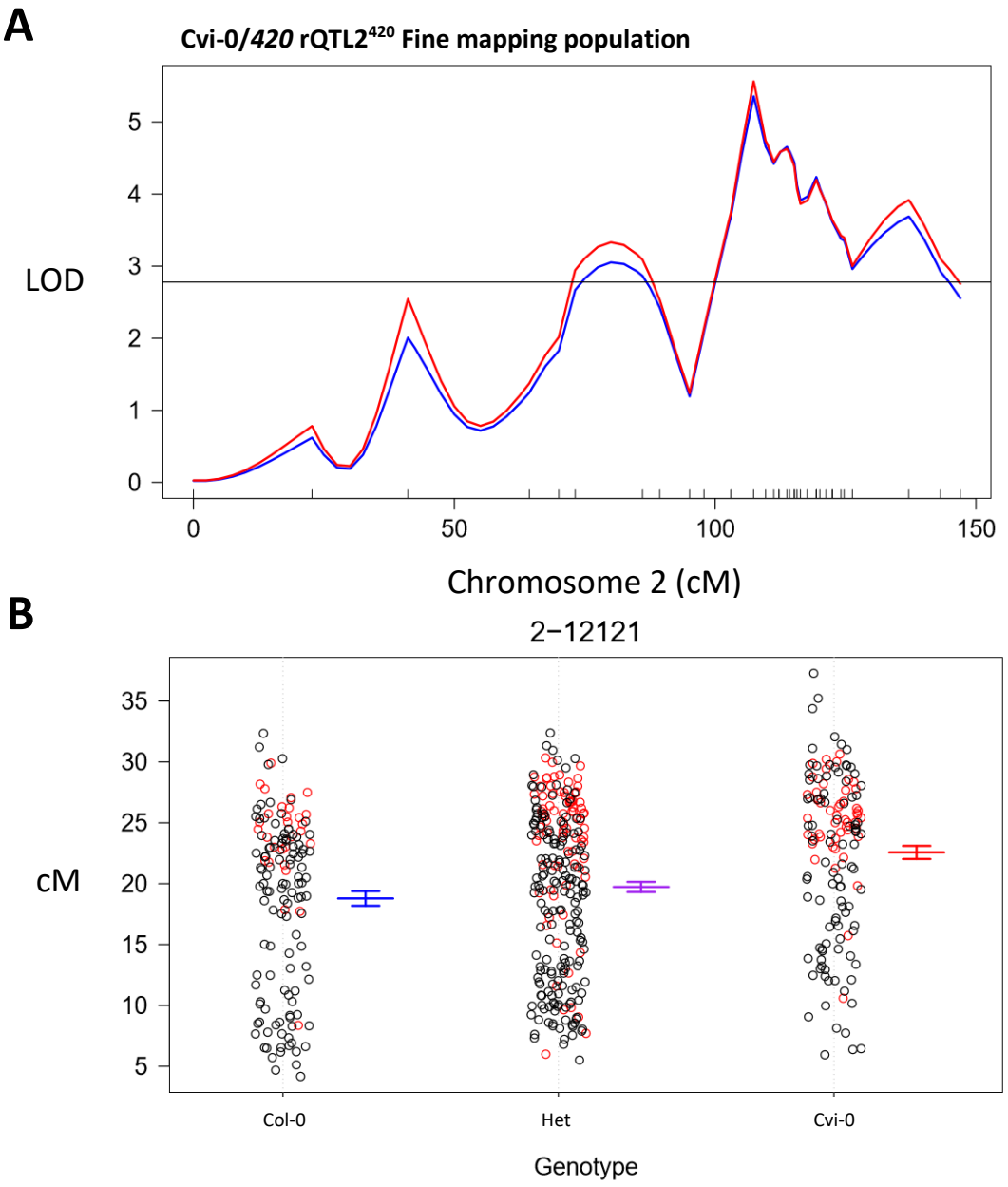


Figure 28. Association of recombination phenotype with marker genotype in the Cvi-0/*420* rQTL2^{420} fine mapping population and rQTL2^{420} effect plot by allele. **A.** Logarithm of odds (LOD) scores for association between *420* recombination frequency and genotyping markers in Cvi-0/*420* rQTL2^{420} fine mapping population. Red line denotes association through Hayley-Knott regression, blue line denotes mapping using an expectation-maximisation algorithm (R/qtl). Marker positions indicated by tick marks on the genetic map of chromosome 2 (cM) on the x-axis. 95% significance threshold indicated by horizontal black line. **B**. Effect plot showing recombination frequency (cM) of individuals with Col/Col, Col/Cvi or Cvi/Cvi genotypes at the marker (12,121,783bp) most strongly associated with recombination for rQTL2^{420} in the fine mapping population. Mean (±SEM) given for each genotype. Col-0 19.1cM (± 0.62); Het 20.1cM (± 0.48); Cvi-0 22.9cM (± 0.66). Differences between Col-0 and Het genotypes, Col-0 and Cvi-0 genotypes, and Het and Cvi-0 genotypes were significant (Two sample t-Test assuming unequal variance, p=0.0229, p=3.47 x 10^{-5} and p=5.13 x 10^{-4} respectively). Missing genotypes (shown in red) filled in by a random imputation that is conditional on the flanking marker genotypes.

4.4 Confirmation of rQTL2^{420} effect and preliminary analysis of region-specificity

To confirm the presence of rQTL2^{420}, a NIL (BC_1F_5) was generated from self-fertilisation of the BP14-2 13-4 4-2 line (Figure 26). This BC_1F_5 line showed Col-0 genotype at all markers tested across chromosomes 1,3,4, and 5 and across chromosome 2 up to marker '13188', and from marker '13427'. The interval between these markers was fixed in the Cvi-0 genotype (Figure 29). This line was then self-fertilised to generate replicates of the rQTL2^{420}NIL (BC_1F_6) to confirm the effect of rQTL2^{420} and its position within this interval. Replicates of rQTL2^{420}NIL had a mean *420* recombination rate of 26.2cM, significantly higher than the Col-0/*420* F_1 controls replicate mean (19.5cM)($\chi^2(1) = 119.1$, $p=2 \times 10^{-27}$, Table 26, Figure 30). This confirms that rQTL2^{420} is located between markers '13188' and '13427', and that the Cvi-0 allele causes an average increase of 6.7cM in the *420* interval, which is a 34% increase in recombination relative to Col-0/*420* F_1 controls.

While rQTL2^{420} has a confirmed effect on recombination in the subtelomeric *420* interval, it is unclear whether this effect can also be observed in other regions of the genome. A QTL on the same arm of chromosome 2, where the Cvi-0 allele was also associated with higher recombination, was detected in the Cvi-0/*CEN3* F_2 mapping population, suggesting that rQTL2^{420} may also affect recombination near the centromere (Figure 17C and Figure 18). However, rQTL2^{420} was not detected in the Cvi-0/*5.11* F_2 mapping population (Figure 25C) which also measures recombination across a centromeric interval. This could be due to a differential effect between chromosomes due to different centromeric *cis*-contexts such as repeat number or level of heterochromatin (Ito *et al*. 2007), or it could be due to *5.11* being a sex-averaged seed recombination estimate compared to *CEN3* which is a pollen-FTL. To determine whether rQTL2^{420} affects centromeric recombination, rQTL2^{420}NIL was crossed to the *5.11* and *CEN3* FTLs. Unfortunately, technical issues with growth conditions during the generation of rQTL2^{420}NIL/*CEN3* lines prevented data being collected for the *CEN3* interval, and also prevented the generation of F_2 lines for either interval with rQTL2^{420} in a homozygous Cvi-0 genotype. However, as the Cvi-0 allele of rQTL2^{420} appeared to act semi-dominantly, effects on recombination were predicted to be observable in the F_1 lines. Analysis of rQTL2^{420}NIL/*5.11* F_1 lines demonstrated no significant difference from Col-0/*5.11* F_1 controls ($\chi^2(1) = 4.038$, $p=0.089$, Table 27 and Figure 30). This limited data from a single interval suggests that if rQTL2^{420} does affect recombination at the centromere, it has either a chromosome-specific or sex-specific effect. Further work analysing additional intervals must be performed to confirm this effect.

4.5 Identification and analysis of candidate loci for rQTL2^{420}

rQTL2^{420}NIL crossover measurements demonstrate the presence of a recombination modifier between markers '13188' and '13427' on chromosome 2. This interval contains 87 genes (Salk Genome

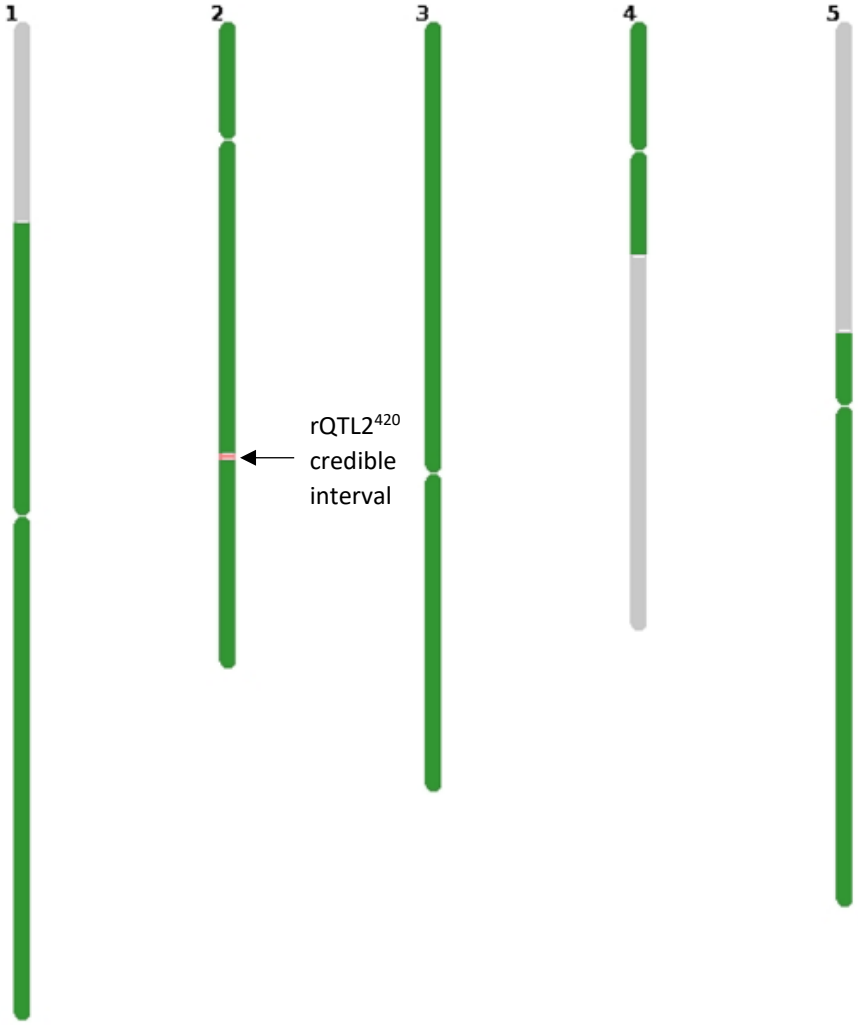


Figure 29: Chromosome map showing the genotype of the rQTL2^{420}NIL. Col-0 genotype is shown in green, Cvi-0 genotype in pink. Grey areas show regions of the genome where appropriate markers were not generated and therefore the genotype is unknown. Generated using TAIR Chromosome map tool.

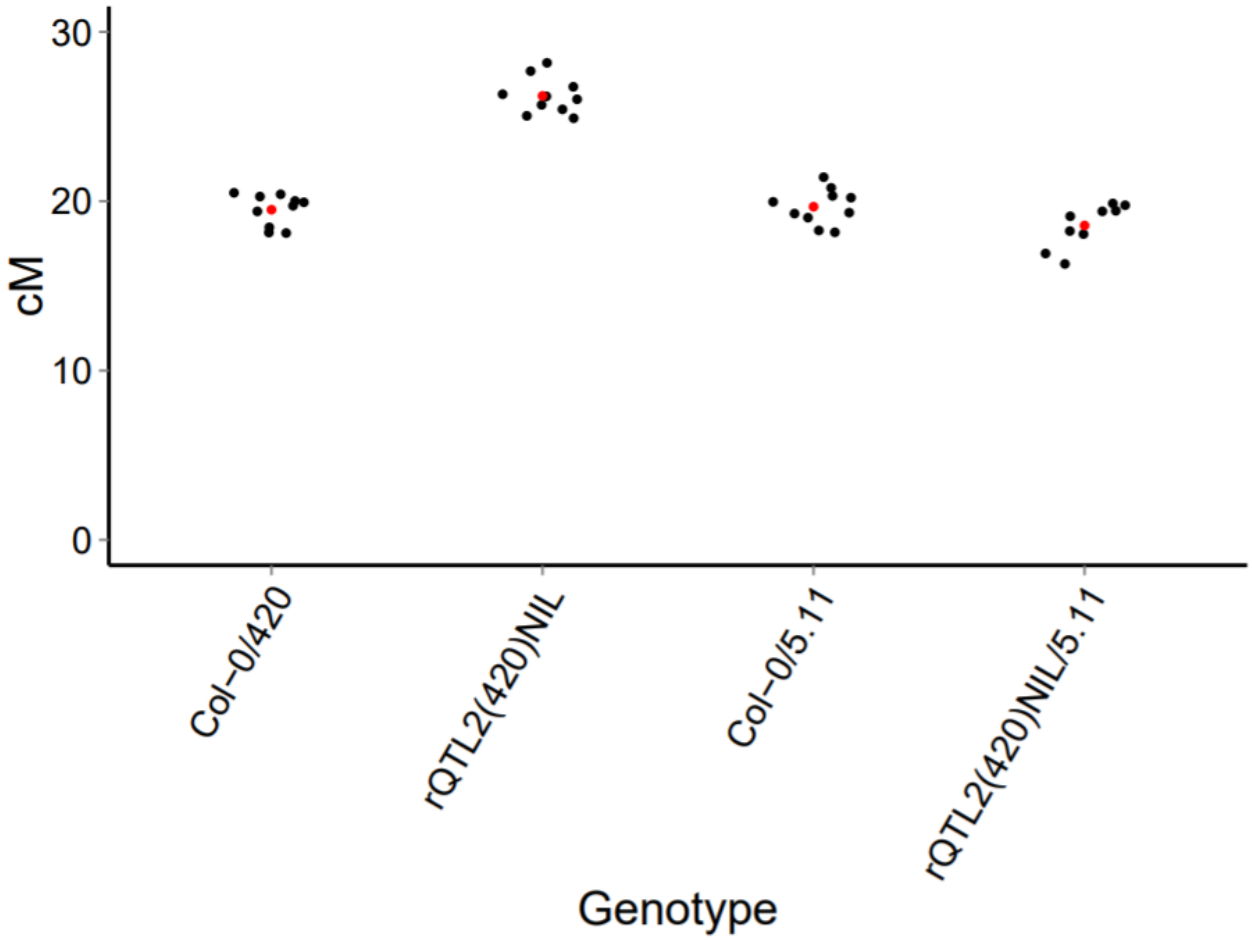


Figure 30: Genetic distance (cM) as measured in *420* and *5.11* seed-based FTL-intervals. Measurements were made in Col-0/*420* and Col-0/*5.11* F_1 individuals, and in rQTL2^{420}NIL and rQTL2^{420}NIL/*5.11* individuals (Tables 26 and 27: Total 29281 and 28820 seeds respectively). Replicates are in black, with the mean of each genotype highlighted in red. Differences between Col-0 lines and rQTL2^{420}NILs are significant for the *420* interval ($\chi^2(1) = 119.1$, $p=2 \times 10^{-27}$) but not the *5.11* interval ($\chi^2(1) =4.038$, $p=0.089$).

Express Browser, Alonso-Blanco *et al*. 2016), and information from The Arabidopsis Information Resource (TAIR, Berardini *et al*. 2015) for loci within the interval showed that many of these gene products are predicted to localise to the chloroplast, mitochondria or other non-nuclear cellular component and were therefore considered less likely to affect recombination in the nucleus. Furthermore, gene expression patterns (Arabidopsis Information Portal Araport11, Krishnakumar *et al*. 2015) demonstrated that some loci in this interval were not expressed during flower development. As meiosis-specific gene expression information is not yet available for most genes, due to difficulty in separating meiocytes in Arabidopsis from surrounding somatic tissue, floral expression was used as a proxy for expression during meiosis. Elimination of these loci reduced the candidate list to 39 genes encoding nuclear proteins with floral expression.

Sequence data from the 1,001 Genomes Project (Alonso-Blanco *et al*. 2016) was then used to identify candidate genes from this list containing polymorphisms between Col-0 and Cvi-0 accessions. Non-polymorphic (no sequence polymorphisms identified within the gene or predicted promoter/terminator regions) genes were removed from the candidate list. However, as the Cvi-0 and Col-0 accessions have extremely divergent genomes (Alonso-Blanco *et al*. 2016), most loci contain polymorphisms between the two accessions. Yet, many of these polymorphisms are shared between the Cvi-0 and Can-0 genomes. As rQTL2^{420} was not identified in the Can-0/*420* F_2 mapping population, the causal polymorphism was assumed not to be shared between the Cvi-0 and Can-0 accessions. Subsequent elimination of genes whose polymorphisms were shared between Cvi-0 and Can-0 further reduced the candidate list to 30 genes. However, it was still possible that the causal variant underlying the QTL could be shared between Can-0 and Cvi-0 accessions, as the effect on recombination phenotype may not manifest in Can-0 due to interaction of the modifier with other factors in the Can-0 genetic background such as potential *cis*-acting features, thereby preventing detection of the variant as a QTL. To ensure that viable candidates were not overlooked, although an interaction would be unlikely to completely negate any observable effect on recombination, consequently making the genes improbable candidates, these eliminated candidate genes were placed on a secondary list for consideration in the event that modifiers were not confirmed from the primary candidate list.

Additional removal of loci where polymorphisms were synonymous amino-acid substitutions outside potential regulatory elements left 17 loci on the primary candidate list. While it is possible that these synonymous mutations could affect protein translation or other processes (Goymer 2007), analysis of genes with non-synonymous mutations or polymorphisms in regulatory elements was given priority, as the probability of them having a significant effect on recombination was higher (Sandor *et al*. 2012; Brachi *et al*. 2010; Marais *et al*. 2003).

Specific candidates were then prioritised for analysis based on gene functions that suggested a potential role in recombination, for example DNA-binding activity (*At2g31210*) or a putative role in DNA repair (*At2g31320*), or an unknown function (Salk 1001 genomes browser, Alonso-Blanco *et al*. 2016; TAIR seqviewer found at: https://seqviewer.arabidopsis.org/). This resulted in 4 candidates being removed from the primary candidate list as they had predicted metabolic functions, leaving 13 potential candidates. As recombination is indirectly affected by a variety of factors, including broad influences such as temperature (Francis *et al*. 2007; Berkemeier and Linnert 1987; Stern 1926), the 4 genes with functions that did not appear to directly pertain to recombination were placed on a secondary candidate list for later consideration, as it is possible that they could have an indirect effect on crossovers. Therefore, while these criteria for candidate selection did leave some room for error, they served to create a viable primary candidate list that was short enough to easily assess using experimental methods, while leaving open the option for later consideration of alternative candidates if primary candidates were not confirmed.

While reciprocal transformation of accession-specific alleles is a more conclusive test of candidate loci, the cloning and transformation steps are laborious to complete for multiple candidate loci. Therefore, to identify the gene underlying rQTL2^{420}, analysis of T-DNA insertion mutants in the Col-0 genetic background was first performed (Alonso *et al*. 2003). T-DNA lines disrupting candidate genes were identified using the Salk T-DNA Express tool (The Salk Institute Genomic Analysis Laboratory) and ordered from the Nottingham Arabidopsis Stock Centre (NASC, Table 2). Where possible, lines were chosen with insertions close to the candidate polymorphisms. Lines were reciprocally crossed to the *420* FTL reporter line as both the male and female parent, and F_1 lines were measured for *420* recombination (Figure 31). As no significant differences were detected between crosses utilising the mutant as the male or female parent for any lines, reciprocal F_1 line data was pooled and analysis was performed comparing each mutant F_1 to Col-0/*420* F_1 controls.

T-DNA mutant/*420* F_1 lines were analysed for 10 candidate loci (Table 28) as, of the 13 loci on the primary candidate list, one candidate had no available insertion lines, and T-DNA insertion lines for an additional two candidates exhibited distorted segregation of the mutant allele that prevented generation of sufficient F_1 lines for analysis (data not shown). As T-DNA insertion mutant alleles may not have identical effects on gene function compared to the accession specific alleles underlying rQTL2^{420}, the phenotype of the F_1 lines was not necessarily expected to match the observed rQTL2^{420}NIL recombination rate. Therefore, observation of a deviation from wild-type Col-0/*420* F_1 recombination rates in either direction was tested for. Subsequent reciprocal transformation of accession alleles would confirm whether the candidate was the locus underlying rQTL2^{420}. This

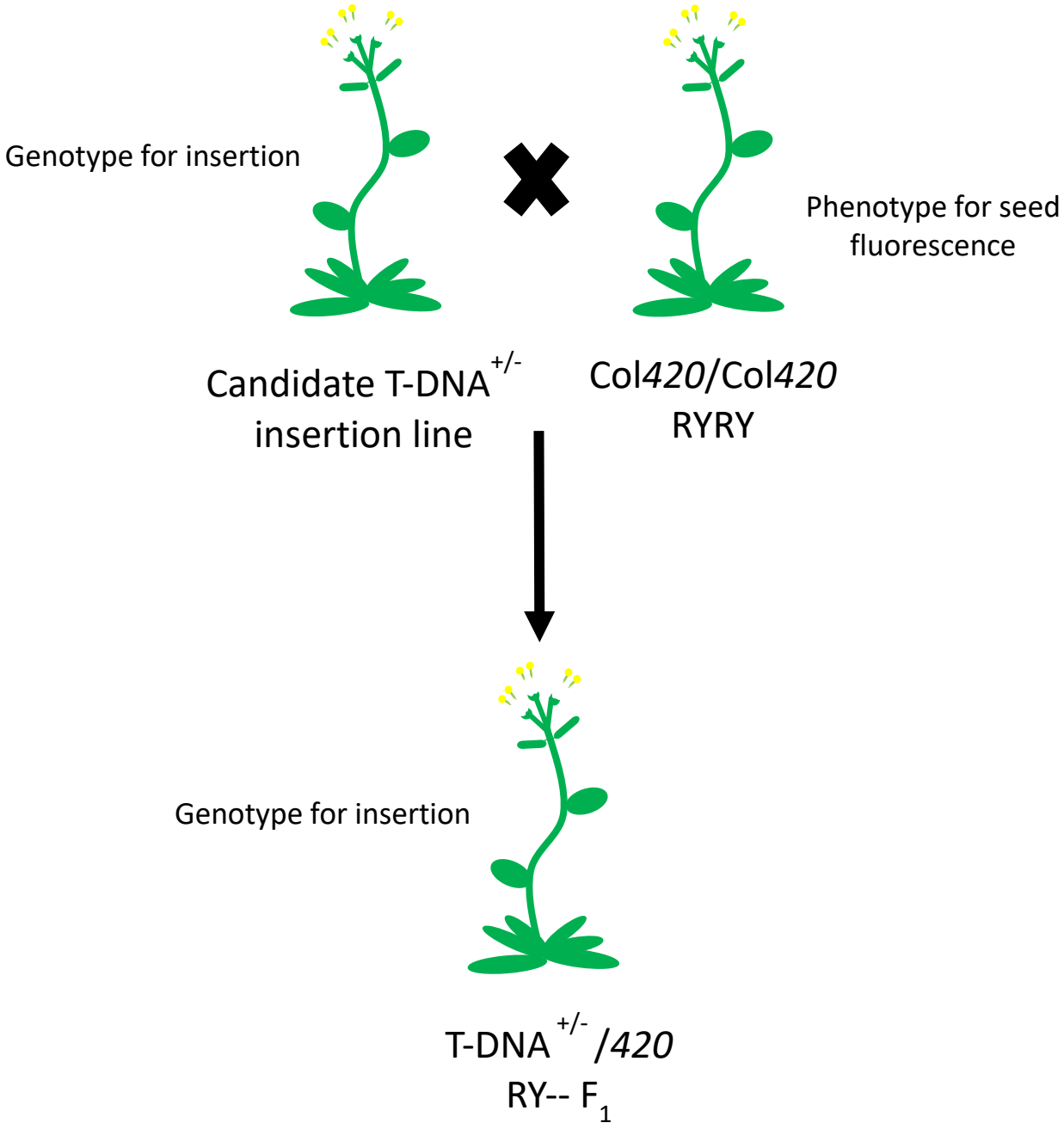


Figure 31: Schematic of candidate T-DNA crossing scheme. Hemizygous T-DNA insertion mutants were identified by PCR-based genotyping (see Materials and Methods) and crossed to a Col*420* FTL line fixed for both seed-fluorescence transgenes. F_1 progeny were genotyped to confirm presence of the insertion before being measured for *420* recombination frequency. The black arrow represents generation of new seedlings. The black cross represents a cross-fertilisation.

strategy offers the potential to identify multiple loci affecting the recombination phenotype, in the case that the QTL peak corresponds to several tightly linked loci.

Twelve T-DNA/*420* F_1 crosses were analysed, corresponding to one mutant line each for genes *At2g31010, At2g31210, At2g31270, At2g31450, At2g31070, At2g31130, At2g31370* and *At2g31410*, and two mutant lines for genes *At2g31320* and *At2g31510* based on insertion line availability and observed mutant-allele segregation patterns (Table 29, Figure 32).

Measurement of recombination in T-DNA/*420* F_1 lines revealed that eleven out of the twelve lines analysed did not have a significant effect on *420* recombination when compared to control Col-0/*420* F_1 lines (Two-tailed χ^2 tests, see Table 29 for p-values). However, the Salk 082541/*420* cross F_1 line, with an insertion in a central exon of the *ARIADNE7* (*ARI7*) gene *At2g31510* showed a significant decrease in recombination compared to Col-0/*420* F_1 replicates ($\chi^2(1) = 44.0666$, $p=6.35 \times 10^{-11}$). The mean recombination rate of Salk 082541/*420* F_1 replicates was 15.4 cM, compared to 18.5 cM in the growth-matched Col-0/*420* F_1 controls (17% decrease relative to Col-0/*420* F_1, Figure 32). Interestingly a second T-DNA line with an insertion in *At2g31510*, Salk 027620/*420* F_1, did not show a recombination rate that was significantly different from Col-0/*420* F_1 controls ($\chi^2(1) = 0.264647$, $p=0.6069$). This line had been selected due to the insertions proximity to a Cvi-0 specific polymorphism in the last intron but it is possible that this insertion did not disrupt the gene sufficiently to cause a change in recombination.

ARI7 (*At2g31510*) is a RING/U-box superfamily protein that has a predicted nucleic acid binding function (TAIR, Berardini *et al*. 2015), and zinc-finger and RING-type protein domains. It is expressed in a wide range of plant structures and growth stages, including flowers, inflorescence meristems and plant embryos (Araport11, Krishnakumar *et al*. 2015). Homology with ARIADNE proteins from other organisms suggests an interaction with the ubiquitin ligase complex and a role in protein ubiquitination (Mladek *et al*. 2003). Additionally, the presence of a nuclear localisation sequence suggests that the ARI7 protein is located within the nucleus, in close proximity to the genetic material (Mladek *et al*. 2003). While an obvious role in recombination is not immediately apparent, the complexity of the recombination process means that ARI7 could have an indirect effect, perhaps through modulation of the stability of recombination proteins, as it is predicted to be involved in protein ubiquitination which often marks proteins for degradation by proteasomes. The Cvi-0 specific polymorphisms in *ARI7* identified by the 1,001 Genomes Project (Alonso-Blanco *et al*. 2016) are within the first and last introns of the gene. Interestingly, an alternative splice variant of the gene (*At2g31510.2,* Salk 1001 genomes browser, Alonso-Blanco *et al.* 2016) removes the final exon due to splicing in the last intron. Hence, it is possible that the Cvi-0 specific polymorphism in the final intron

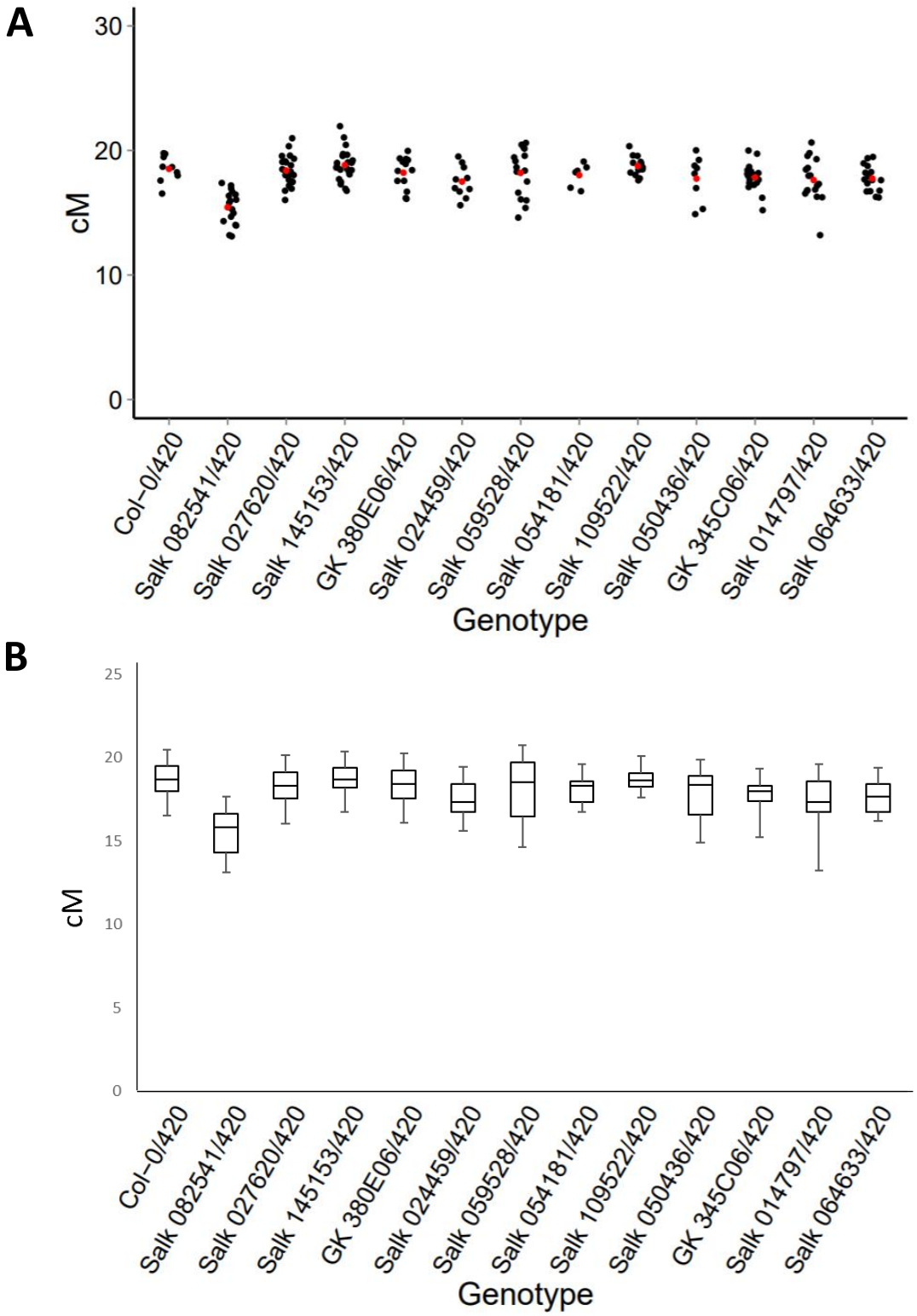


Figure 32: Genetic distance (cM) as measured in the *420* seed-based FTL-interval for rQTL2^{420} T-DNA insertion candidate lines. Measurements were made in Col-0/*420* and T-DNA/*420* F_1 individuals (Table 29: Total 279137 seeds). **A**. Dotplot. Replicates are in black, with the mean of each genotype highlighted in red. **B.** Same data represented in boxplots to highlight differences between genotypes. Interquartile range represented by box, median represented by midline. Minimum and maximum phenotype values for each genotype represented by whiskers. Differences between Col-0/*420* controls and T-DNA/*420* lines are only significant for the Salk 082541/*420* cross ($\chi^2(1) = 44.0666$, $p=6.35 \times 10^{-11}$). (See Table 28 for candidate loci information and Table 2 for T-DNA lines and NASC ID).

affects the splice junction, giving the polymorphism a potential regulatory function. This speculation would need to be confirmed through comparison of *At2g31510* mRNA sequences in Cvi-0 and Col-0 genetic backgrounds. Therefore, while a potential candidate has been identified for rQTL2^{420}, the causal polymorphism and mechanism of action remain unclear.

While the data from Salk 082541/*420* F_1 lines suggests that *ARI7* may act as a modifier of meiotic recombination in the *420* interval, the lack of effect observed in the Salk 027620/*420* F_1 lines makes the situation less clear and additional work analysing expression level, protein location and splice variants in these lines needs to be performed. Additionally, these F_1 measurements were made with the T-DNA insertion in a hemizygous state. While the Cvi-0 allele of rQTL2^{420} appears to behave semi-dominantly, a T-DNA insertion allele may not behave the same way. The Salk 027620 T-DNA insertion appears to have no effect on recombination while in a hemizygous state, but it is unclear whether this is because the insertion has no effect on the phenotype, or because the insertion allele is not dosage sensitive and the remaining Col-0 wild-type allele is sufficient to compensate for recombination function. To clarify this, and to quantify the effect of the Salk 082541 T-DNA insertion on recombination in the *420* interval, measurements need to be performed in lines containing the T-DNA insertions in a homozygous state. Furthermore, observation of a significant effect in the T-DNA insertion mutant does not confirm that *ARI7* is the gene underlying rQTL2^{420}. Reciprocal transformation of Col-0 and Cvi-0 alleles must be performed in future to confirm allele-specific effects, before further work into characterising a mechanism of action can be performed.

4.6 Fine-mapping rQTL5^{420}

A second potentially novel significant modifier of recombination was identified on chromosome 5 in the Cvi-0/*420* F_2 population (LOD 19.72, F-test p=6.33 x 10^{-14}), explaining 15% of the variance in the population (Figure 20C). This modifier also behaved semi-dominantly, with heterozygous individuals exhibiting an intermediate phenotype, although in this case the Col-0 allele was associated with higher recombination (Figure 21C). To identify the locus underlying rQTL5^{420}, a strategy utilising backcrossing into Col-0 with MAS was performed.

The F_2 population had identified a broad QTL peak spanning approximately 4Mb across chromosome 5, but further recombination events in this region were required in order to refine this interval. Additionally, the effects of rQTL1^{420}, rQTL2^{420} and possible small effect loci were segregating in the population (Figure 20C), so backcrossing to remove these effects and any potential *cis*-acting heterozygosity influence was desirable. An individual from the F_2 population (line 24-4) was selected and self-fertilised to generate an F_3 line which was backcrossed to Col-0. MAS was used to identify

progeny lines with a predominantly Col-0 genetic background, with the exception of the rQTL5^{420} region, to be taken forward for the generation of fine-mapping populations (Figure 33).

The line Cvi*420*F_3(F_2 24-4)/Col 1-6 (BC_1F_1, Figure 34) was identified that had chromosomes 2, 3, 4 and most of chromosome 5 in the Col-0 genotype, although chromosome 1 was still segregating (Figure 33). Self-fertilisation of this line generated a small BC_1F_2 fine-mapping population where screening identified 18 individuals with informative recombination events, which were then scored for recombination in the *420* interval (Table 30). A line from this population with low recombination frequency of 13.23 cM (26% decrease relative to Col-0/*420* F_1) and a recombination event within the QTL region contained chromosome 5 in the Col-0 homozygous genotype from marker '22313' (position 22,313,360 bp) onwards (BC_1F_2 line Cvi*420*F_3(F_2 24-4)/Col 1-6 1-6, Figure 34). This line was self-fertilised to generate a line that was fixed in the Col-0 background across chromosomes 1, 2, 3 and 4, and fixed Col-0 across chromosome 5 up to marker '20437' (position 20,437,184 bp) and from marker '22313', with the region between the markers fixed in the Cvi-0 genotype (Figure 33, BC_1F_3 line Cvi*420*F_3(F_2 24-4)/Col 1-6 1-6 1-3, Figure 34). This line had a low recombination frequency of 13.96 cM (22% decrease relative to Col-0/*420* F_1), indicating that rQTL5^{420} is a locus between markers '20437' and '22313'. This rQTL5^{420}NIL (BC_1F_3) was then self-fertilised to generate replicate recombination measurements to quantify the effect of rQTL5^{420} and confirm its position within this interval (Table 26).

To generate additional recombinants a second F_2 line (individual 2-3) was self-fertilised, and the F_3 was backcrossed to Col-0 in parallel to generate a BC_1F_1 line that had the Col-0 genotype for all markers tested across chromosomes 1, 2, 3 and 4, and was heterozygous for all markers across chromosome 5 (Figure 33, line Cvi*420*F_3(F_2 2-3)/Col 1-5, Figure 34). Self-fertilisation of this line generated a BC_1F_2 fine-mapping population where genotyping identified 24 lines with informative recombination events in the region surrounding rQTL5^{420} that would narrow the interval size, that were then scored for *420* recombination frequency (Table 31). Two of these lines (Cvi*420*F_3(F_2 2-3)/Col 1-5 10-2 and Cvi*420*F_3(F_2 2-3)/Col 1-5 10-6, Figure 34) were taken forward to generate further BC_1F_3 mapping populations, and 30 additional recombinants were identified and scored (Figure 33 and Table 32).

Fine mapping data from these additional 72 lines was combined with Cvi-0/*420* F_2 population data to generate a population with increased mapping power to resolve the QTL. This information was used to create a QTL map to refine the rQTL5^{420} peak on chromosome 5 using R/qtl (Broman *et al*. 2003; Arends *et al*. 2010. Figure 35A). rQTL5^{420} still demonstrated semi-dominant behaviour with the Col-0 genotype associating with higher recombination (Figure 35B). However, as was observed for rQTL2^{420} fine mapping, the LOD scores across the QTL peak indicated a fairly broad interval with markers above

Figure 33. SSLP and CAPS genotyping data for lines generated for fine-mapping of rQTL5[420]. Parent line denotes the line that was self-fertilised to generate the individual, the Generation column denotes the generation that the individual line belongs to, and the cM column denotes the recombination measurement of the *420* FTL interval in the individual. Genotyping markers are indicated by a number representing their location in the genome (chromosome-kilobase) e.g. marker 1-6108 is on chromosome 1 at position 6108789bp. Col-0 genotype is indicated by the letter C, Cvi-0 genotype by the letter V and segregating markers by the letter H for each individual at each genotyping marker.

				Genotyping markers																
Parent line	Individual	Generation	cM	1-6108	1-16908	1-19540	1-30413	2-132	2-12121	2-16339	3-1031	3-4126	3-15949	3-19165	4-230	4-2611	4-7157	5-9437	5-19994	5-25212
Cvi-0/420 F_1	24-4	F_2	5.72	H	V	V	V	V	C	C	H	H	H	H	C	C	C	C	C	H
Cvi-0/420 F_1	2-3	F_2	N/A	C	C	C	C	H	C	C	H	H	V	V	C	C	C	C	H	H
Cvi420F_3 (F_2 24-4)/Col	1-6	BC_1F_1	12.68	C	H	H	H	H	C	C	C	C	C	C	C	C	C	C	C	H
Cvi420F_3 (F_2 2-3)/Col	1-5	BC_1F_1	17.37	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	H	H

Parent line	Individual	Generation	cM	1-19540	1-30413	2-132	2-16339	3-19165	5-19994	5-20312	5-20437	5-20685	5-20780	5-20899	5-21349	5-21562	5-21906	5-22313	5-22402	5-25212
Cvi420F_3 (F_2 24-4)/Col 1-6	1-6	BC_1F_2	13.23	H	H	C	C	C	C	C	C	H	H	H	H	H	H	C	C	C
Cvi420F_3 (F_2 24-4)/Col 1-6 1-6	1-3	BC_1F_3	13.96	C	C	C	C	C	C	C	C	V	V	V	V	V	V	C	C	C
Cvi420F_3 (F_2 2-3)/Col 1-5	10-2	BC_1F_2	18.29	C	C	C	C	C	V	V	H	H	H	H	H	H	H	H	H	H
Cvi420F_3 (F_2 2-3)/Col 1-5	10-6	BC_1F_2	18.00	C	C	C	C	C	H	H	H	H	H	H	H	V	V	V	V	V

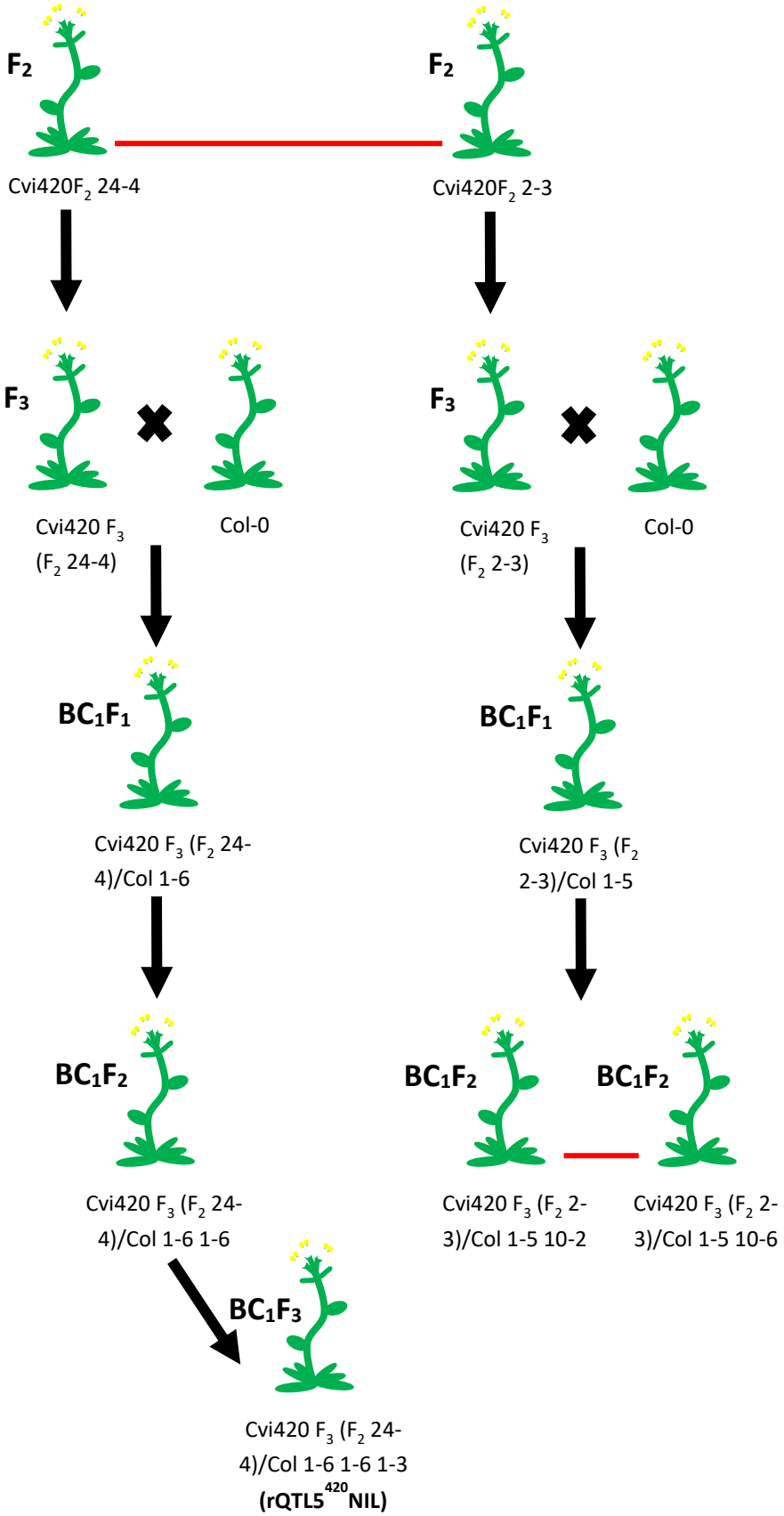


Figure 34: Schematic of rQTL5^{420} fine mapping plant pedigree. Plants were genotyped using accession-specific SSLP and CAPS markers (see Appendix) and measured for recombination in the *420* FTL interval. For genotyping see Figure 33 and 36. For scoring see Table 30, Table 31 and Table 32. Black arrows represent generation of new seedlings. Black crosses represent a cross-fertilisation. Sibling relationships are represented by red lines between plants.

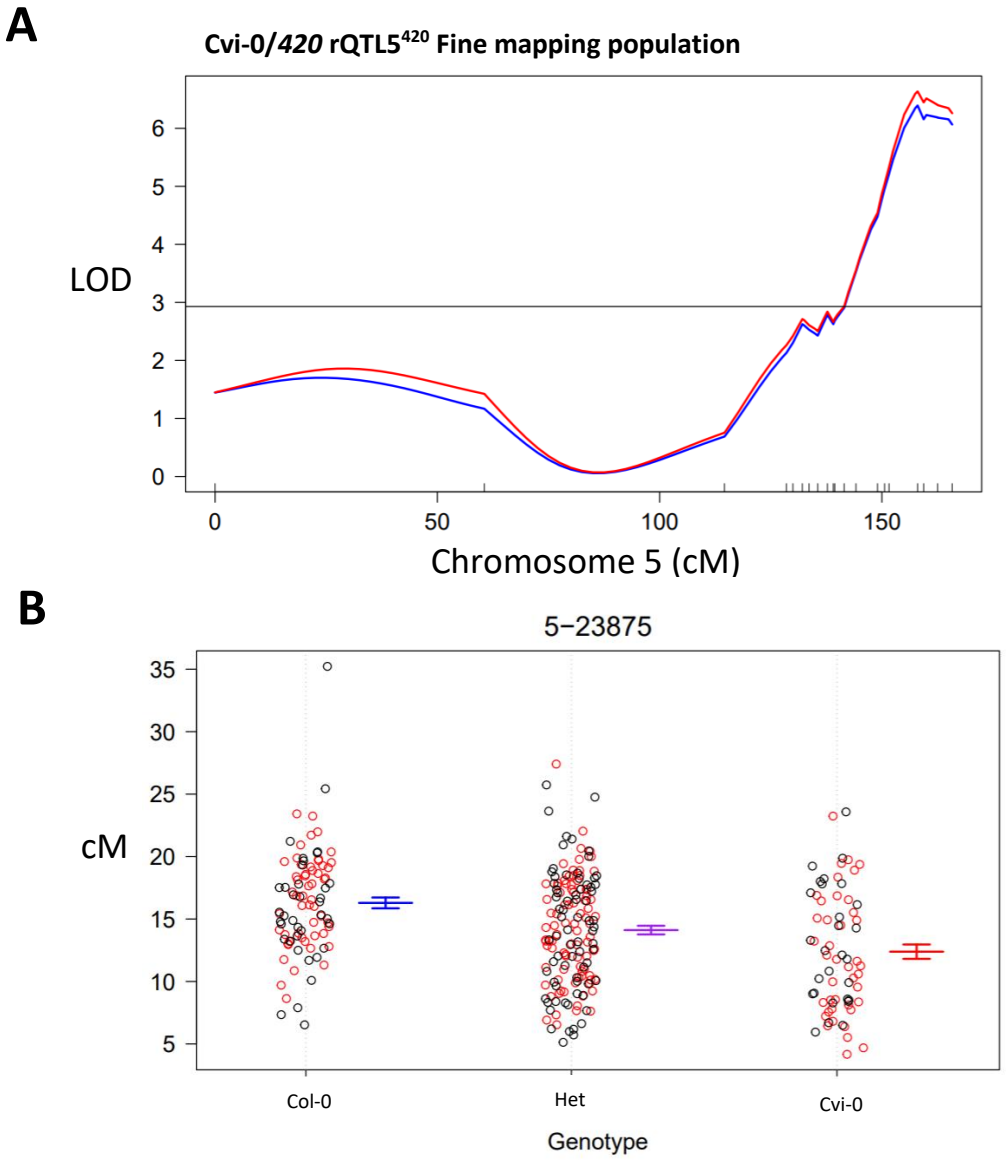


Figure 35. Association of recombination phenotype with marker genotype in the Cvi-0/*420* rQTL5^{420} fine mapping population and rQTL5^{420} effect plot by allele. **A.** Logarithm of odds (LOD) scores for association between *420* recombination frequency and genotyping markers in Cvi-0/*420* rQTL5^{420} fine mapping population. Red line denotes association through Hayley-Knott regression, blue line denotes mapping using an expectation-maximisation algorithm (R/qtl). Marker positions indicated by tick marks on the genetic map of chromosome 5 (cM) on the x-axis. 95% significance threshold indicated by horizontal black line. **B**. Effect plot showing recombination frequency (cM) of individuals with Col/Col, Col/Cvi or Cvi/Cvi genotypes at the marker (23,875,653 bp) most strongly associated with recombination for rQTL5^{420} in the fine mapping population. Mean (±SEM) given for each genotype. Col-0 16.2cM (± 0.49); Het 13.9cM (± 0.4); Cvi-0 12.2cM (± 0.58). Differences between Col-0 and Het genotypes, Col-0 and Cvi-0 genotypes, and Het and Cvi-0 genotypes were significant (Two sample t-Test assuming unequal variance, $p=4.85 \times 10^{-4}$, $p=9.03 \times 10^{-7}$ and $p=0.0207$ respectively). Missing genotypes (shown in red) filled in by a random imputation that is conditional on the flanking marker genotypes.

the LOD significance threshold (Threshold LOD 2.93, Alpha level 0.05, 1000 permutations) (Table 33). This mapping was not sufficient to define the credible region to below 4 Mb.

Again, while the rQTL5^{420} interval was reduced by fine-mapping recombinants, it still contained too many genes to perform candidate testing (Salk 1001 genomes browser, Alonso-Blanco *et al*. 2016; TAIR seqviewer; Glazier, Nadeau and Aitman 2002). Additional information was provided from the genotyping patterns of fine-mapping recombinants across chromosome 5 (Figure 36). Line Cvi*420*F$_3$(F$_2$ 2-3)/Col 1-5 10-6 11-3 (a BC$_1$F$_3$ line) is fixed in the Col-0 genotype along chromosome 5 up to marker '20899' (position 20,899,170 bp), yet still has a low recombination phenotype of 16.5 cM (8% decrease relative to Col-0/*420* F$_1$), suggesting that rQTL5^{420} must be present in the Cvi-0 or heterozygous genotype and therefore in the region of chromosome 5 beyond marker '20899'. Similarly, the sibling line Cvi*420*F$_3$(F$_2$ 2-3)/Col 1-5 10-6 13-1 (a BC$_1$F$_3$ line) is fixed in the Col-0 genotype along chromosome 5 up to marker '21349' (position 21349815bp) and has a low recombination rate of 14.4 cM (20% decrease relative to Col-0/*420* F$_1$) indicating that rQTL5^{420} is past marker '21349'. BC$_1$F$_2$ lines Cvi*420*F$_3$(F$_2$ 2-3)/Col 1-5 2-4 and Cvi*420*F$_3$(F$_2$ 2-3)/Col 1-5 4-2 both exhibit 420 recombination rates that are comparable with the Col-0/*420* F$_1$ average of 18 cM, indicating that the Col-0 allele of rQTL5^{420} is present. These lines are fixed in the Col-0 genotype along chromosome 5 up to marker '21906' (position 21906244bp), suggesting that rQTL5^{420} must be in the region before marker '21906'. Finally, BC$_1$F$_2$ lines Cvi*420*F$_3$(F$_2$ 24-4)/Col 1-6 2-1 and 5-5 have low *420* recombination rates of 15.3 cM (15% decrease relative to Col-0/*420* F$_1$) and 14.1 cM (22% decrease relative to Col-0/*420* F$_1$) respectively, indicating that the Cvi-0 allele of rQTL5^{420} is present. Both of these lines are fixed in the Col-0 genotype along chromosome 5 up to marker '20437' and from marker '21562' (position 21,562,910 bp) onwards, suggesting that the rQTL5^{420} loci lies between these two markers (Figure 36). Combined information from these independent lines suggests that rQTL5^{420} is a locus between markers '21349' and '21562'. This region does not directly correspond to the peak marker identified by R/qtl (Figure 35A). However, the QTL association map was generated from all combined data across the genome, and is potentially influenced by the segregation of small effect loci, rQTL1^{420} and rQTL2^{420} in the F$_2$ individuals that were combined with fine-mapping lines to generate this map. This explains the slight peak shift observed in fine-mapping lines, where the effect of other loci is removed by backcrossing into the Col-0 background.

The interval between markers '21349' and '21562' was just 213 kb wide, and only contained 31 candidate genes (Salk 1001 genomes browser, Alonso-Blanco *et al*. 2016; TAIR seqviewer). This list of genes was examined, as previously, for floral gene expression and the presence of polymorphisms between Col-0 and Cvi-0 accessions. As this would generate a sufficiently short list of potential candidates, no further fine-mapping populations for rQTL5^{420} were generated.

Figure 36. SSLP and CAPS genotyping data for lines generated for further fine-mapping of rQTL5[420]. Parent line denotes the line that was self-fertilised to generate the individual, the Generation column denotes the generation that the individual line belongs to, and the cM column denotes the recombination measurement of the *420* FTL interval in the individual. Genotyping markers are indicated by a number representing their location in the genome (chromosome-kilobase) e.g. marker 1-6108 is on chromosome 1 at position 6108789bp. Col-0 genotype is indicated by the letter C, Cvi-0 genotype by the letter V and segregating markers by the letter H for each individual at each genotyping marker.

				Genotyping markers																	
Parent line	Individual	Generation	cM	1-19540	1-30413	2-132	2-16339	3-19165	5-19994	5-20312	5-20437	5-20685	5-20780	5-20899	5-21349	5-21562	5-21906	5-22313	5-22402	5-25212
Cvi420F$_3$ (F$_2$ 2-3)/Col 1-5	2-4	BC$_1$F$_2$	19.04	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	H	H
Cvi420F$_3$ (F$_2$ 2-3)/Col 1-5	4-2	BC$_1$F$_2$	20.94	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	H	H
Cvi420F$_3$ (F$_2$ 2-3)/Col 1-5 10-6	11-3	BC$_1$F$_3$	16.54	C	C	C	C	C	C	C	C	C	C	C	H	V	V	V	V	V
Cvi420F$_3$ (F$_2$ 2-3)/Col 1-5 10-6	13-1	BC$_1$F$_3$	14.39	C	C	C	C	C	C	C	C	C	C	C	C	V	V	V	V	V
Cvi420F$_3$ (F$_2$ 24-4)/Col 1-6	2-1	BC$_1$F$_2$	15.33	H	H	C	C	C	C	C	C	H	H	H	H	C	C	C	C	C
Cvi420F$_3$ (F$_2$ 24-4)/Col 1-6	5-5	BC$_1$F$_2$	14.07	V	V	C	C	C	C	C	C	H	H	H	H	C	C	C	C	C

4.7 Confirmation of rQTL5^{420} effect and preliminary analysis of region-specificity

To confirm the effect of rQTL5^{420} and its location on chromosome 5, BC_1F_4 replicates of the rQTL5^{420}NIL, generated from self-fertilisation of fine-mapping line Cvi*420*F_3(F_2 24-4)/Col 1-6 1-6 1-3 (BC_1F_3, Figure 34), were scored for recombination in the *420* interval. The rQTL5^{420}NIL showed Col-0 genotype at all markers tested across chromosomes 1, 2, 3 and 4 and across chromosome 5 up to marker '20437' and from marker '22313'. The interval between these markers was fixed in the Cvi-0 genotype (Figure 37). Replicates of the rQTL5^{420}NIL had a mean *420* recombination rate of 13.13 cM, significantly lower than the Col-0/*420* F_1 control replicate mean of 19.5 cM ($\chi^2(1) = 167.18$, $p=6.11 \times 10^{-38}$, Table 26, Figure 38). This confirms that rQTL5^{420} corresponds to a locus located between markers '20437' and '22313' on chromosome 5, and that the Cvi-0 allele causes an average decrease of 6.4 cM in the *420* interval, which is a 33% decrease in recombination relative to Col-0/*420* F_1 controls.

However, it is unclear whether rQTL5^{420} affects recombination generally across the genome, or if the effect is isolated to specific regions. The rQTL5^{420} peak marker is less than 1 Mb away from the peak marker identified for rQTL5^{CEN3} in the Cvi-0/*CEN3* F_2 mapping population, and in both cases the Col-0 allele is associated with higher recombination (Figures 18 and 21). It is possible that both of these QTL correspond to the same locus, thereby suggesting that rQTL5^{420} may also have an effect on recombination near the centromere. However, as was observed for rQTL2^{420}, although a similar QTL peak was detected in the Cvi-0/*CEN3* F_2 mapping population, it was not detected in the Cvi-0/*5.11* F_2 mapping population which also measures recombination across a centromeric interval (Figure 25C). This suggests that these modifiers may have chromosome-specific or sex-specific effects on recombination.

To determine whether rQTL5^{420} has an effect on centromeric recombination, the rQTL5^{420}NIL was crossed to the *5.11* and *CEN3* FTLs. As previously mentioned for analysis of the rQTL2^{420}NIL, technical issues resulted in the loss of NIL/*CEN3* lines, but sufficient data was collected for analysis of rQTL5^{420}NIL/*5.11* lines. The mean of rQTL5^{420}NIL/*5.11* replicates demonstrated a significant, though minor (2 cM), decrease in recombination when compared to Col-0/*5.11* F_1 controls (11% decrease relative to Col-0/*5.11* F_1, $\chi^2(1) = 12.62$, $p=7.63 \times 10^{-4}$, Table 27, Figure 38). While this data is from analysis of a single interval, it does suggest that rQTL5^{420} could be having an effect on recombination around centromeres. It is unclear why this effect was not observed in the Cvi-0/*5.11* F_2 population, although it is possible that the minor effect was masked by other segregating small effect loci and *cis*-acting factors. Further work analysing additional intervals must be performed to confirm this effect.

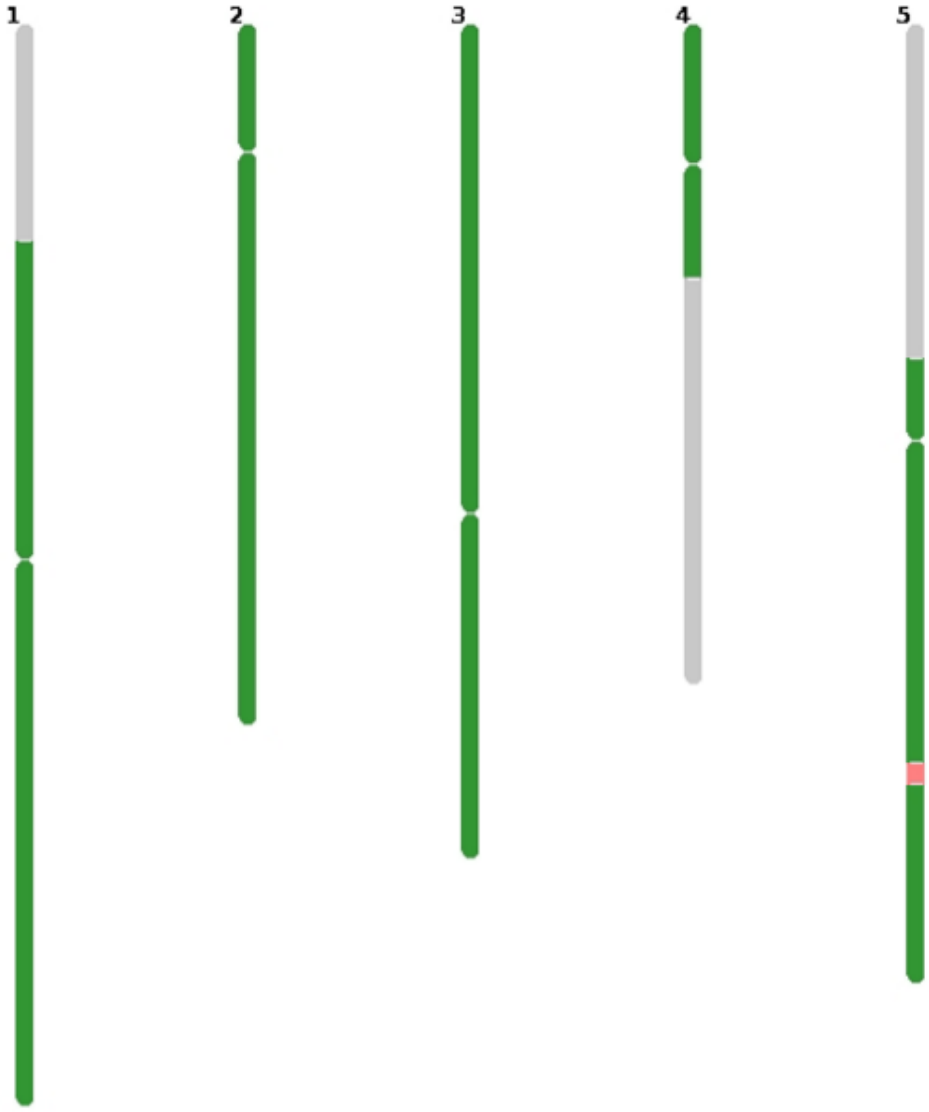


Figure 37: Chromosome map showing the genotype of the rQTL5^{420}NIL. Col-0 genotype is shown in green, Cvi-0 genotype in pink. Grey areas show regions of the genome where appropriate markers could not be generated and therefore the genotype is unknown. Generated using TAIR Chromosome map tool.

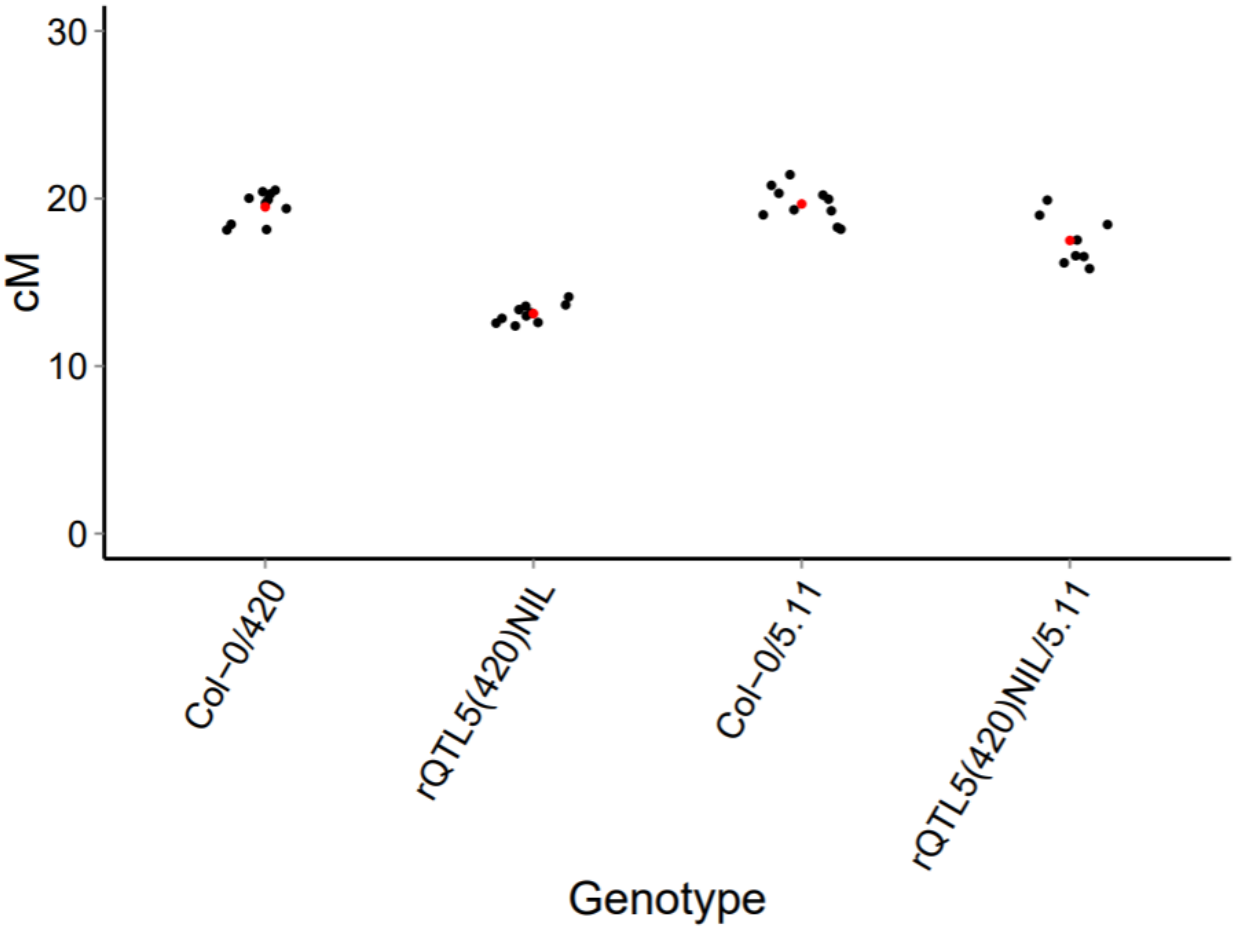


Figure 38: Genetic distance (cM) as measured in *420* and *5.11* seed-based FTL-intervals. Measurements were made in Col-0/*420* and Col-0/*5.11* F_1 individuals, and in rQTL5^{420}NIL and rQTL5^{420}NIL/*5.11* individuals (Tables 26 and 27: Total 31123 and 26962 seeds respectively). Replicates are in black, with the mean of each genotype highlighted in red. Differences between Col-0 lines and rQTL5^{420}NILs are significant for the *420* interval ($\chi^2(1) = 167.18$, $p = 6.11 \times 10^{-38}$) and the *5.11* interval ($\chi^2(1) = 12.62$, $p = 7.63 \times 10^{-4}$).

4.8 Identification and analysis of candidate loci for rQTL5^{420}

Analysis of the rQTL5^{420}NIL demonstrated the presence of a recombination modifier between markers '20437' and '22313' on chromosome 5 that has different alleles in the Cvi-0 and Col-0 genetic backgrounds. Additional information from fine-mapping individuals (Figure 36) narrowed this interval further to 213 kb between markers '21349' and '21562'. This interval contains 31 loci (Salk Genome Express Browser, Alonso-Blanco *et al*. 2016), a number that was subsequently narrowed down to generate a candidate list for analysis. Genes whose products were predicted to localise to the chloroplast, mitochondria or other non-nuclear components were removed from the list, as were genes whose expression patterns (Araport11, Krishnakumar *et al*. 2015) showed that they were not expressed during floral development. Attempts to remove loci that did not contain polymorphisms between Col-0 and Cvi-0 did not reduce the size of the candidate list, as most genes were polymorphic between the two accessions (Alonso-Blanco *et al*. 2016). However, as rQTL5^{420} was not identified in the Can-0/*420* F_2 population, this suggests that the allele is not shared between Cvi-0 and Can-0. Therefore, loci where polymorphisms between Cvi-0 and Col-0 were shared between Cvi-0 and Can-0 were removed from the candidate list, leaving 10 candidate loci. Additional removal of two retrotransposons and four loci with synonymous mutations further reduced the candidate list to 4 loci, all of which has plausible functional relevance to recombination, such as DNA-binding or regulatory functions (Table 34)(Salk 1001 genomes browser, Alonso-Blanco *et al*. 2016; TAIR seqviewer).

To determine which of these candidate loci was the gene underlying rQTL5^{420}, T-DNA insertion mutants disrupting each of these four loci were identified using the T-DNA Express Tool (The Salk Institute Genomic Analysis Laboratory) and ordered from NASC (Table 3). Mimicking the strategy used to identify the candidate gene underlying rQTL2^{420}, insertion mutants were reciprocally crossed to the *420* reporter line as both the male and female parent, and recombination was measured in F_1 lines. No significant differences were detected between F_1 lines with the mutant as the male or female parent, so reciprocal F_1 line data was pooled and each mutant F_1 was compared to Col-0/*420* F_1 controls. T-DNA insertion mutant alleles were not expected to affect recombination in exactly the same way as the accession-specific alleles underlying rQTL5^{420}, therefore deviation from wild-type Col-0/*420* F_1 recombination rates in either direction was tested for. Four T-DNA/*420* F_1 crosses were analysed, with one mutant line each for genes *At5g52800*, *At5g52830, At5g52880* and *At5g53140*. Lines were selected based on availability, observed mutant-allele segregation and proximity of insertion to Cvi-0 specific polymorphisms (Table 35, Figure 39).

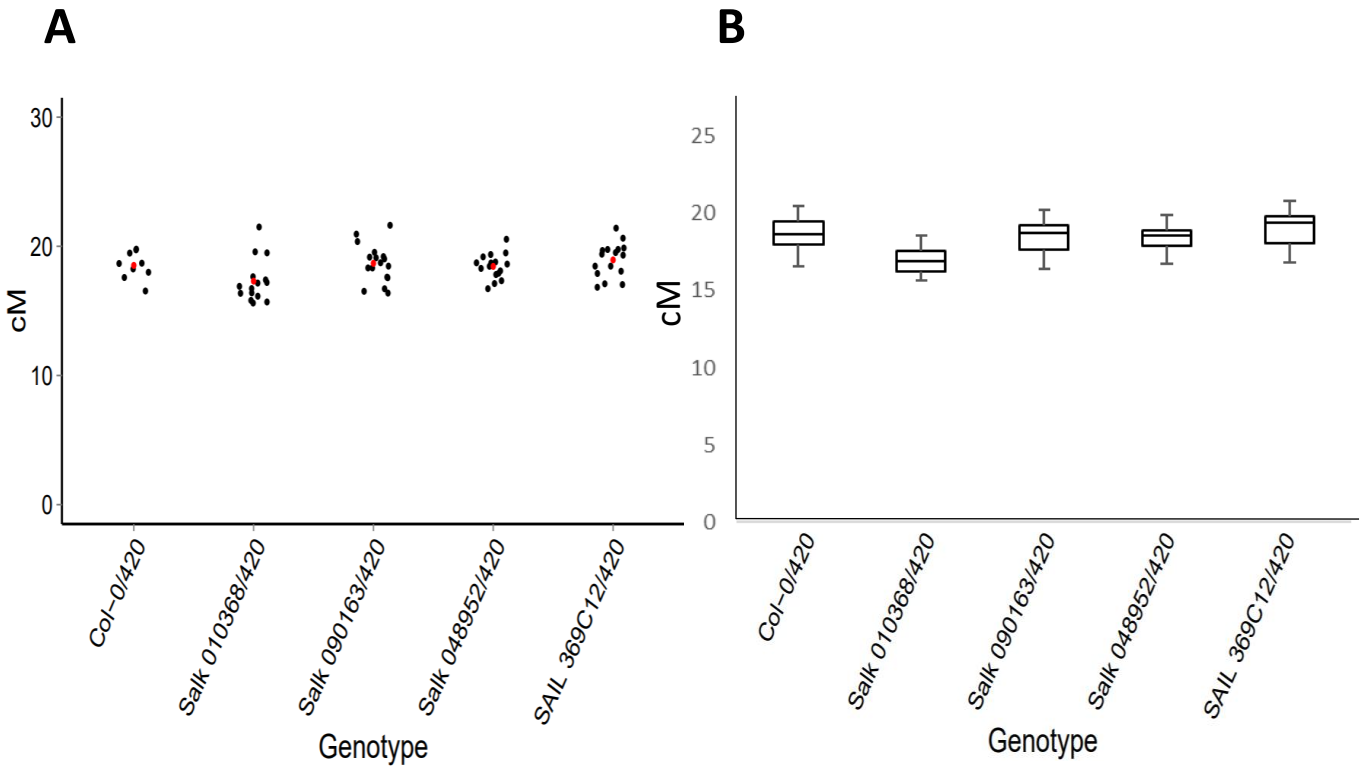


Figure 39: Genetic distance (cM) as measured in the *420* seed-based FTL-interval for rQTL5^{420} T-DNA insertion candidate lines. Measurements were made in Col-0/*420* and T-DNA/*420* F_1 individuals (Table 35: Total 107820 seeds). **A**. Dotplot. Replicates are in black, with the mean of each genotype highlighted in red. **B.** Same data represented in boxplots to highlight differences between genotypes. Interquartile range represented by box, median represented by midline. Minimum and maximum phenotype values for each genotype represented by whiskers. Differences between Col-0/*420* controls and T-DNA/*420* lines are only significant for the Salk 010368/*420* cross ($\chi^2(1) = 5.989$, p=0.0288). See Table 34 for candidate loci information and Table 3 for T-DNA lines and NASC ID.

Measurement of recombination in T-DNA/*420* F_1 lines revealed that three out of the four lines analysed did not have a significant effect on *420* recombination when compared to Col-0/*420* F_1 lines (Two-tailed χ^2 tests, see Table 35 for p-values). However, the Salk 010368/*420* F_1 lines, which contain an insertion upstream of gene *At5g53140* showed a significant, though minor, decrease in recombination compared to Col-0/*420* F_1 replicates ($\chi^2(1) = 5.989$, p=0.0288). The mean recombination rate of Salk 010368/*420* F_1 replicates was 17.3 cM, which was 1.2 cM lower than the growth-matched Col-0/*420* F_1 control mean of 18.5 cM (6% decrease relative to Col-0/*420* F_1, Figure 39). The Salk 010368 line was chosen as it was the closest available insertion to the Cvi-0 specific polymorphisms observed in the predicted promoter region of *At5g53140* (Alonso-Blanco *et al*. 2016). However, the insertion is further upstream than the putative promoter and so further work looking at the expression of *At5g53140* in the mutant line would need to be performed to confirm that the insertion is affecting this gene and not an alternative regulatory element. Additionally, as the observed effect on recombination is minor, analysis of additional insertion mutants in the *At5g53140* gene would be required to confirm this effect. Measurement of recombination in homozygous Salk 010368 T-DNA insertion lines could also clarify if the size of the effect is due to dominance or dosage effects. Reciprocal transformation of the Col-0 and Cvi-0 alleles of the gene and measurement of the effect on recombination in the *420* interval would also be required to confirm *At5g53140* as the gene underlying rQTL5^{420}.

The *At5g53140* gene encodes a protein phosphatase 2C family protein with predicted involvement in protein dephosphorylation, based on protein homology (Kerk *et al*. 2002). It is expressed in a variety of plant structures and growth stages, including flowers, inflorescence meristems and plant embryos (Araport11, Krishnakumar *et al*. 2015). Like *ARI7*, the rQTL2^{420} candidate, *At5g53140* does not have an obvious role in recombination, meaning that it could have an indirect effect, potentially through modification of a protein directly involved in the recombination process.

Therefore, preliminary data suggests that *At5g53140* may have a role in modulation of recombination frequency in the *420* interval, but further work confirming the effect of the Salk 010368 T-DNA insertion on the expression of the gene is required before this gene can be properly assessed as a candidate for rQTL5^{420}. If the insertion was demonstrated to affect *At5g53140*, subsequent experiments involving reciprocal transformation of Col-0 and Cvi-0 alleles would need to be performed to confirm allele-specific effects and validate *At5g53140* as the gene underlying rQTL5^{420}.

4.9 Fine-mapping CanQTL4

One significant potentially novel modifier of meiotic recombination frequency was identified on chromosome 4 in the Can-0/*420* F_2 QTL mapping population (LOD 13.6, F-test, p=8.1 x 10^{-11}), explaining

32% of the variance in the F_2 population (Figure 22C). The Can-0 genotype at the peak marker was associated with higher recombination, and the QTL appeared to behave recessively, with no significant difference between heterozygous and Col-0 genotype recombination rates (Two sample t-Test assuming unequal variances p=0.0697, Figure 23B). This QTL, CanQTL4, did not overlap the location of any known modifiers of meiotic recombination frequency (Lawrence *et al*. 2017), so it was assumed to be a novel modifier that could be identified through fine-mapping. The peak identified in the F_2 population was very broad, spanning over 5 Mb, meaning that many additional recombination events breaking up this region were required to allow identification of candidate loci for CanQTL4. However, generation of recombinants in this region of the chromosome would be difficult, due to the proximity of centromeric repeats and large amounts of heterochromatin suppressing recombination. Additionally, the effect of *cis*-acting heterozygosity observed in the Can-0/*420* F_2 population on chromosome 3 (Figure 24, Ziolkowski *et al*. 2015) had to be removed as it caused additional variation within the population that reduced the mapping power available for locating CanQTL4. Therefore, CanQTL4 was introgressed into the Col-0 background through backcrossing with MAS as had previously been performed for $rQTL2^{420}$ and $rQTL5^{420}$, to remove the effect of heterozygosity and small-effect loci that produce additional variation in the population, and to refine the QTL interval.

An F_2 line with QTL4 segregating in a predominantly Col-0 genetic background (individual 95-1) was self-fertilised and an F_3 line was backcrossed to Col-0 to generate a small BC_1F_1 fine-mapping population (Figures 40 and 41). Genotyping identified 17 individuals with informative recombination events breaking up the large QTL interval on chromosome 4, which were then scored for *420* recombination (Table 36). Three lines from this population that were genotyped as Col-0 at all markers across chromosomes 1 and 2, and most markers across chromosomes 3 and 5, with different chromosome 4 genotyping patterns were taken forward to generate BC_1F_2 fine-mapping populations (Figure 40, BC_1F_1 lines Can*420*F_3(F_2 95-1)/Col 4-1, Can*420*F_3(F_2 95-1)/Col 12-2 and Can*420*F_3(F_2 95-1)/Col 19-5, Figure 41). 81 additional BC_1F_2 lines containing recombination events in the QTL region were identified and scored for *420* recombination frequency (Table 37). Fine mapping data from these additional 98 recombinants was combined with Can-0/*420* F_2 population data to create a QTL map to refine the CanQTL4 interval on chromosome 4 using R/qtl (Broman *et al*. 2003; Arends *et al*. 2010. Figure 42A). CanQTL4 still demonstrated recessive behaviour, with the Can-0 allele associating with higher recombination (Figure 42B). Unfortunately, the low number of recombination events in the QTL region prevented resolution of the QTL interval beyond 4.5 Mb (Table 38, Threshold LOD 2.76, Alpha level 0.05, 1000 permutations).

As analysis of the genotyping patterns across QTL regions in fine-mapping lines had been useful to narrow QTL intervals for $rQTL2^{420}$ and $rQTL5^{420}$, this approach was repeated for CanQTL4. Informative

Figure 40. SSLP and CAPS genotyping data for lines generated for fine-mapping of CanQTL4. Parent line denotes the line that was self-fertilised to generate the individual, the Generation column denotes the generation that the individual line belongs to, and the cM column denotes the recombination measurement of the *420* FTL interval in the individual. Genotyping markers are indicated by a number representing their location in the genome (chromosome-kilobase) e.g. marker 2-14176 is on chromosome 2 at position 14176271bp. Col-0 genotype is indicated by the letter C, Can-0 genotype by the letter N and segregating markers by the letter H for each individual at each genotyping marker

				Genotyping markers																	
Parent line	Individual	Generation	cM	1-1242	1-10443	1-19540	1-26357	2-2346	2-14176	3-1031	3-10695	3-12356	3-19165	4-5	4-2618	4-5258	4-10721	4-17158	5-3750	5-25340
Can-0/420 F_1	95-1	F_2	16.83	C	C	C	C	C	C	H	H	C	C	H	H	H	C	C	C	H
Can420F_3(F_2 95-1)/Col	4-1	BC_1F_1	12.13	C	C	C	C	C	C	H	C	C	C	C	C	H	C	C	C	H
Can420F_3(F_2 95-1)/Col	12-2	BC_1F_1	26.02	C	C	C	C	C	C	C	C	C	C	N	N	H	C	C	C	C
Can420F_3(F_2 95-1)/Col	19-5	BC_1F_1	20.65	C	C	C	C	C	C	C	C	C	C	H	H	H	C	C	C	C

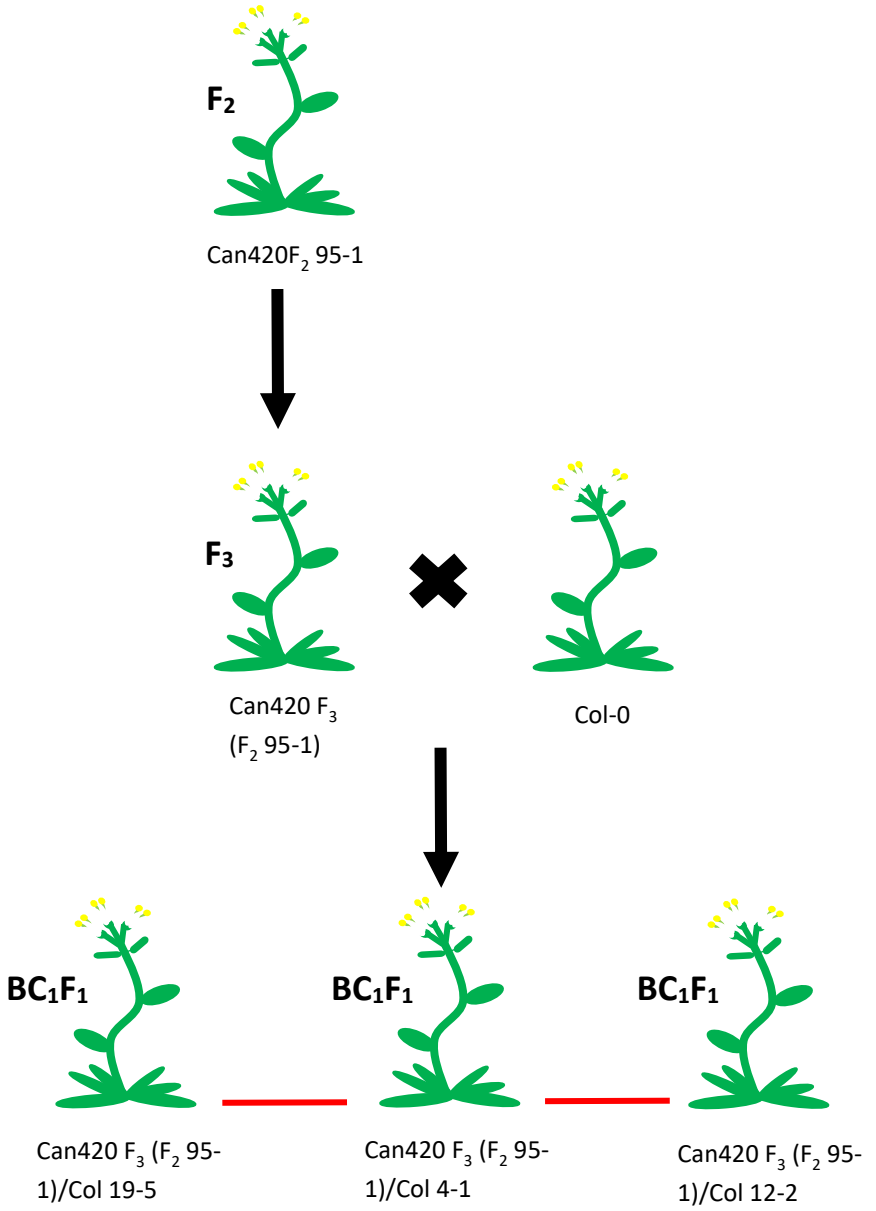


Figure 41: Schematic of CanQTL4 fine mapping plant pedigree. Plants were genotyped using accession-specific SSLP and CAPS markers (see Appendix) and measured for recombination in the *420* FTL interval. For genotyping see Figures 40 and 43. For scoring see Table 36 and Table 37. Black arrows represent generation of new seedlings. Black crosses represent a cross-fertilisation. Sibling relationships are represented by red lines between plants.

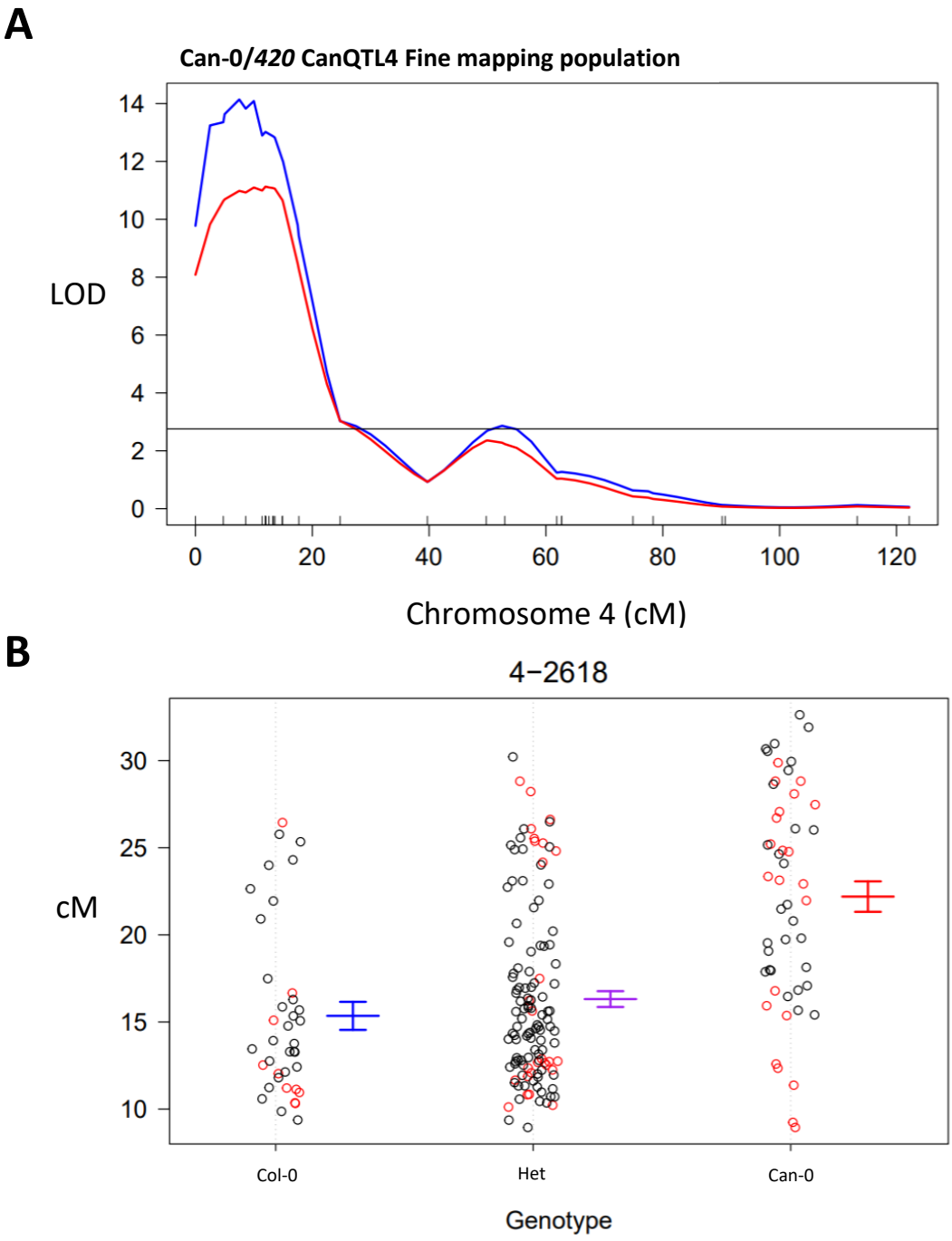


Figure 42. Association of recombination phenotype with marker genotype in the Can-0/*420* CanQTL4 fine mapping population and CanQTL4 effect plot by allele. **A.** Logarithm of odds (LOD) scores for association between *420* recombination frequency and genotyping markers in Can-0/*420* CanQTL4 fine mapping population. Red line denotes association through Hayley-Knott regression, blue line denotes mapping using an expectation-maximisation algorithm (R/qtl). Marker positions indicated by tick marks on the genetic map of chromosome 5 (cM) on the x-axis. 95% significance threshold indicated by horizontal black line. **B**. Effect plot showing recombination frequency (cM) of individuals with Col/Col, Col/Can or Can/Can genotypes at the marker (2,618,095bp) most strongly associated with recombination for CanQTL4 in the fine mapping population. Mean (±SEM) given for each genotype. Col-0 15.3cM (± 0.84); Het 16.1cM (± 0.45); Can-0 22.8cM (± 0.99). Differences between Col-0 and Can-0 genotypes, and Het and Can-0 genotypes were significant (Two sample t-Test assuming unequal variance, $p=3.37 \times 10^{-7}$ and $p=1.85 \times 10^{-7}$ respectively). The difference between Col-0 and Het genotypes was not significant (p=0.44). Missing genotypes (shown in red) filled in by a random imputation that is conditional on the flanking marker genotypes.

recombinant lines were identified and their genotyping patterns across chromosome 4 were assessed to identify the most likely position for CanQTL4 and limit the number of potential candidate loci to be analysed (Figure 43). BC_1F_2 lines Can*420*F_3(F_2 95-1)/Col 12-2 6-1 and 12-4 had high *420* recombination frequencies of 29.9 cM (66% increase relative to Col-0/*420* F_1) and 27 cM (50% increase relative to Col-0/*420* F_1) respectively, and genotyping patterns across chromosome 4 that were fixed as Col-0 from marker '5258' (position 5258448bp) onwards, and Can-0 for the rest of the chromosome. As the Can-0 allele of CanQTL4 is associated with higher recombination, these lines suggest that CanQTL4 is located before marker '5258'. The Can*420*F_3(F_2 95-1)/Col 12-2 5-2 line (a BC_1F_2 line) demonstrated a *420* recombination rate more similar to Col-0/*420* F_1 controls (22cM), suggesting that in this line CanQTL4 is either in the heterozygous state or the Col-0 genotype. As this line is fixed in the Can-0 genotype along chromosome 4 up to marker '4871' (position 4,871,784 bp), this suggests that CanQTL4 is located after marker '4871'. Together, these lines define the interval between markers '4871' and '5258' as the most likely location for CanQTL4. Further recombinants were not generated to refine this interval, as the late flowering phenotype of the Can-0 accession present in fine-mapping populations due to fixation of the *FRIGIDA* flowering time gene on chromosome 4 in the Can-0 genotype limited the number of generations that could be grown (Simon *et al*. 2008). As a result of this, generation of a CanQTL4NIL was not completed, so additional backcrosses using MAS will be required to generate a NIL in future. This will allow confirmation of the effect and location of CanQTL4 independent of the possible effects of other small effect loci in the Can-0 background, as current lines have both a large Can-0 introgression segment on chromosome 4 containing multiple genes and additional heterozygous tracts on non-target chromosomes that may be affecting recombination (see Figures 40 and 43).

4.10 Identification and analysis of candidate loci for CanQTL4

While the fine-mapping of CanQTL4 was not as advanced as that completed for the two QTL identified in the Cvi-0/*420* F_2 mapping population, the putative interval was narrowed to 387 kb between markers '4871' and '5258', and contained just 94 loci (Salk 1001 genomes browser, Alonso-Blanco *et al*. 2016; TAIR seqviewer). The location of CanQTL4 near the centromere of chromosome 4 limited the number of recombinants generated from fine-mapping populations, but it did offer an advantage for the generation of a candidate loci list. Many of the loci present within this interval were non-coding repeat sequences or retrotransposons that were removed from the candidate list (TAIR seqviewer). After additional removal of a few genes which were not expressed during flower development (Araport11, Krishnakumar *et al*. 2015), only 11 loci were left. The classification of Can-0 as a relict accession (Alonso-Blanco *et al*. 2016) means that its sequence is extremely divergent from the other accessions sequenced by the 1,001 Genomes Project, including Col-0. This divergence means that the

Figure 43. SSLP and CAPS genotyping data for lines generated for further fine-mapping of CanQTL4. Parent line denotes the line that was self-fertilised to generate the individual, the Generation column denotes the generation that the individual line belongs to, and the cM column denotes the recombination measurement of the *420* FTL interval in the individual. Genotyping markers are indicated by a number representing their location in the genome (chromosome-kilobase) e.g. marker 2-14176 is on chromosome 2 at position 14176271bp. Col-0 genotype is indicated by the letter C, Can-0 genotype by the letter N and segregating markers by the letter H for each individual at each genotyping marker.

				Genotyping markers															
Parent line	Individual	Generation	cM	1-26357	2-14176	3-11649	4-5	4-230	4-1160	4-2618	4-4636	4-4844	4-4871	4-5153	4-5258	4-5742	4-6781	4-7807	4-10721	5-3750
Can420F_3(F_2 95-1)/Col 12-2	5-2	BC_1F_2	22.92	C	C	C	N	N	N	N	N	N	N	C	C	C	C	C	C	C
Can420F_3(F_2 95-1)/Col 12-2	6-1	BC_1F_2	29.88	C	C	C	N	N	N	N	N	N	N	N	C	C	C	C	C	C
Can420F_3(F_2 95-1)/Col 12-2	12-4	BC_1F_2	27.07	C	C	C	N	N	N	N	N	N	N	N	C	C	C	C	C	C

QTL interval contains a large number of polymorphisms between the Col-0 and Can-0 accessions (Salk 1001 genomes browser, Alonso-Blanco *et al*. 2016). To eliminate irrelevant polymorphisms, the Can-0 sequence was compared to the Cvi-0 sequence (Salk 1001 genomes browser), as CanQTL4 was not identified in the Cvi-0/*420* F_2 mapping population (Figure 20C), suggesting that the polymorphism underlying the QTL is not shared between the two accessions. Removal of genes where polymorphisms were not Can-0 specific, or were synonymous mutations, left 5 candidate loci. Removal of a final gene with a function in lipid metabolism (Salk 1001 genomes browser, Alonso-Blanco *et al*. 2016) that was unlikely to be involved in recombination gave a short-list of four candidates for further analysis (Table 39).

The use of T-DNA insertion mutants to identify genes with an effect on recombination was performed as for $rQTL2^{420}$ and $rQTL5^{420}$. Unfortunately, the location of these genes in a region of the genome with high levels of repeat sequences and heterochromatin meant that few insertional mutagenesis lines were available. Of the four candidates identified, T-DNA lines were only available for two. Two mutant lines were identified for each of these genes using the T-DNA Express tool (The Salk Institute Genomic Analysis Laboratory) and ordered from NASC (Table 4). T-DNA lines were then reciprocally crossed to the *420* reporter line as both the male and female parent, and F_1 lines were measured for *420* recombination. No significant differences were detected between crosses using the mutant as the male or female parent, so reciprocal F_1 line data was pooled and analysis was performed comparing each mutant F_1 to Col-0/*420* F_1 controls (Table 40, Figure 44).

Scoring of T-DNA/*420* F_1 lines revealed that three out of the four lines analysed did not have a significant effect on *420* recombination when compared to control Col-0/*420* F_1 lines (Two-tailed χ^2 tests, see Table 40 for p-values). The GABI-Kat line 219G07, with an insertion in exon three of the *TPR8* gene *At4g08320*, showed a significant increase in recombination compared to Col-0/*420* F_1 replicates ($\chi^2(1) = 5.858$, p=0.031). However, the mean recombination rate of GK 219G07/*420* F_1 replicates at 19.9 cM was only 1.4 cM higher than growth-matched Col-0/*420* F_1 replicates (8% increase relative to Col-0/*420* F_1, Figure 44), and a second insertion line available for this gene (SAIL 731H04) did not show a significant effect on recombination when crossed to *420* ($\chi^2(1) = 2.896$, p=0.1776). While this does not preclude *At4g08320* acting as a modifier of recombination, as the SAIL 731H04 insertion is in the 3' end of the gene and may have a different effect, this minor effect does not seem to support *At4g08320* as the candidate gene underlying CanQTL4, unless the allele is exceptionally weak in comparison to the Can-0 allele effect. Another possibility is that, like the Can-0 allele of CanQTL4, the GK 219G07 T-DNA insertion allele is recessive. Therefore, future analysis of *420* recombination in homozygous T-DNA insertion mutants is required to clarify this and quantify the effect of the GK 219G07 insertion, and determine whether *At4g08320* is a likely candidate for CanQTL4. Further

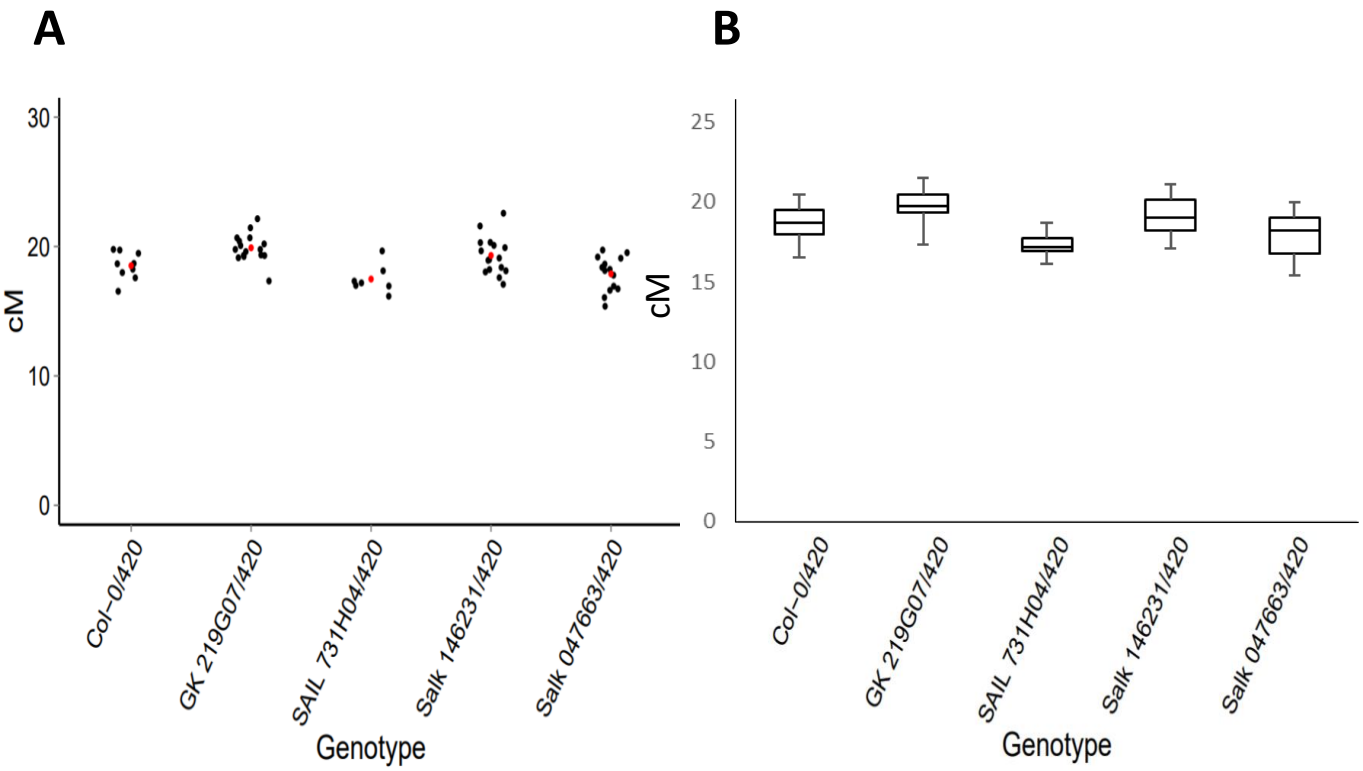


Figure 44: Genetic distance (cM) as measured in the *420* seed-based FTL-interval for CanQTL4 T-DNA insertion candidate lines. Measurements were made in Col-0/*420* and T-DNA/*420* F_1 individuals (Table 40: Total 86713 seeds). **A**. Dotplot. Replicates are in black, with the mean of each genotype highlighted in red. **B**. Same data represented in boxplots to highlight differences between genotypes. Interquartile range represented by box, median represented by midline. Minimum and maximum phenotype values for each genotype represented by whiskers. Differences between Col-0/*420* controls and T-DNA/*420* lines are only significant for the GK 219G07/*420* cross ($\chi^2(1) = 5.858$, p=0.031). See Table 4 for T-DNA lines and NASC ID.

experiments to determine the effect of these insertions on the *At4g08320* gene need to be performed before any conclusions can be made about it's possible function as a recombination modifier.

TPR8 (*At4g08320*) is a carboxylate clamp-tetratricopeptide repeat protein which, based on protein structure, has the potential to interact with Hsp90/Hsp70 as a co-chaperone (Prasad *et al*. 2010). The protein is broadly expressed in a variety of plant structures and growth stages (Araport11, Krishnakumar *et al*. 2015). As the gene is highly polymorphic between the Col-0 and Can-0 accessions (Salk 1001 genomes browser, Alonso-Blanco *et al*. 2016), it would not be possible to identify the causal polymorphism without extensive future experiments, which would not be pursued until the effect of *At4g08320* was confirmed. Therefore, while a potential candidate has been identified for CanQTL4, the minor effect observed in GK 219G07/*420* F_1 lines and the lack of any obvious mechanism of action requires additional work to be performed to confirm *At4g08320* as the locus underlying CanQTL4. Reciprocal transformation of Can-0 and Col-0 alleles of *At4g08320* and subsequent observation of a stronger effect on recombination could support the theory that the GK 219G07 insertion allele has a particularly weak effect. However, reciprocal allele transformation analysis would be complicated by the dominance effect of the Col-0 allele observed at the CanQTL4 peak marker in the Can-0/*420* F_2 population. Therefore, identification of a confirmed *At4g08320* knockout line, through analysis of gene expression, would be required to allow comparison of Col-0 and Can-0 allele transgenic complementation in the mutant line (Mitchell-Olds and Schmitt. 2006; Koornneef *et al*. 2004; Alonso-Blanco and Koornneef 2000; Filiault *et al.* 2008; Maloof 2003; Tessadori *et al.* 2009). This would allow direct comparison of accession specific effects without the complication of dominance, and confirm *At4g08320* as a modifier of recombination that varied between Col-0 and Can-0 accessions.

Future work to characterise CanQTL4 would include generation of a NIL to confirm the location and effect of the modifier, analysis of NIL crosses to alternative intervals to determine region specificity, measurement of recombination in homozygous T-DNA insertion lines to determine dominance effects, and complementation tests to confirm *At4g08320* as the gene underlying CanQTL4.

4.11 Conclusions

Three significant potentially novel modifiers of meiotic recombination frequency were identified in accession cross F_2 mapping populations and taken forward for further fine-mapping and identification of potential candidate genes. The effect and location of $rQTL2^{420}$ and $rQTL5^{420}$ were confirmed, and a preliminary analysis of region-specificity was performed. Candidate genes with T-DNA insertion lines demonstrating a significant effect on recombination in the *420* interval were identified for both QTL, and future work should be focused on confirmation of candidates through transformation of accession-specific alleles into different genetic backgrounds and analysis of their effects on

recombination. Preliminary characterisation of CanQTL4 was restricted by a reduced number of mapping generations caused by a late-flowering phenotype, and limited recombination in the region surrounding the QTL due to proximity to the centromere. However, a refined candidate interval was identified that needs to be confirmed by generation and analysis of a NIL, and a candidate gene whose T-DNA insertion mutant had a significant, though minor, effect on recombination was identified. Future work to confirm this candidate gene through allele complementation is required.

Chapter 5 – Characterisation of the HEI10 meiotic E3 ligase as a modifier of meiotic recombination

5.1 Summary

A significant large-effect modifier of meiotic recombination frequency, rQTL1^{420}, was identified on chromosome 1 in an *Arabidopsis thaliana* accession-cross F_2 QTL mapping population. Back-crossing into the Col-0 accession background was performed to refine the rQTL1^{420} location and quantify its individual effect on crossover frequency. This mapping identified the gene *HEI10*, which encodes a conserved meiotic E3 ligase, as a potential candidate gene underlying rQTL1^{420}. Transformation of accession-specific *HEI10* alleles into an FTL-reporter system revealed that *HEI10* exerts a dosage-sensitive effect on crossover frequency. In this chapter, data is presented showing that *HEI10* is the gene underlying rQTL1^{420} in the Col/Cvi population, but that it may be one of multiple linked loci affecting recombination. Additional analysis of a *HEI10* overexpressor line suggests that the dosage-sensitive increase in recombination also causes a reduction in crossover interference.

5.2 Introduction: HEI10; a meiosis-specific E3 ubiquitin-ligase required for Class I crossovers

The Arabidopsis HEI10 protein is a structural and functional homolog of the budding yeast Zip3 ZMM protein which functions in synapsis initiation (Chelysheva *et al*. 2012). It also shares many similarities with the mammalian HEI10 protein, which has been shown to have *in vitro* E3 ubiquitin-ligase activity, and to be involved in meiotic recombination (Ward *et al.* 2007). Work with *hei10* knockout mutants in Arabidopsis has revealed that a function in meiotic recombination is conserved, as loss of HEI10 results in an 85-90% reduction in crossovers and the presence of univalents during metaphase I of meiosis (Chelysheva *et al*. 2012). This substantial reduction in crossovers in *hei10* knockout lines, combined with the protein's co-localisation with the MLH1 protein on meiotic chromosomes, a marker of class I crossovers, suggests a role in ZMM-dependent class I crossover formation. This was confirmed by analysis of *hei10 zmm* double mutants which showed no further reduction in recombination, compared to *zmm* single mutants (Chelysheva *et al*. 2012).

A role in the ZMM pathway suggests a function related to stabilisation of early recombination intermediates, but retention of HEI10 foci on chromosomes suggests that it may also have a later role, possibly by promoting dHJ resolution as class I crossovers (Chelysheva *et al*. 2012). In mice, RNF212 and HEI10 have been shown to stabilise chromosome binding of MSH4/MSH5 heterodimers and promote crossover resolution via a SUMO-ubiquitin relay, although the direct protein targets are still unknown (Qiao *et al*. 2014; Rao *et al*. 2017). As plants only contain one *RNF212/HEI10* ortholog, it is likely that analogous functions are carried out by HEI10 in Arabidopsis (Chelysheva *et al*. 2012).

5.3 Background: Natural variation and dosage-sensitivity of *HEI10* in meiotic recombination

HEI10 has previously been identified as a modifier of meiotic recombination frequency that varies between natural Arabidopsis accessions (Ziolkowski *et al*. 2017). QTL mapping in a Col/L*er* F_2 population had revealed a large peak (LOD 40.2) on chromosome 1 that, after subsequent fine-mapping, was shown to overlap the meiotic recombination gene *HEI10* (Ziolkowski *et al*. 2017). As HEI10 is a conserved protein known to promote crossovers in a range of eukaryotic organisms (Bhalla *et al*. 2008; Fledel-Alon *et al*. 2011; De Muyt *et al*. 2014), and polymorphisms in *RNF212* and *HEI10* had previously been shown to associate with variation in recombination rate in mammalian populations (Kong *et al*. 2008; Sandor *et al*. 2012; Fledel-Alon *et al*. 2011), Ziolkowski *et al* (2017) assessed it as a candidate for the QTL.

Complementation of the null *hei10* allele, both by crossing to Col and to an introgression line with $HEI10^{Ler}$ in a Col background demonstrated that this mutation has a dosage-sensitive effect, unlike other *zmm* knockout mutants such as *msh4* (Ziolkowski *et al*. 2017). This effect varied depending on the *HEI10* allele used to complement the knockout line. While both $HEI10^{Col}$ and $HEI10^{Ler}$ alleles caused an increase in recombination over the null line when crossed to *hei10-2*, $HEI10^{Col}$ had a significantly greater effect, consistent with this allele being more active in promoting recombination (Ziolkowski *et al*. 2017). This was further supported by transformation of the $HEI10^{Col}$ and $HEI10^{Ler}$ alleles into the Col-0/*420* FTL reporter line (Ziolkowski *et al*. 2017). T_1 plants for both constructs had a recombination rate that was substantially and significantly higher than wild type controls, due to additional copies of the gene (Ziolkowski *et al*. 2017). Additionally, $HEI10^{Col}$ T_1s had significantly higher *420* recombination than $HEI10^{Ler}$ T_1s, demonstrating the effect of the different alleles (Figure 45) (Ziolkowski *et al*. 2017). These additional crossovers were shown to be located primarily in euchromatic regions, especially in the subtelomere (Figure 45). While cytological observation of MLH1 foci in a *HEI10* overexpressor line (line C2, a $HEI10^{Col}$ T_1) showed that crossovers were elevated generally across the genome, crossover mapping via genotyping-by-sequencing revealed that increases were minor in the pericentromere (Ziolkowski *et al*. 2017).

One non-synonymous and three synonymous intragenic variants between Col-0 and L*er* were identified in the *HEI10* sequence, and additional polymorphisms were identified in the promoter sequence (Ziolkowski *et al*. 2017). The L*er* sequence polymorphisms were compared with the Cvi-0 and Bur-0 *HEI10* alleles (Figure 46A), as similarly positioned QTL peaks were observed in Col/Bur and Col/Cvi F_2 mapping populations (Ziolkowski *et al*. 2017). Shared polymorphisms were then compared to the Ct-1 allele, as no chromosome 1 QTL peak was observed in a Col/Ct F_2 population (Ziolkowski *et al*. 2015), thereby revealing 12 shared variants between L*er*, Cvi-0 and Bur-0 that were not shared

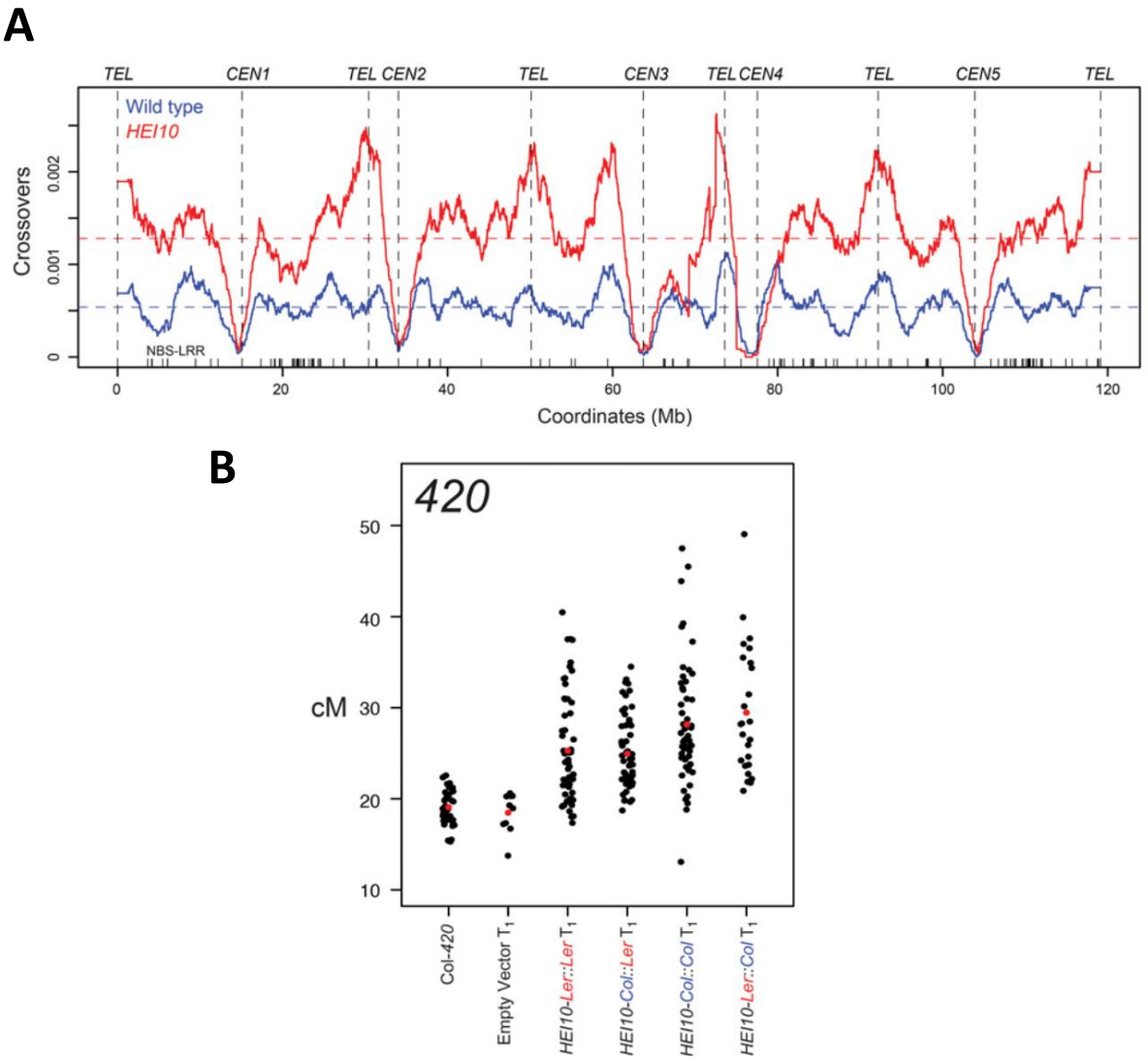


Figure 45. *HEI10* dosage-dependent increases in recombination. Adapted from Ziolkowski *et al*. (2017). **A**. Crossover frequency along the five Arabidopsis chromosomes in wild-type (blue) and *HEI10* overexpressor (red) Col/L*er* F_2 populations. Mean values are shown by the dotted horizontal lines, and telomere (TEL) and centromere (CEN) positions are indicated by vertical dotted lines and labels. **B**. Genetic distance (cM) as measured in the *420* FTL-interval in empty vector, $HEI10^{Ler}$ T_1s, $HEI10^{Col}$ T_1s and promoter swap construct T_1s, and untransformed Col-0/*420* controls. Replicates are in black, with the mean of each genotype highlighted in red.

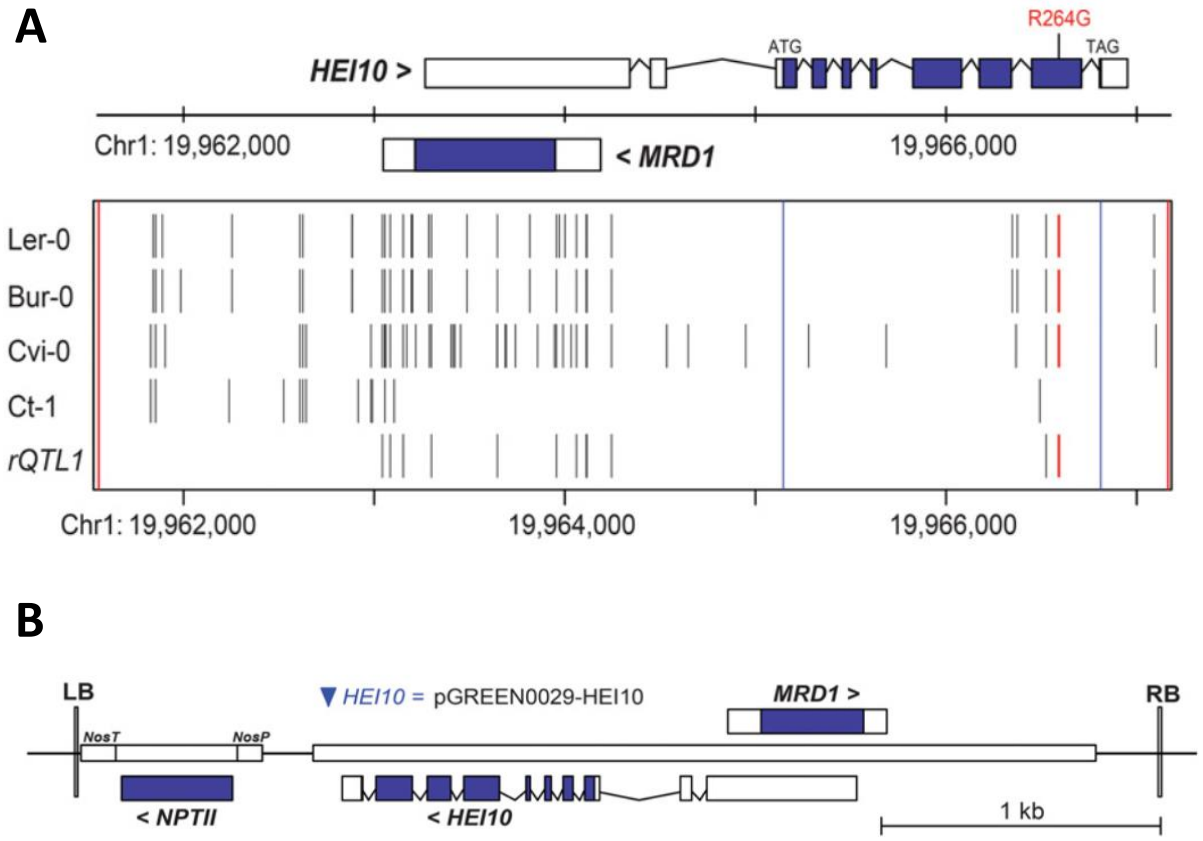


Figure 46. *HEI10* polymorphisms and transformation construct. Adapted from Ziolkowski *et al*. (2017). A. Polymorphisms in the *HEI10* region of chromosome 1. *HEI10* and *MRD1* gene annotations are plotted as black boxes, with coding regions shown in blue. Blue vertical lines indicate HEI10 ATG and TAG codons. Black ticks show the positions of L*er*-0, Bur-0, Cvi-0 and Ct-1 polymorphisms, identified by Sanger sequencing, and polymorphisms in L*er* rQTL1. Red ticks indicate the presence of the non-synonymous substitution R264G. **B**. Diagram illustrating the *HEI10* transgenic construct used for Agrobacterium transformation of Col-0/*420* lines. LB denoted the T-DNA left-border sequence, RB denotes the T-DNA right border sequence.

with Ct-1 and could therefore potentially underlie the QTL. These included a non-synonymous amino acid substitution, the R264G variant, in the C-terminal region which is predicted by homology to be involved in E3 substrate recognition (Deshaies and Joazeiro. 2009).

Analysis of *HEI10* expression levels in Col and L*er* through quantitative RT-PCR by Ziolkowski *et al*. (2017) revealed no significant differences. Furthermore, transformation of $HEI10^{Col}$ and $HEI10^{Ler}$ promoter swap constructs into Col-0/*420* supported this lack of accession-specific effect, suggesting that the observed allelic differences were unlikely to be due to expression differences, and were more likely attributable to intragenic variants such as R264G (Ziolkowski *et al*. 2017). This variant was suggested to alter substrate recognition and increase SUMO/ubiquitin transfer by HEI10 during recombination, resulting in the Col-0 allele more effectively promoting crossover resolution (Ziolkowski *et al*. 2017). However, additional experiments were required to confirm this as the causal variant underlying the QTL in L*er*, Cvi-0 and Bur-0 populations.

5.4 Fine-mapping rQTL1^{420} in the Cvi-0/*420* F_2 population

In this project, QTL mapping in the Cvi-0/*420* F_2 population had identified three significant QTL, the largest of which was found on chromosome 1 (LOD 44.02, F-test $p<2 \times 10^{-16}$) that explained 44.4% of the variance in recombination in the population (Figure 20C). The Col-0 allele at the peak marker was associated with higher recombination compared with the Cvi-0 allele, and heterozygous individuals showed an intermediate phenotype, suggesting semi-dominance of the locus (Figure 21A). This broad peak overlapped the *HEI10* gene, which was known to cause variation in crossovers between Arabidopsis accessions (Ziolkowski *et al*. 2017). Several polymorphisms in the Cvi-0 allele of *HEI10* had already been identified that differed from Col-0 but were shared with L*er*-0, including the putative causal variant proposed for the L*er*-0 allele effect, R264G (Ziolkowski *et al*. 2017). This suggested that *HEI10* was a strong candidate for the gene underlying rQTL1^{420}, therefore further experiments prioritised confirmation of this locus over fine-mapping to identify alternative candidates.

Additional work was performed to confirm the activity of rQTL1^{420}. As had previously been performed for rQTL2^{420}and rQTL5^{420}, rQTL1^{420} was backcrossed into the Col-0 background to isolate it from the effects of other QTLs, small effect loci and *cis*-acting heterozygosity (see Chapters 3 and 4). Generation of a NIL would then allow confirmation of the effect of rQTL1^{420} on recombination in the *420* interval and, through crosses to additional FTL intervals, analysis of the effect of rQTL1^{420} on different regions of the genome. Two F_2 lines from the Cvi-0/*420* population were selected (lines 13-1 and 27-3) that had large proportions of the genome in the Col-0 background, with the exception of rQTL1^{420} (Figure 47). These were then backcrossed to Col-0 and self-fertilised through multiple generations utilising MAS to generate further fine-mapping lines with the Cvi-0 allele of rQTL1^{420} in a background that was

Figure 47. SSLP and CAPS genotyping data for lines generated for fine-mapping of rQTL1^{420}. Parent line denotes the line that was self-fertilised to generate the individual, the Generation column denotes the generation that the individual line belongs to, and the cM column denotes the recombination measurement of the *420* FTL interval in the individual. Genotyping markers are indicated by a number representing their location in the genome (chromosome-kilobase) e.g. marker 1-6108 is on chromosome 1 at position 6108789bp. Col-0 genotype is indicated by the letter C, Cvi-0 genotype by the letter V and segregating markers by the letter H for each individual at each genotyping marker.

				Genotyping markers																
Parent	Individual	Generation	cM	1-6108	1-16908	1-18587	1-19540	1-22709	1-24743	1-30413	2-132	2-8799	2-16339	3-1031	3-15949	3-21008	4-230	4-7157	5-9437	5-19994	5-24192
Cvi-0/420 F_1	13-1	F_2	6.54	H	H	H	H	V	H	C	C	C	C	H	H	H	V	H	H	H	H
Cvi-0/420 F_1	27-3	F_2	6.18	H	H	V	V	V	H	V	H	C	C	H	C	C	C	H	C	C	H
Cvi420F_3(F_2 27-3)/Col	1-1	BC_1F_1	14.46	C	C	H	H	H	C	H	C	C	C	C	C	C	C	C	C	C	C
Cvi420F_3(F_2 27-3)/Col	1-2	BC_1F_1	14.80	C	H	H	H	H	H	H	C	C	C	C	C	C	C	C	C	C	C
Cvi420F_3(F_2 13-1)/Col	1-4	BC_1F_1	11.84	C	H	H	H	H	C	C	C	C	C	C	C	C	H	C	C	C	C
Cvi420F_3(F_2 27-3)/Col 1-1	1-4	BC_1F_2	14.81	C	C	V	V	V	C	C	C	C	C	C	C	C	C	C	C	C	C
Cvi420F_3(F_2 13-1)/Col 1-4	1-8	BC_1F_2	13.03	C	V	V	V	C	C	C	C	C	C	C	C	C	C	C	C	C	C

Col-0 for all markers genotyped across other chromosomes and the majority of chromosome 1 (Figures 47 and 48 and Table 41). This removed the effects of segregating rQTL2^{420} and rQTL5^{420} and produced an rQTL1^{420} NIL that could be used to characterise the individual effect of this QTL independently of the other large-effect modifiers (Alonso-Blanco and Koornneef 2000).

Analysis of the genotyping pattern of these lines revealed that rQTL1^{420} must be between markers '16908' (position 16,907,783 bp) and '22709' (position 22,709,104 bp) on chromosome 1 (Figure 47). Line Cvi*420*F$_3$ (F$_2$ 27-3)/Col 1-1 1-4 (a BC$_1$F$_2$ line) was fixed for the Col-0 genotype along chromosome 1 to marker '16908' and distal from marker '24743' (position 24,743,355 bp) and showed a low recombination rate in the *420* interval of 14.8 cM (18% decrease relative to Col-0/*420* F$_1$). This suggested that the Cvi-0 allele of rQTL1^{420} must be present, thereby locating the QTL between these markers. BC$_1$F$_2$ line Cvi*420*F$_3$(F$_2$ 13-1)/Col 1-4 1-8 was fixed in the Col-0 genotype distal to marker '22709', and again exhibited low *420* recombination (13 cM, 28% decrease relative to Col-0/*420* F$_1$), indicating that rQTL1^{420} is present in the Cvi-0 genotype and must therefore be located proximal to marker '22709'. As *HEI10* falls within this interval and was the most promising candidate, further fine-mapping was not performed.

5.5 Confirmation of rQTL1^{420} effect and preliminary analysis of region-specificity

The fine-mapping line Cvi*420*F$_3$(F$_2$ 27-3)/Col 1-1 1-4 (BC$_1$F$_2$ line, see Figures 47 and 49), which contained rQTL1^{420} fixed in the Cvi-0 genotype in a predominantly Col-0 genetic background, was used as a NIL and self-fertilised to generate rQTL1^{420}NIL replicates (BC$_1$F$_3$), in order to confirm the effect of rQTL1^{420} and its position between markers '16908' and '24743'. Replicates of rQTL1^{420}NIL had a mean *420* recombination rate of 16.4 cM, which was significantly lower than the Col-0/*420* F$_1$ controls (mean 19.5 cM) ($\chi^2(1) = 33.19$, $p=1.67 \times 10^{-8}$, Table 42 and Figure 50). This confirms that a significant modifier of recombination is located within this interval, and that the Cvi-0 allele causes an average decrease of 3.1 cM in the *420* interval, which is a 16% decrease in recombination relative to Col-0/*420* F$_1$ controls.

However, the effect observed in the rQTL1^{420}NIL (BC$_1$F$_3$) is smaller than that previously observed in alternative rQTL1^{420} fine mapping lines (Figure 47 and Table 41), such as the Cvi*420*F$_3$(F$_2$ 13-1)/Col 1-4 1-8 line (a BC$_1$F$_2$ line) which showed a *420* recombination rate of 13 cM. The 3.1 cM decrease observed in the rQTL1^{420}NIL is not sufficient to explain the huge QTL peak observed in the Cvi-0/*420* F$_2$ population (Figure 20C). As no epistasis was detected with the other large effect QTL in the F$_2$ population, a possible explanation for the reduced effect in the rQTL1^{420}NIL is that potential interactions between rQTL1^{420} and heterozygosity or small effect loci in the genome increased the effect of rQTL1^{420} on recombination. Loss of these factors through gradual introgression into the Col-

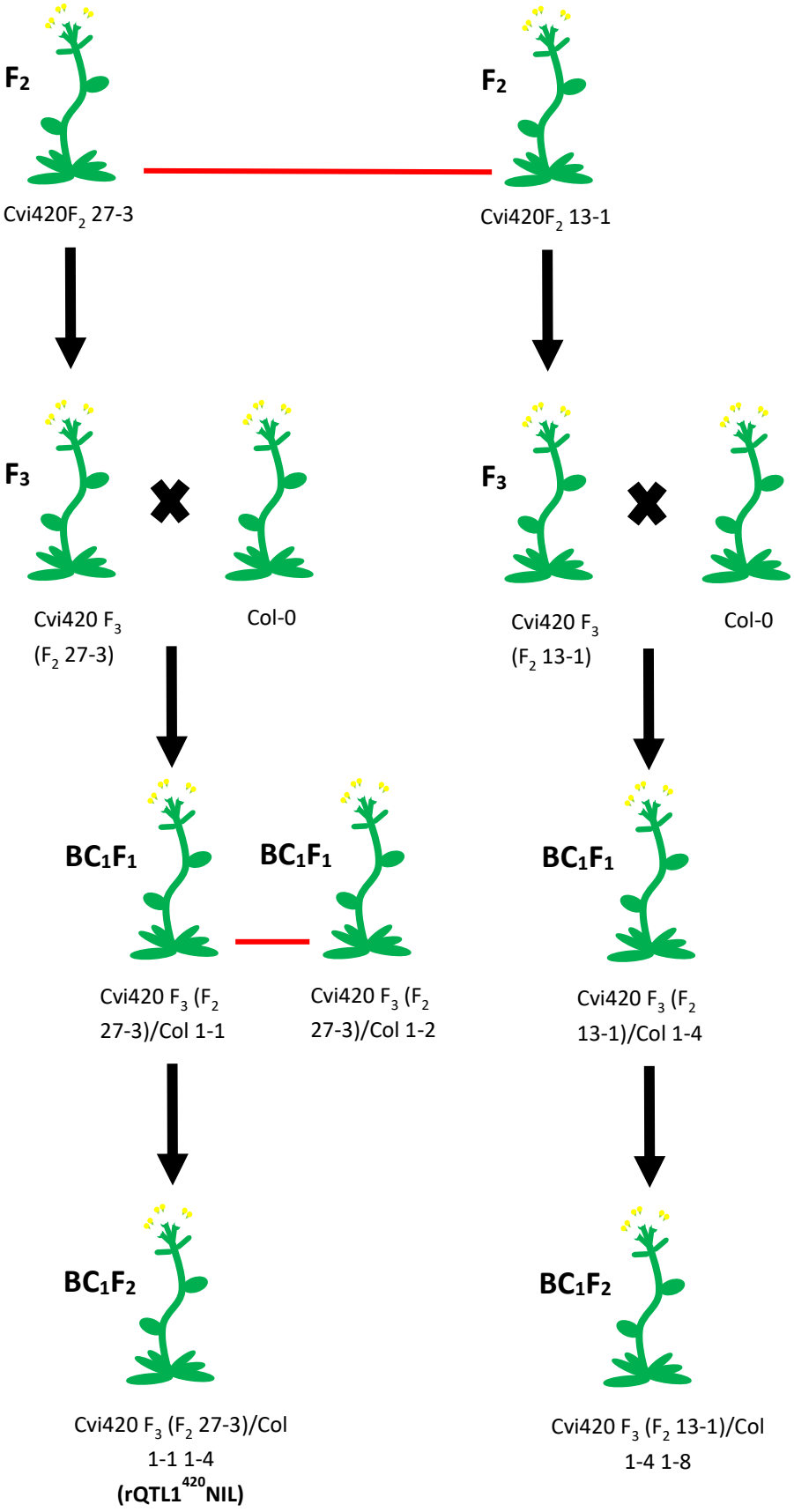


Figure 48: Schematic of rQTL1^{420} fine mapping plant pedigree. Plants were genotyped using accession-specific SSLP and CAPS markers (see Appendix) and measured for recombination in the *420* FTL interval. For genotyping see Figure 47. For scoring see Table 41. Black arrows represent generation of new seedlings. Black crosses represent a cross-fertilisation. Sibling relationships are represented by red lines between plants.

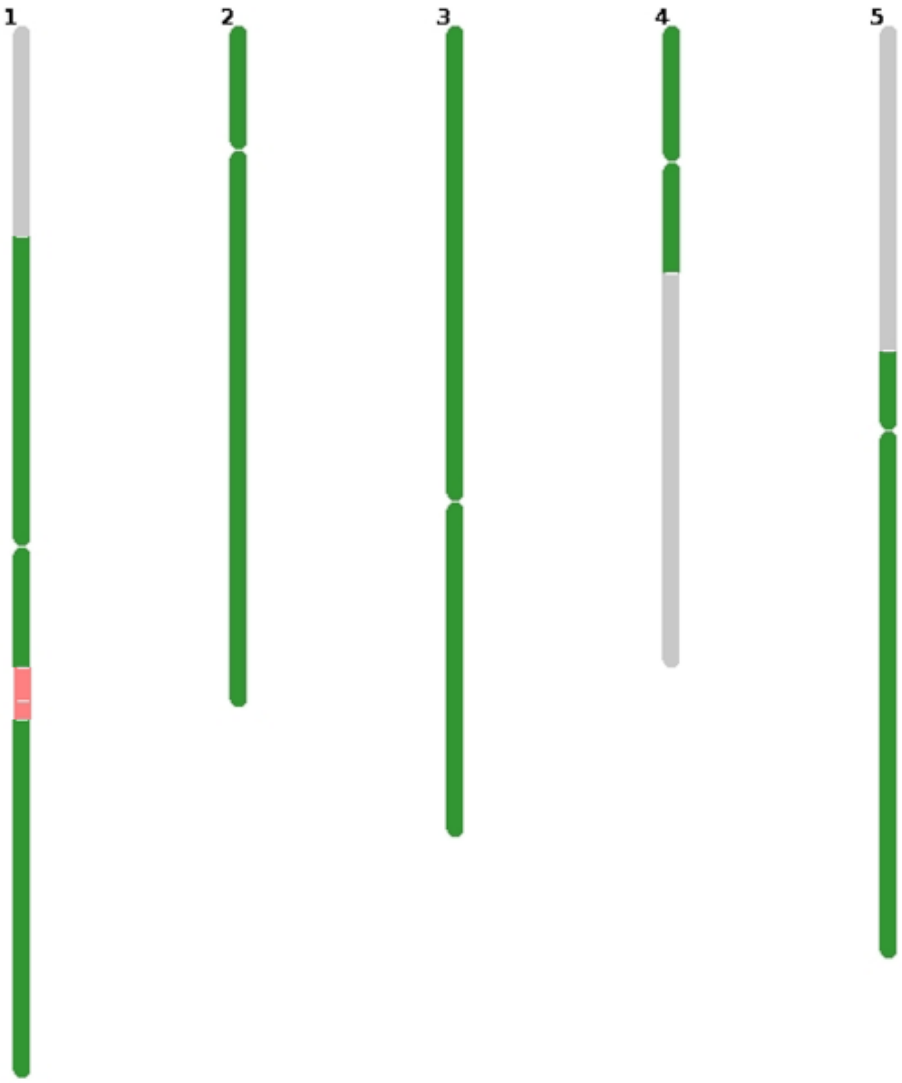


Figure 49: Chromosome map showing the genotype of the rQTL1^{420}NIL. Col-0 genotype is shown in green, Cvi-0 genotype in pink. Grey areas show regions of the genome where appropriate PCR-based markers were not designed and therefore the genotype is unknown. Generated using TAIR Chromosome map tool.

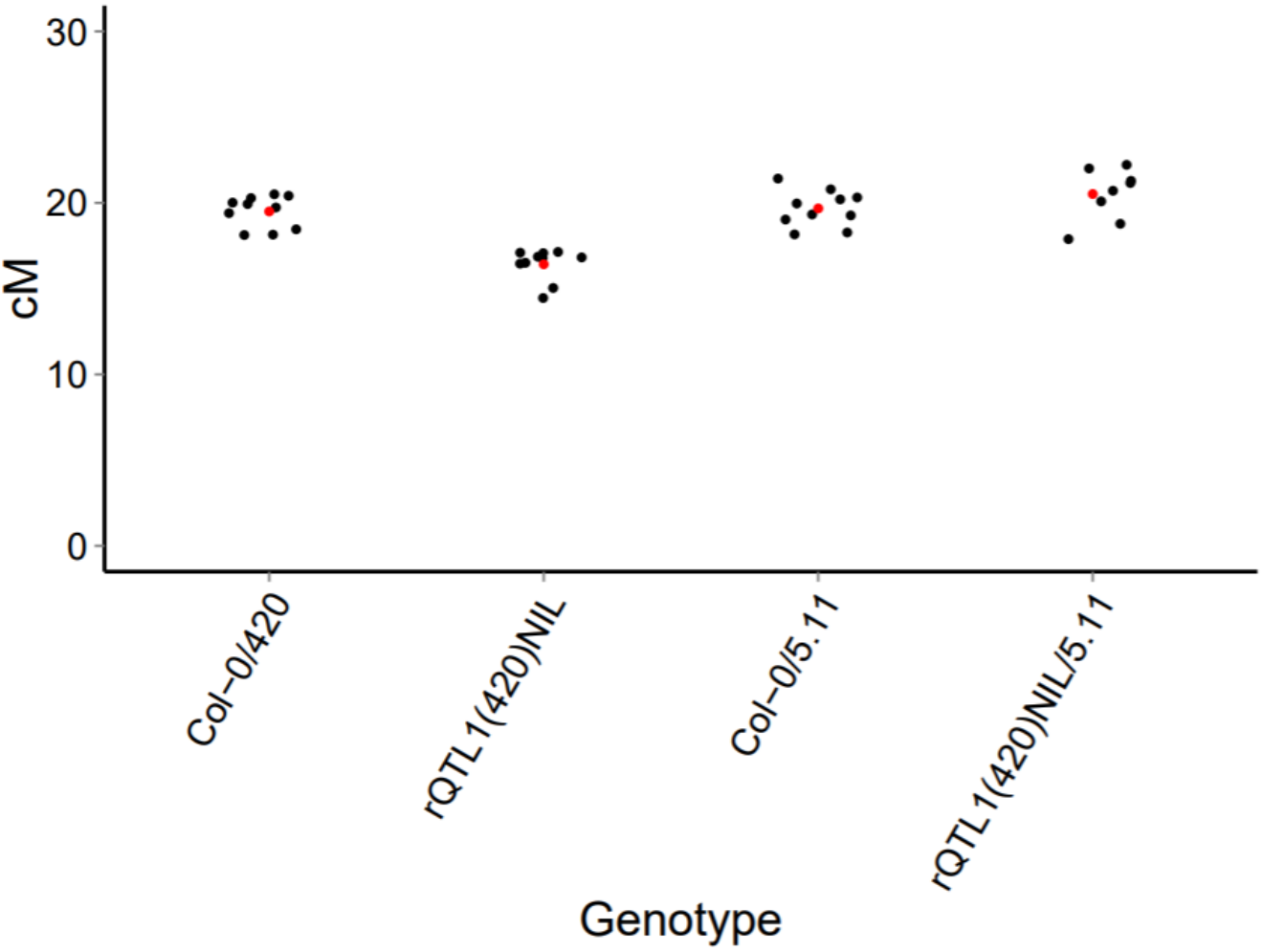


Figure 50: Genetic distance (cM) as measured in *420* and *5.11* seed-based FTL-intervals. Measurements were made in Col-0/*420* and Col-0/*5.11* F_1 individuals, and in rQTL1^{420}NIL and rQTL1^{420}NIL/*5.11* individuals (Tables 42 and 43: Total 29889 and 27602 seeds respectively). Replicates are in black, with the mean of each genotype highlighted in red. Differences between Col-0 lines and rQTL1^{420}NILs are significant for the *420* interval ($\chi^2(1) = 33.19261$, $p=1.67 \times 10^{-8}$) but not the *5.11* interval ($\chi^2(1) = 1.95$, $p=0.3252$).

0 background could therefore have reduced the effect of rQTL1^{420}. Alternatively, it is also possible that the rQTL1^{420} effect could be due to multiple linked loci affecting recombination. Recombination events during the generation of the rQTL1^{420}NIL could have separated these QTL, and fixed one in the Col-0 background, thereby decreasing the overall reduction of *420* recombination observed in these lines. The lower recombination observed in the Cvi*420*F_3(F_2 13-1)/Col 1-4 1-8 line (a BC_1F_2 line) could be due to multiple QTL being present in the Cvi-0 genotype. Consistent with this theory, this line does not have the same chromosome 1 genotype as the rQTL1^{420}NIL (BC_1F_3). Specifically, Cvi*420*F_3(F_2 13-1)/Col 1-4 1-8 has a larger section of the chromosome proximal to marker '18587' in the Cvi-0 genetic background (Figure 47). This could potentially be the location of an additional rQTL, although this would need to be verified by generation of lines with a recombination event between the potential QTL, with the more distal QTL being fixed in the Col-0 genotype.

While no obvious recombination modifiers are present within this region, there are several potential candidates with DNA-binding functions or functions related to known recombination modifiers, including transcription factors, a ubiquitin protein ligase, a ubiquitin-conjugating E2-protein and a RING/U-box superfamily protein among others (Salk 1001 genomes browser, Alonso-Blanco *et al*. 2016; TAIR seqviewer found at: https://seqviewer.arabidopsis.org/). Alternatively, if the alleles of the second QTL had contrasting effects to the first, whereby the Col-0 allele was associated with lower recombination, it could possibly be located distal to the first QTL, between markers '19540' and '24743' - a region which appears to be in the Cvi-0 genotype in the rQTL1^{420}NIL, and predominantly in the Col-0 genotype in line Cvi*420*F_3(F_2 13-1)/Col 1-4 1-8 (Figure 47). Located in this region is the *RECQ4B* gene, which is known to affect recombination frequency in Arabidopsis (Seguela-Arnaud *et al*. 2015). A number of non-synonymous polymorphisms between Col-0 and Cvi-0 accessions are present within this gene, making it a promising candidate for a second QTL (Salk 1001 genomes browser, Alonso-Blanco *et al*. 2016). However, this is purely speculative as additional lines and further genotyping would be required to determine the presence of a second QTL, and its possible location.

Although rQTL1^{420} demonstrated a significant effect on recombination in the subtelomeric *420* interval, its effect on recombination in other areas of the genome was unknown. No QTL were observed on the distal arm of chromosome 1 in Cvi-0/*CEN3* and Cvi-0/*5.11* F_2 mapping populations (Figure 17C and Figure 25C) which measured recombination in centromeric intervals, suggesting that rQTL1^{420} may not affect recombination around the centromere. To confirm that the rQTL1^{420} effect on the centromere was not being masked by other segregating QTL with antagonistic effects in the F_2 population, the rQTL1^{420}NIL was crossed to the *CEN3* and *5.11* FTLs. While technical issues prevented the generation of data from the *CEN3* cross, analysis of rQTL1^{420}NIL/*5.11* replicates revealed that rQTL1^{420} did not have a significant effect on recombination in the *5.11* interval when compared to Col-

0/*5.11* F_1 controls ($\chi^2(1) = 1.95$, p=0.3252, Table 43 and Figure 50). While this experiment only addresses recombination across the centromere of chromosome 5, the data from the *CEN3* F_2 population suggests this is likely to be the case on other chromosomes. This effect is consistent with rQTL1^{420} corresponding to the *HEI10* locus, as recombination in *HEI10* overexpressor lines is predominantly increased in the euchromatic arms and sub-telomere, with little effect observed in the pericentromeres (Ziolkowski *et al*. 2017). Further work analysing alternative centromeric FTL intervals could be performed to confirm this effect.

5.6 Analysis of *HEI10* as a potential candidate gene underlying rQTL1^{420}

Introgression of rQTL1^{420} into the Col-0 background had revealed the general location of the QTL as being between markers '16908' and '22709' – a region which contained the *HEI10* gene. The position of the QTL in the F_2 population, with a peak marker less than a megabase away from the *HEI10* gene (Figure 21A), combined with shared polymorphisms between Cvi-0 and the L*er* accession, where the *HEI10* allele had been demonstrated to cause variation in recombination between Col and L*er* accessions (Ziolkowski *et al*. 2017), made *HEI10* a strong candidate for rQTL1^{420}. Sequencing of the Cvi-0 allele had identified 30 polymorphisms between Cvi-0 and Col-0, including the R264G amino acid substitution which was suggested to be the causal variant underlying the allelic differences between Col and L*er* (Figure 46A) (Ziolkowski *et al*. 2017). Therefore, assessment of *HEI10* as a potential recombination modifier in the Cvi-0 background was performed through transformation of transgenic copies of the Cvi-0 allele of *HEI10* (*HEI10^{Cvi}*) into the Col-0/*420* reporter system.

This strategy had previously been used by Ziolkowski *et al*. (2017) to demonstrate the dosage sensitivity of *HEI10* (Figure 45) and quantify the difference in effect caused by additional copies of the *HEI10^{Col}* and *HEI10^{Ler}* alleles. This approach cloned alleles with their endogenous promoters and, as additional promoter swap constructs had shown that variation was likely caused by intragenic variants (Ziolkowski *et al*. 2017), cloning constructs for analysis of *HEI10*^{Cvi} were designed containing the Cvi-0 promoter. To produce comparable data, the *HEI10*^{Col} and *HEI10*^{Ler} pGREEN0029 constructs generated by Ziolkowski *et al*. (2017) were used as transformation controls. The same primers were used for amplification of *HEI10*^{Cvi} from Cvi-0 genomic DNA and insertion into the binary pGREEN0029 vector (Figure 46B, see Materials and Methods). An additional construct had been produced in the lab (gift from E. Lawrence) using site-directed mutagenesis to replace the Col R264 amino acid with the L*er* 264G amino acid in the *HEI10^{Col}* construct, to test the isolated effect of the R264G variant, and determine whether this is the causative polymorphism in the L*er* allele. The *HEI10^{Cvi}*, *HEI10^{Col}*, *HEI10^{Ler}* and *HEI10^{R264G}* transgenes were transformed into Col-0/*420* FTL hemizygous plants using the GV3101

Agrobacterium tumefaciens strain through the floral dip method (Ziolkowski *et al*. 2017; Zhang *et al*. 2006).

All four *HEI10* T_1 populations showed significantly higher mean recombination levels than untransformed Col-0/*420* controls (increase relative to Col-0/*420* controls: $HEI10^{Col}$ 55%; $HEI10^{Ler}$ 16%; $HEI10^{Cvi}$ 45%; $HEI10^{R264G}$ 36%)($\chi^2(1)$ tests: $HEI10^{Col}$ p= 9.18 x 10^{-88}; $HEI10^{Ler}$ p= 1.11 x 10^{-8}; $HEI10^{Cvi}$ p= 2.95 x 10^{-43}; $HEI10^{R264G}$ p=3.42 x 10^{-42}, Table 44 and Figure 51). However, reduced numbers of replicates resulting from low transformation efficiencies prevented a full comparison of allelic effects. Large amounts of variation were observed within T_1 construct populations in the original experiment (Ziolkowski *et al*. 2017), which was attributed to variation in transgene copy numbers and insertion position effects affecting the level of expression. This meant that a large number of T_1s needed to be measured for recombination for each construct, to allow an accurate comparison of effects. Fifty replicates per construct were used for the original $HEI10^{Col}$ and $HEI10^{Ler}$ comparison (Ziolkowski *et al*. 2017), but here fewer replicates were generated: $HEI10^{Col}$ n=20, $HEI10^{Ler}$ n=12, $HEI10^{R264G}$ n=26 and $HEI10^{Cvi}$ n=9 (Table 44). Considerable variation was observed within T_1 populations (Standard Deviation: $HEI10^{Col}$ 6.9; $HEI10^{Ler}$ 3.4; $HEI10^{Cvi}$ 5.9; $HEI10^{R264G}$ 7.6), supporting the need for additional data.

Preliminary data showed that the $HEI10^{Ler}$, $HEI10^{Cvi}$ and $HEI10^{R264G}$ T_1 populations all differed significantly from the $HEI10^{Col}$ T_1 population recombination measurements (mean decrease relative to $HEI10^{Col}$ mean: $HEI10^{Ler}$ 25%; $HEI10^{Cvi}$ 7%; $HEI10^{R264G}$ 13%) ($\chi^2(1)$ tests: $HEI10^{Ler}$ p= 5.11 x 10^{-43}; $HEI10^{Cvi}$ p= 7.54 x 10^{-3}; $HEI10^{R264G}$ p=2.41 x 10^{-17}). This indicates that HEI10 is a recombination modifier that varies between the Col-0 and Cvi-0 accessions, and that the R264G variant is likely involved in this variation. However, while the mean *420* recombination rates of the $HEI10^{Col}$ and $HEI10^{Ler}$ T_1 populations differed by 7.2 cM, the $HEI10^{Cvi}$ T_1 population mean was only 1.9 cM lower than $HEI10^{Col}$, suggesting that despite sharing many polymorphisms, Cvi-0 and L*er*-0 *HEI10* alleles may have differing effects on recombination. Additional T_1 measurements would need to be made to verify this theory, as only nine $HEI10^{Cvi}$ T_1s were analysed.

Furthermore, while the $HEI10^{R264G}$ T_1 population was significantly different from the $HEI10^{Col}$ T_1 population, the difference in mean *420* recombination rate of 3.6 cM was less than that observed between $HEI10^{Col}$ and $HEI10^{Ler}$. Again, additional measurements are required to corroborate this, but initial data suggests that while the R264G variant has a role in variation in recombination, it may not be solely responsible for differences observed between $HEI10^{Col}$ and $HEI10^{Ler}$ alleles. The effect observed in $HEI10^{R264G}$ T_1 lines is similar to that observed in $HEI10^{Cvi}$ T_1s, suggesting that this may be the variant underlying differences between $HEI10^{Col}$ and $HEI10^{Cvi}$ T_1s, but the disparity between

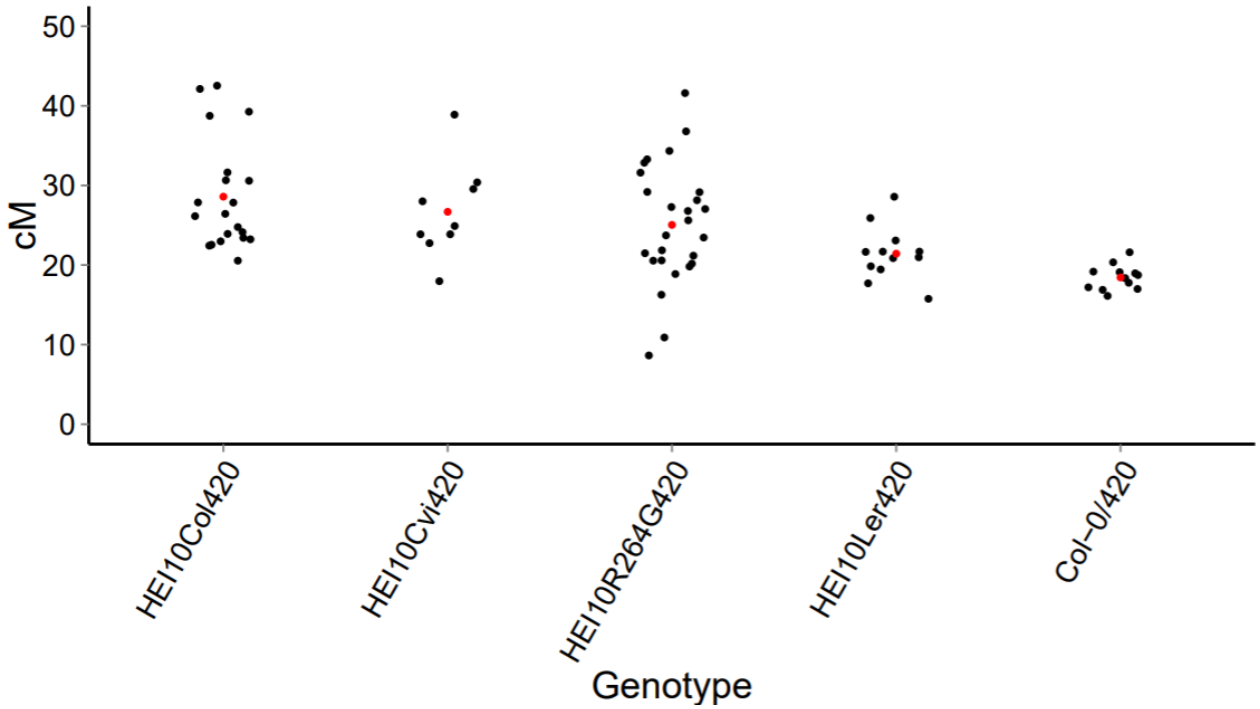


Figure 51: Genetic distance (cM) as measured in the *420* seed-based FTL-interval for Col-0/420 F_1 lines and *HEI10* T_1 lines. Measurements were made in Col-0/*420* F_1 individuals, and in *HEI10^{Col}420, HEI10^{Ler}420, HEI10^{Cvi}420* and *HEI10^{R264G}420* T_1 individuals (Table 44: Total 123154 seeds). Replicates are in black, with the mean of each genotype highlighted in red. Differences between Col-0/*420* lines and *HEI10* T_1s are significant for all four constructs (*HEI10^{Col}420* $\chi^2(1) = 395.7708$, $p=9.18 \times 10^{-88}$; *HEI10^{Ler}420* $\chi^2(1) = 33.98373$, $p=1.11 \times 10^{-8}$; *HEI10^{Cvi}420* $\chi^2(1) = 191.5251$, $p=2.95 \times 10^{-43}$; *HEI10^{R264G}420* $\chi^2(1) = 186.644$, $p=3.44 \times 10^{-42}$).

HEI10^{Ler} and *HEI10*^{R264G} T_1s suggests that variation in L*er* may be due to additional variants acting in combination with the R264G polymorphism (Figure 46A). However, this theory is speculative until further T_1 lines can be generated to substantiate the observed differences in recombination.

5.7 Observation of increased crossover-coincidence in *HEI10* overexpressor lines

Substantial increases in recombination were observed by Ziolkowski *et al*. (2017) across the genome in *HEI10* overexpressor lines (Figure 45), with the overall number of crossovers in the C2 *HEI10*^{Col} T_1 line being double the wild-type Col-0 count as measured by MLH1 foci counts and genotyping-by-sequencing of F_2 populations (Ziolkowski *et al.* 2017; Serra *et al.* bioRxiv). While these crossovers still followed a non-Poisson distribution, the distance between recombination events was reduced, indicating an increase in crossover coincidence that could be due to a reduction in crossover interference compared to wild-type plants (Serra *et al*. bioRxiv). This observation was unexpected as HEI10 acts in the Class I crossover pathway and these ZMM-dependent events are known to exhibit crossover interference (Chelysheva *et al*. 2012; Mercier *et al.* 2015).

These observations by Serra *et al*. (bioRxiv) were made in segregating F_2 populations, which preclude an analysis of crossover interference as individuals are produced from a combination of two independent meioses, and it is not possible to determine if double-crossover (DCO) events occurred on the same chromosome (in *cis*) or on the homologous chromosome (in *trans*). As crossover interference is believed to be propagated as a physical signal along the chromosome (Kleckner *et al*. 2004), only analysis of DCOs in *cis* are informative for estimation of its effect. Therefore, while general estimates about the average distance between crossover events can be made from F_2 populations, analysis of the distance between two events on the same chromosome in the same meiosis is required for calculation of crossover interference (Serra *et al.* bioRxiv).

Three-colour FTL intervals can be used to measure recombination in two adjacent intervals in the genome on the same chromosome, allowing estimation of crossover interference (Yelina *et al*. 2013). Three genetically linked transgenes that express different colours of fluorescent protein (eYFP, dsRed and eCFP) define two intervals on a chromosome, which can be measured for recombination between the transgenes (Yelina *et al*. 2013; Francis *et al.* 2007). Analysis of inheritance of fluorescence of all three colours in pollen using flow-cytometry can be used to estimate meioses that resulted in no recombination, a single crossover event in one interval, or recombination events in both intervals (double-recombinants, Figure 11) (Yelina *et al*. 2013; Ziolkowski *et al*. 2015).

In a pair of intervals where the central transgene is for yellow fluorescence, identification of pollen grains in the following eight fluorescence classes from a plant with hemizygous fluorescence

transgenes in a *cis*-configuration can be used to calculate the proportion of double-crossover events across this region. 'RYB' and '---' are parental/non-recombinant classes, 'RY-', '-YB', 'R—' and '—B' denote single-crossover events in either interval, and -Y- and R-B classes result from a double crossover event (where R is the dsRed transgene, Y is the eYFP transgene, B is the eCFP transgene and – is absence of transgene). The genetic distance (cM) of the first interval (RY) can be calculated using pollen counts for each fluorescence class with the formula 100*(R--+-YB+R-B+-Y-)/T, where T is total pollen. The genetic distance of the second interval (YB) can be calculated as 100*(-Y-+RY-+--B+R-B)/T. The observed number of double-recombinants is the sum of the -Y- and R-B fluorescence classes. The expected number of DCOs in the absence of interference can be calculated from the genetic distances of the two intervals as: ($cM_{RYinterval}$/100)*($cM_{YBinterval}$/100)*T. The ratio of observed DCOs to expected DCOs (O/E) gives the coefficient of coincidence (CoC), which can then be used to calculate the level of crossover interference operating between the two adjacent intervals (1-CoC) (Yelina *et al*. 2013; Ziolkowski *et al*. 2015).

5.8 Analysis of crossover interference in *HEI10* overexpressor lines using three-colour flow cytometry

In this experiment, I attempted to clarify whether the effect of *HEI10* on crossover distribution was due to an effect on crossover interference by measuring recombination using three-colour flow cytometry, thereby avoiding the analysis issues demonstrated by other experiments measuring the distance between crossovers in F_2 populations (Serra *et al.* bioRxiv). The *HEI10* overexpressor line C2 was crossed to the *I3bc* FTL which measures recombination in two adjacent intervals in the sub-telomere of chromosome 3, where recombination is predicted to increase with higher *HEI10* dosage (Ziolkowski *et al.* 2017; Serra *et al*. bioRxiv). A significant increase in recombination was observed for both the *I3b* and *I3c* intervals compared to wild-type (*I3b* 84% increase relative to Col-0/*i3b* F_1, *i3c* 114% increase relative to Col-0/*i3c* F_1, Figure 52, Tables 45 and 46, *i3b* cM $\chi^2(1) = 23575.54$, $p<2.2 \times 10^{-16}$, *i3c* cM $\chi^2(1) = 10549.36$, $p<2.2 \times 10^{-16}$). Calculation of the coefficient of coincidence from the ratio of observed to expected DCOs (Table 46) revealed a much higher co-occurrence of crossovers in both intervals compared to wild type. Subsequent calculation of crossover interference (1-CoC), revealed a significant decrease in interference in *HEI10* overexpressor lines (two-sample t-test assuming unequal variances, $p=2.66 \times 10^{-6}$), from a mean of 0.64 in wild-type replicates to a mean of 0.34 in *HEI10* overexpressor lines (Figure 52), consistent with estimates of distances between DCOs made from F_2 data (Serra *et al*. bioRxiv).

While this increase in coincidence of crossovers does not equal that observed in *recq4a recq4b* mutants, which show an increase in the Class II interference-independent crossover pathway (Seguela-

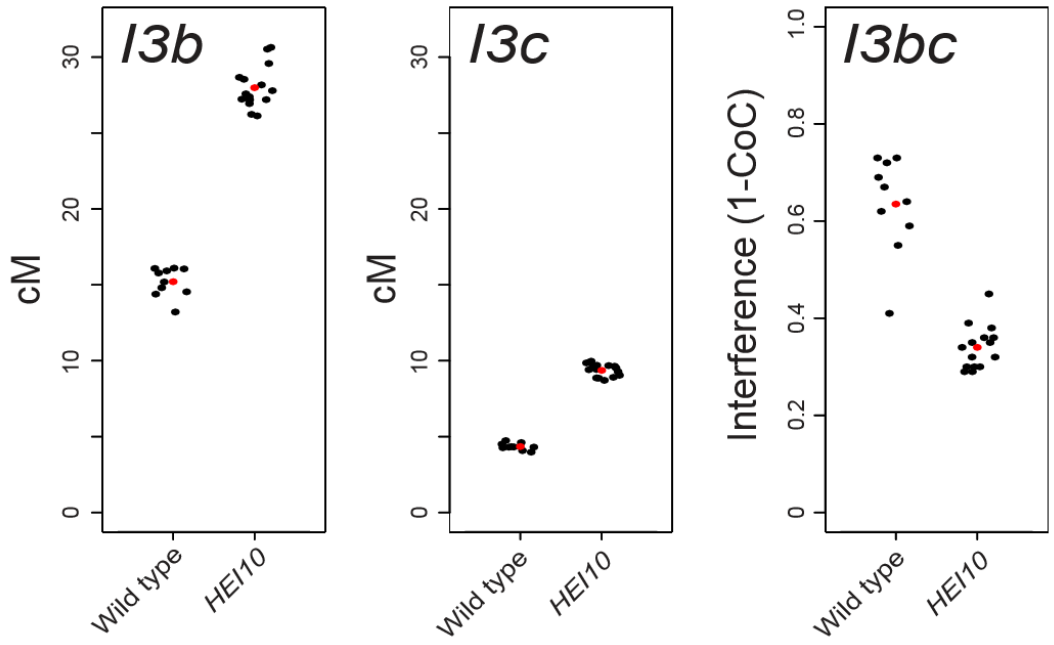


Figure 52. Comparison of wild-type and *HEI10* overexpressor genetic distance (cM) as measured in *I3b* and *I3c* pollen-based FTL-intervals, and comparison of crossover interference (1-CoC) between *I3b* and *I3c* intervals. Measurements were made in Col-0/*I3bc* and *HEI10* overexpressor/*I3bc* F_1 individuals (Table 45: Total 1153540 pollen grains), and crossover interference was calculated from the ratio of observed to expected DCO events in these two intervals (Table 46). Replicates are in black, with the mean of each genotype highlighted in red. Differences between wild-type and *HEI10* overexpressor lines are significant for both intervals (i3b cM $\chi^2(1) = 23575.54$, $p<2.2 \times 10^{-16}$, i3c cM $\chi^2(1) = 10549.36$, $p<2.2 \times 10^{-16}$), and for crossover interference (two-sample t-Test assuming unequal variances, $p=2.66 \times 10^{-6}$).

Arnaud *et al*. 2015; Seguela-Arnaud *et al*. 2016; Serra *et al*. bioRxiv), it is still a significant effect caused by an increase in dosage of a gene involved in the Class I crossover pathway (Chelysheva *et al.* 2012). The mean interference value of 0.34 in the *HEI10* overexpressor lines suggests that some interference is still acting, but that this is considerably lower than that found in the wild-type. However, the mechanism underlying the HEI10-induced reduction of interference is unclear.

HEI10 is believed to promote class I crossover repair at recombination intermediates through SUMO or ubiquitin transfer to unknown targets (Serra *et al*. bioRxiv). This would fit with what is known about HEI10 function in *Sordaria*, where it integrates signals from the synaptonemal complex (SC) and associated recombination complexes to modulate development of crossovers via SUMOylation or ubiquitination functions (De Muyt *et al*. 2014). This link to the SC, which includes chromosomal axis proteins, could be relevant to the apparent effect of HEI10 on crossover interference, as interference is modelled as a mechanical force transmitted by the chromosome axis or SC (Kleckner *et al*. 2004). The chromosome axis or SC could be SUMO/ubiquitin targets of HEI10, providing a mechanism for HEI10 to potentially alter transmission of the interference signal.

Modification of interference via SUMOylation of the axis or SC would be consistent with evidence from yeast that axis protein Red1 (homolog of ASY3 in Arabidopsis) and axis-associated topoisomerase II are SUMOylated, and that loss of this SUMOylation through mutation of their SUMOylation sites affects crossover interference (Zhang *et al.* 2014; Cheng *et al.* 2006). Experiments indicate that ubiquitin-mediated removal of SUMOylated proteins is also required for wild-type interference levels, although it is unclear if these processes act on the same proteins, and if so, whether they are sequential in one pathway or in competition with each other (Zhang *et al.* 2014). The SUMOylation of Red1 in yeast appears to be dependent on the Zip3 SUMO E3 ligase (Cheng *et al.* 2006; Eichinger and Jentsch 2010), which is structurally and functionally related to the Arabidopsis HEI10 protein (Chelysheva *et al.* 2012). If HEI10 targets are similar to those found for Zip3 in yeast, it is plausible that SUMOylation or ubiquitination of axis proteins by HEI10 could form the basis of the observed influence of HEI10 over interference in *HEI10* overexpressor lines. Alternatively, HEI10 action at recombination intermediates may modify recombination complexes and decrease their sensitivity to the interference signal (Serra *et al*. bioRxiv). Future work identifying the SUMOylation/ubiquitination targets of HEI10 in Arabidopsis could provide additional information about how HEI10 influences crossover interference.

5.9 Conclusions

Fine-mapping of rQTL1^{420} identified a broad region of chromosome 1 as the likely location of the recombination modifier. Analysis of the rQTL1^{420}NIL demonstrated that presence of this region in the Cvi-0 genotype in a predominantly Col-0 genetic background had a significant effect on recombination in the *420* sub-telomeric interval, but not the *5.11* centromeric interval. However, this effect was not as large as that found in earlier mapping lines, suggesting that additional factors may act in combination with rQTL1 to affect recombination. Preliminary analysis of the *HEI10* recombination modifier as a candidate for rQTL1^{420} showed that Col-0 and Cvi-0 alleles of *HEI10* have significantly different effects on recombination. Additional analysis of an amino acid swap allele supported the theory that the R264G variant found in L*er* and Cvi-0 alleles of *HEI10* is linked to variation in recombination, although quantification of its effect suggests additional variants may also contribute to accession-specific differences in recombination. Characterisation of the *HEI10* dosage-dependent effect on recombination revealed that increasing expression results in a reduction in crossover interference, although the molecular mechanism underlying this phenomenon is currently unknown. Future work should focus on confirmation of the *HEI10*^{Cvi} effect on recombination and identification of targets that could provide insight into its molecular mechanism of recombination modification.

Chapter 6 - General Discussion

6.1 Summary

Previous studies of *Arabidopsis thaliana* accessions had revealed significant variation in crossover frequency (Lopez *et al*. 2012; Esch *et al*. 2007; Sanchez-Moran *et al*. 2002; Salome *et al*. 2012; Ziolkowski *et al*. 2015). This suggested the presence of natural modifiers of recombination that varied between accessions, altering their recombination profiles. Identification of these natural modifiers could not only benefit understanding of recombination mechanisms in general, but also understanding of the role of recombination and its potential effect on natural selection and adaptation in natural populations (Bauer *et al.* 2013; Wijnker *et al.* 2013; Hill and Robertson 1966; Roze and Barton 2006). QTL mapping in accession-cross populations has revealed several significant modifiers of recombination frequency that vary between Cvi-0, Col-0 and Can-0 accessions. Subsequent experiments identified potential candidate loci underlying these modifiers, and confirmed that the *HEI10* E3 ubiquitin ligase polymorphisms associate with variation in recombination between Col-0 and Cvi-0 accessions. Further analysis of *HEI10* revealed a role in crossover interference, which may be related to its function in influencing recombination frequency across the genome.

6.2 Identification of modifiers of meiotic crossover frequency in *Arabidopsis thaliana* accessions

Measurement of recombination in accession-FTL F_1 lines through observation of fluorescence-segregation in the products of meiosis showed significant variation both between accession crosses, and between different intervals of the genome in the same accession cross (see Figures 14, 15 and 16). Genetic distances of 11.8 cM and 11.1 cM in the *420* FTL interval in Cvi-0/*420* F_1 and Can-0/*420* F_1 lines compared to the Col-0/*420* F_1 mean of 18 cM demonstrated that large effect modifiers varying between the Cvi-0 and Can-0 accessions and Col-0 were acting on recombination in the subtelomere of chromosome 3 (see Figure 14 and Table 5). Analysis of additional FTL intervals revealed that recombination was increased in Cvi-0/FTL crosses measuring recombination across centromeric intervals, but decreased across euchromatic sub-telomeric intervals compared to Col-0/FTL homozygotes, suggesting that modifiers may exhibit region-specific effects on recombination (see Figure 15). This is consistent with observations made in multiple accession-FTL crosses (Ziolkowski *et al*. 2015), where individual accessions frequently exhibited differential increases and decreases in different intervals compared to Col-0/FTL measurements. Region-specific modifiers have also been observed in other species, including *Schizosaccharomyces pombe*, *Neurospora crassa* and *Zea mays* (De Veaux *et al*. 1994; Catcheside 1977; Yandeau-Nelson *et al*. 2006), suggesting that this is a wide-spread phenomenon. The reason for this effect is unclear, although it is likely due, at least in part, to the interaction of *trans*-acting modifiers with the local *cis*-context of sequence and chromatin

features. This could have evolved as a mechanism causing modifiers to specifically promote recombination in euchromatic gene-rich areas of the genome, thereby having a direct effect on genetic-linkage of loci and limiting potentially deleterious recombination in centromeric repeats (Salome *et al*. 2012; Parket *et al*. 1995).

Interestingly, large increases in recombination across centromeric intervals in Cvi-0/FTL F_1 lines compared to Col-0/FTL homozygotes result in Cvi-0/FTL lines showing a higher overall level of recombination across the genome when interval measurements are summed (Tables 5, 6, 7, 8, 9 and 10. Sum of the means of the six intervals measured). When looking only at male-specific recombination in pollen-FTLs, the same effect is observed, suggesting that male recombination rates are higher in Cvi-0/FTL lines. This contrasts with data obtained from chiasma counts in pollen-mother cells (PMCs), where Cvi-0 measurements were lower than those observed in Col-0 (Lopez *et al*. 2012). There are several possible reasons for the observed disparity in recombination estimates, including region-specific effects, technical issues, general heterozygosity effects and combined effects of *trans*-modifier alleles resulting in an increase in recombination in hybrids. It is possible that various factors could be having a region-specific effect on recombination whereby, overall, male recombination is low in Cvi-0, but it is high in *CEN3* and *I2f* intervals. Alternatively, as chiasma counts were measured in homozygous Cvi-0 lines, in contrast to FTL-estimates made in Col-0/Cvi-0 hybrids, this may also be consistent with alleles of modifiers from different accessions combining to increase recombination in hybrids. For example, if the Cvi-0 allele of one modifier was associated with higher recombination than the Col-0 allele, while the Cvi-0 allele of a second modifier was associated with lower recombination, the homozygous Cvi-0 line would have a balanced recombination rate. However, in a Col-0/Cvi-0 hybrid the alleles associated with higher recombination would be present for both modifiers and, depending on dominance and the relative effects of each modifier, their combination could result in an increase in recombination in the hybrid relative to homozygous Cvi-0 lines. A similar effect was observed for two modifiers affecting the *420* interval that varied between the Col-0 and L*er*-0 accessions, where the Col-0 alleles of each modifier had opposite effects on recombination frequency (Ziolkowski *et al.* 2017). The combination of alleles in the Col-0/L*er*-0 hybrid resulted in a lower level of recombination within the *420* interval than was observed in the homozygous Col-0 line, due to the semi-dominance of alleles for the first modifier where the Col-0 allele was associated with higher recombination, and the dominance of the Col-0 allele for the second modifier, which was associated with lower recombination (Ziolkowski *et al.* 2015; Ziolkowski *et al.* 2017).

It is also possible that the presence of heterozygosity itself across the hybrid genome could be affecting recombination in the FTL measurements. There is some evidence that recombination increases in the heterozygous backgrounds of Arabidopsis F_1 hybrids compared to homozygous

parental accessions (Barth *et al.* 2001, Ziolkowski *et al*. 2015), however the variation in recombination frequently does not correlate with the level of polymorphism between parental lines (Salome *et al*. 2012; Ziolkowski *et al*. 2015). This suggests that the increase in recombination observed in Col-0/Cvi-0 hybrids is unlikely to be a product of the heterozygosity itself but is more likely to be a result of interactions between the combined Cvi-0 and Col-0 *cis* and *trans* modifiers of recombination. There is also a small possibility that the difference between PMC chiasma counts and FTL measurements could be due to the difference in technique. While chiasmata are frequently counted at the metaphase I stage of meiosis in PMCs (Lopez *et al*. 2012), before any chromosome segregation, measurements made using FTLs require viable pollen or seed expressing the fluorescent protein. This could create a bias against cells with especially low levels of recombination that may not ensure the obligate crossover between each pair of homologous chromosomes, and also cells with significantly altered crossover distributions that may result in entanglements or segregation difficulties. This selection for viable meiosis products could create a bias for cells with higher recombination rates or particular crossover distributions, which could skew the subsequent recombination estimate for the line, thereby making chiasma counting and FTL-scoring difficult techniques to compare. While this bias would be expected to have a more considerable effect on mutant lines with severely reduced recombination rates and subsequent segregation and viability issues, such as *zmm* mutants (Higgins *et al.* 2004; Mercier *et al.* 2015), it is possible that it could result in a slightly artificially raised FTL-based recombination estimate in other lines, including natural accessions.

QTL mapping in segregating accession/FTL F_2 populations identified several significant modifiers of meiotic recombination frequency varying between Col-0 and Cvi-0 accessions, and one significant modifier varying between the Col-0 and Can-0 accessions (Figures 17C, 20C and 22C). Despite having similar recombination patterns across FTL intervals measured in F_1 lines (Ziolkowski *et al*. 2015), including the *420* interval (Figure 14), Cvi-0 and Can-0 accessions contained different modifiers affecting recombination in the *420* interval. This is consistent with a previous analysis made of their genome sequences (Alonso-Blanco *et al*. 2016), as both accessions were characterised as 'relicts' due to their extreme sequence differences both from each other, and from the other 1133 accessions analysed. These 'relict' populations are believed to be the result of earlier expansions in habitat range that became isolated populations by surviving climate change in glacial refugia, resulting in genomes that have diverged significantly from other natural populations due to minimal gene flow (Alonso-Blanco *et al.* 2016; Durvasula *et al*. 2017). Notably, this genome analysis demonstrated that Can-0 and Cvi-0 are substantially different from each other, and are therefore unlikely to share variants in recombination modifiers that differ between accessions (Alonso-Blanco *et al*. 2016).

Both accessions appear to have independently evolved similar recombination patterns, including decreased recombination in the sub-telomeric *420* interval and an overall increase in recombination across measured intervals compared to Col-0, using different molecular mechanisms, as the modifiers underlying these patterns are not the same. As the development of similar CO patterns does not appear to be the result of a close genetic relatedness, it could reflect adaptation to a common environment, as both accessions are found in arid disturbed habitats. There is evidence to suggest that there may be an advantage to increased recombination in disturbed habitats, as changing conditions may create a requirement for rapid adaptation to new environments (Charlesworth *et al*. 1976; Hadany *et al*. 2008; Webster and Hurst 2012; El-Soda *et al.* 2014). Recombination can influence the rate of adaptation by affecting the efficacy of natural selection (Morrell *et al*. 2004; Marais *et al*. 2004). CO events can reduce the linkage between variants that may have different effects on fitness, such as one being deleterious and the other advantageous under current conditions, allowing selection to act on each variant individually, thereby improving the efficiency of natural selection and increasing the rate of adaptation (Marais and Charlesworth 2003; Morrell *et al*. 2004; Roze and Barton 2006; Webster and Hurst 2012). Increases in the rate of recombination would result in a concomitant decrease in Hill-Robertson interference, allowing the rapid generation of new variant combinations and promoting the progression of adaptation, which could be advantageous in disturbed habitats that are subject to change (Hill and Robertson 1966; Barton and Charlesworth 1998; Webster and Hurst 2012).

However, it is also possible that the increase in recombination is not the result of selection acting on the recombination phenotype, but rather a by-product of selection acting on an unrelated trait. Recombination is known to increase as a by-product of selection on other traits, as seen in domesticated animals, and it is possible that the overall increase in recombination observed in the Cvi-0 and Can-0 accessions, compared to many other accessions, is the result of selection acting on a different aspect of meiosis (Otto and Barton 2001; Smukowski and Noor 2011; Webster and Hurst 2012; Bomblies *et al*. 2015; Ziolkowski *et al*. 2015). For example, both accessions grow in regions with high ambient temperatures, and as high temperatures are known to affect the stability of chromosome axis proteins and the SC, it is possible that this results in increased selection acting on these proteins to improve their thermotolerance (Wright *et al*. 2015; Morgan *et al*. 2017; Lloyd *et al.* 2018). As the structure of the chromosome axis and SC is also known to influence recombination, selection for increased thermotolerance, which could affect axis formation and synapsis, may result in an indirect alteration of recombination patterns (Sanchez-Moran *et al*. 2007; Ferdous *et al*. 2012). Analysis of recombination in accessions from a range of habitats could determine whether there are

clear associations between environmental factors and recombination rates in *A.thaliana*, and therefore clarify which selection pressures are likely to underlie variation in recombination.

While significant variation in recombination was observed in F_2 mapping populations measuring recombination across centromeric FTL-intervals (Figures 17A and 25A), a large proportion of this variation was not attributable to segregating large-effect modifiers. The Cvi-0/*CEN3* F_2 population, which showed a 33.1 cM range in recombination rates, revealed four significant large-effect modifiers (Figure 17C). However, these only explained 60.8% of the variance in the population and explained much less of the variance individually than modifiers identified in the Cvi-0/*420* F_2 population (Figure 20C), meaning that considerable effects were likely contributed by variation in small-effect loci and *cis*-acting factors. These additional factors had an even larger effect in the Cvi-0/*5.11* F_2 population, where a range of 14.1 cM was observed, yet no significant modifiers were identified (Figure 25C). Some of this unexplained variation can probably be attributed to background variation, as differences in recombination frequency were observed between genetically identical replicates of Cvi-0/*CEN3* and Cvi-0/*5.11* F_1 lines (Figure 16). However, the variation between F_1 replicates was considerably smaller than the amount of unexplained variation observed in the F_2 populations, particularly for the Cvi-0/*5.11* cross, suggesting that most of this variation cannot be explained by background variation or minor technical variances and is therefore likely to be caused by differences in *cis*-acting factors and small-effect loci.

While small effect loci and *cis*-effects presumably influence recombination across the genome, this effect appears to be particularly pronounced in centromeric intervals, and the reason for this is unclear, though it may relate to epigenetic differences. Substantial differences between Col-0 and Cvi-0 accessions have been observed in heterochromatin levels, centromeric repeat numbers and maintenance of DNA methylation (Tessadori *et al*. 2009; Ito *et al*. 2007; Pignatta *et al*. 2014). Yelina *et al*. (2015) observed that segregating *met1* and *ddm1* methylation mutant populations had more variable recombination frequencies, consistent with stochastic epigenetic divergence. Therefore, it is possible that variation in DNA methylation levels, or other heterochromatic marks such as $H3K9^{me2}$, between Cvi-0 and Col-0, could contribute to variation in recombination in hybrids. This effect would be likely to be more pronounced in centromeric intervals, due to higher levels of DNA methylation and heterochromatin associated histone marks (Yelina *et al.* 2012; Yelina *et al.* 2015; Yao and Schnable 2005) and could account for some of the variation observed in Cvi-0/*5.11* and Cvi-0/*CEN3* F_2 populations. However, comparison of methylation profiles and recombination rates in individuals from segregating populations would need to be performed before any conclusions could be drawn about this effect.

Aside from potential variation in *cis*-acting factors, a possible alternative explanation for the variation observed in F_2 populations measuring recombination across centromeric intervals is that it is caused by multiple segregating small-effect *trans*-acting modifiers that are not sufficiently significant to be detected by QTL mapping in these populations (Broman and Sen 2009; Mitchell-Olds and Schmitt 2006). Sequence variants in modifiers affecting recombination across centromeric intervals may be under stronger selection pressures than those in genes affecting sub-telomeric recombination, due to the potential for more deleterious consequences if centromeric recombination is altered (Ellermeier *et al*. 2010). It is possible that random variants having a large effect on recombination frequency would be more likely to be strongly selected against if they affected recombination at the centromere, than if they acted at the sub-telomere. As a consequence of this selection pressure, variants affecting recombination around the centromere that had moved to fixation in an accession population would be more likely to have small effects on CO frequency, and therefore be harder to identify by QTL mapping. The variation in F_2 populations measuring centromeric recombination could therefore be due to a cumulative effect of many small-effect modifier variants, as opposed to the few large-effect variants that explain most of the variation in the Cvi-0/*420* F_2 population (Mitchell-Olds and Schmitt 2006). This hypothesis is consistent with the Cvi-0/Col-0 populations analysed here, measuring recombination across both centromeric and sub-telomeric intervals, but mapping in additional populations and different accessions would be required to confirm this trend. Further information about natural accession populations and selection pressures affecting recombination would also be required to determine whether this explanation is credible.

Another feature that was apparent in some F_2 populations but did not appear in others was the effect of *cis*-acting heterozygosity (Ziolkowski *et al* 2015). This effect was found to influence recombination in Cvi-0/*CEN3* F_2 and Can-0/*420* F_2 populations but was not apparent in the Cvi-0/*420* F_2 population (Figures 17C, 20C, 22C and 24). The reason for this disparity is unclear, although it is likely to relate to differences between the centromere (*CEN3*) and subtelomere (*420*) in the Cvi-0 background. Potential differences at the centromere in Cvi-0 in the activity of mismatch repair proteins, or in the relative action of class I and class II recombination pathways (Anderson *et al.* 2014), could cause this effect, as the class II (non-interfering) crossover repair pathway has been demonstrated to be less efficient in heterozygous regions of the genome (Ziolkowski *et al*. 2015). Experiments in tomato suggest that patterns of class I and class II COs may differ along chromosomes, with class II events being disproportionately represented in pericentric heterochromatin (Anderson *et al.* 2014). However, no evidence of such differences in Arabidopsis is currently known, and there is no indication that the patterns of these events differ between accessions either, making this theory extremely speculative.

In summary, significant variation in meiotic recombination was observed between Arabidopsis accessions and in different regions of the genome. Using this information, significant *trans*-acting modifiers of recombination frequency were identified that varied between natural accessions. The influence of additional factors, including *cis*-acting heterozygosity and epigenetic variation, may also contribute, although further experiments are required to analyse these effects. Identification of significant recombination QTL led to further experiments designed to determine the effects of individual modifiers and identify candidate loci and polymorphisms.

6.3 Fine-mapping genetic modifiers of meiotic recombination frequency and identification of candidate genes.

QTL identified in segregating accession-cross populations were backcrossed into the Col-0 genetic background to isolate their effects from other large effect loci, small effect loci and *cis*-acting influences on recombination. This confirmed that rQTL2^{420}, a modifier identified on chromosome 2 in the Cvi-0/*420* F_2 population, had a significant effect on recombination in the *420* interval independent of other QTL (Figure 30). However, while the Cvi-0 allele of rQTL2^{420} caused an average increase of 6.7 cM in *420* over Col-0/*420* homozygous lines, it did not cause a significant effect on recombination in the centromeric *5.11* FTL interval, suggesting that this modifier has region-specific effects. Analysis of the Cvi-0 allele of rQTL5^{420} isolated in the Col-0 background also revealed a significant effect on *420* recombination (a decrease of 6.4 cM compared to Col-0/*420* homozygotes). However, rQTL5^{420} was also shown to have a significant effect on recombination in the centromeric interval *5.11*, resulting in a 2 cM decrease when compared to Col-0/*5.11* F_1 controls (Figure 38), suggesting that this modifier may have a more general effect on recombination across the genome.

Identification of individual plants with recombination events in the regions surrounding the QTL allowed refinement of credible intervals from megabase to kilobase scales. Fine-mapping and marker-assisted selection using PCR-based genotyping markers (Figure 27) narrowed the rQTL2^{420} interval to 239 kb, a region containing 87 loci which could potentially be responsible for the observed effect on recombination. This candidate gene list was further reduced using criteria including expression profile, presence of non-synonymous polymorphisms (constituting an amino acid change) and potential regulatory polymorphisms between Col-0 and Cvi-0 accessions, and gene function. While this strategy identified 10 genes that were considered to be the strongest candidates for the QTL, it was still possible that one of the eliminated loci was the causal locus underlying rQTL2^{420}. The complexity of the recombination process means that while a gene may not have an obviously related function, it may have an indirect effect, and therefore removal of candidates based on gene function risks removal of potential bona fide modifiers.

Ideally, additional recombination events would have been used to narrow the QTL intervals further, but the limited recombination observed within this interval means this would require screening of a fine-mapping population of thousands of individuals to identify informative recombinants. Given the labour-intensive nature of screening large populations, and space constraints for plant growth, further fine mapping was not performed. Instead, the ten strongest candidate genes identified were assessed using T-DNA insertion mutants crossed to the FTL-reporter system to determine whether they affected recombination frequency. If no effect was observed in the lines assessed, additional T-DNA lines could be analysed, or further fine-mapping individuals could be screened.

This strategy identified the Salk 082541 line, which has an insertion in the *ARI7* gene (At2g31510), as having a significant effect on recombination when crossed to the *420* reporter line (Figure 32). ARI7 is a RING/U-box superfamily protein with a putative nucleic acid binding function (TAIR, Berardini *et al*. 2015). While it is unknown whether *ARI7* is expressed during meiosis, it is known to be expressed in flowers, inflorescence meristems and plant embryos (Araport11, Krishnakumar *et al*. 2015), and is predicted to localise to the nucleus as it contains a nuclear localisation sequence (Mladek *et al.* 2003). Based on homology to ARIADNE proteins in Drosophila and mammals, Arabidopsis ARI7 is predicted to have a function in protein ubiquitination (Mladek *et al.* 2003), meaning that it may influence recombination via regulation of the stability of meiotic proteins, as has been demonstrated for HEI10 and RNF212 RING-domain containing proteins in mammals (Qiao *et al.* 2014; Rao *et al*. 2017). While ARI7 has a different protein structure to HEI10 and RNF212, all three proteins contain an N-terminal RING domain (Mladek *et al.* 2003; Chelysheva *et al.* 2012; Qiao *et al.* 2014). This domain in ARIADNE proteins in mammals has been shown to interact with E2 ubiquitin-conjugating enzymes (E2 UBCs)(Moynihan *et al.* 1999; Mladek *et al.* 2003). The interaction allows the ARIADNE proteins to accept ubiquitin from the E2 UBCs, which is necessary for their function as ubiquitin protein-ligases that subsequently perform the transferral of ubiquitin to substrate proteins (Mladek *et al.* 2003; Deshaies and Joazeiro 2009). This suggests that the RING domain is likely to be involved in the possible ubiquitination function of ARI7 (Mladek *et al.* 2003). Alternatively, the predicted U-box domain of ARI7 (TAIR, Berardini *et al*. 2015), which is structurally related to the RING-domain, could also function in mediating ubiquitination (Deshaies and Joazeiro 2009).

If the potential effect of *ARI7* on recombination is mediated by its possible ubiquitination function, polymorphisms between the Col-0 and Cvi-0 accessions in the predicted RING or U-box domains could alter ubiquitination function, and thereby affect the influence of *ARI7* on recombination. However, the Cvi-0 specific polymorphisms identified within *ARI7* by the 1001 genomes project are not present in either of these domains (Alonso-Blanco *et al*. 2016). Therefore, while sequencing of the allele of *ARI7* from the Cvi-0 line used for the QTL mapping population may possibly reveal additional

polymorphisms, it is also possible that these domains are not responsible for the effect that ARI7 appears to have on recombination. Sequence information from the 1001 Genomes Project (Alonso-Blanco *et al*. 2016) indicates that there are Cvi-0 specific polymorphisms near a splice junction in the final intron of *ARI7*. This could result in an alteration of splicing in Cvi-0, which may significantly alter protein function or localisation and subsequently the effect on recombination.

An examination was performed of the sequences of *ARI7* alleles from the same 25 accessions used by Ziolkowski *et al.* (2015) for analysis of the variation in recombination between accessions, which were chosen as a small but fairly geographically and genetically diverse group representative of wider variation (Alonso-Blanco *et al*. 2016, Figure 12, Supplementary Figure 1). This revealed that one of these intron polymorphisms – a guanine to thymine SNP – seems to be unique to Cvi-0 (Supplementary Figure 1). Additional Cvi-0 polymorphisms that varied from Col-0 in the region were also found in the Can-0 and Bur-0 accessions (Alonso-Blanco *et al.* 2016, Supplementary Figure 1). As rQTL2^{420} was not identified in QTL mapping populations generated from either Can-0/*420* or Bur-0/*420* crosses (Figure 22C and E. Lawrence 2016, personal communication), these polymorphisms were not considered likely to underlie the observed variation in recombination. Therefore, the G-to-T SNP was considered to be the probable causal variant underlying rQTL2^{420}, providing that *ARI7* is confirmed as the underlying gene. Analysis of the sequence of *ARI7* orthologs in related *Brassicacea* species *Arabidopsis lyrata* and *Arabidopsis halleri* showed that Col-0 contains the ancestral sequence for this region, which suggests that the Cvi-0 variants in this intron may be more recently derived (Supplementary Figure 1). Although both the *ARI7* gene and the causal variant would require confirmation before any assumptions could be made about the evolution of recombination, it seems probable that the Cvi-0 allele, and its associated phenotype of higher recombination, are more recent developments and therefore could potentially be the result of adaptation to the local environment.

Further work is required to confirm *ARI7* as the gene underlying rQTL2^{420}, particularly as a second At2g31510 T-DNA insertion line (Salk 027620), with an insertion in the last intron in close proximity to the putative causal Cvi-0 specific polymorphism, did not show an effect on recombination (Figure 32). This confirmation could be performed through transformation of transgenic *ARI7* alleles from each accession into the *420* reporter system to demonstrate Cvi-0 and Col-0 allele-specific effects on recombination. The semi-dominant effect of the Cvi-0 allele observed at the peak rQTL2^{420} genotyping marker suggests that if *ARI7* is the rQTL2^{420} modifier, transformation of the Cvi-0 allele would result in an observable increase in recombination. If *ARI7* was confirmed as the modifier, additional experiments analysing *ARI7* mRNA sequences in Cvi-0 and Col-0 genetic backgrounds could be performed to test whether the splice junction variant influences the mRNA that is expressed. Alternatively, other Cvi-0 polymorphisms, such as potential regulatory variants in the predicted *ARI7*

promoter region, could be explored to determine if they are responsible for allele-specific effects on recombination.

The same approach was utilised to identify and assess candidates for rQTL5^{420}, although in this case the interval in question was 213 kb wide and contained 31 loci. Reduction of this list, using the same criteria as previously, identified four strong candidate loci which were assessed for an effect on recombination using T-DNA insertion lines crossed to the *420* reporter system. A candidate gene was identified with a significant, although small, effect on recombination, which encodes a PP2C protein phosphatase (At5g53140, Kerk *et al*. 2002). The presence of Cvi-0 specific polymorphisms in the predicted promoter of the gene suggests a potential regulatory difference in the gene between Cvi-0 and Col-0 accessions (1001 Genomes Project, Alonso-Blanco *et al*. 2016). Of the five SNPs identified between Col-0 and Cvi-0, four are also present in the Can-0, L*er*-0, Bur-0 or Ct-1 accessions, or some combination thereof (1001 Genomes Project, Alonso-Blanco *et al*. 2016). As a LOD peak corresponding to rQTL5^{420} was not observed in any of the QTL mapping populations derived from crosses of these accessions to the *420* reporter line, it is not likely that these variants are responsible for the observed variation in recombination between Col-0 and Cvi-0 (Figure 22C; Ziolkowski *et al*. 2015; Ziolkowski *et al.* 2017; E. Lawrence 2016, personal communication). The remaining polymorphism, a thymine to cytosine SNP, is unique to Cvi-0 when the *PP2C* promoter sequence is compared to 24 other representative accessions (Supplementary Figure 1). This polymorphism was therefore considered the most probable candidate for a causal variant affecting recombination.

Although no orthologous sequence was identified in the other *Brassicacea* species analysed, which prevented identification of the ancestral variant, the presence of the thymine base in 96% of the accessions analysed suggests that the Cvi-0 variant at this position was probably more recently derived (JGI Phytozome BLAST, www.phytozome.jgi.doe.gov, 7th March 2018; Supplementary Figure 1). This interpretation of a more recent mutation moving to fixation in Cvi-0 could support the theory that the pattern of recombination in Cvi-0 is due to selection leading to the adaptation of genes to new conditions, although genetic drift of random variation is also possible. However, the effect of this variant on recombination, and of the *PP2C* gene in general, is yet to be confirmed, meaning that this preliminary analysis is speculative. While a regulatory difference between Col-0 and Cvi-0 is possible, this data was collected from analysis of a single T-DNA line with an insertion upstream of the promoter, therefore additional work is required to confirm the effect of this insertion on At5g53140 expression. Analysis of additional T-DNA lines disrupting the gene may also help to clarify whether this is the gene underlying rQTL5^{420}.

Adoption of the same strategy for identification of the locus underlying CanQTL4 had less success. Recombination around this QTL was restricted by proximity to the recombination-suppressed centromere. Furthermore, the late flowering phenotype of the Can-0 accession made generation and identification of recombinants challenging. A putative interval of 387 kb containing 94 loci was identified, and subsequent removal of transposons and non-coding repeat sequences in addition to application of the candidate gene criteria utilised for the Cvi-0 QTL gave a list of four candidate loci. TEs are known to induce formation of local heterochromatin (Caceres *et al*. 2001) and therefore indirectly influence recombination in *cis*. However, while it is possible that transposons could act as *trans* modifiers of recombination, we believed that variation in one of the coding genes was more probable.

Analysis of T-DNA insertion lines demonstrated that the GK-219G07 line, which contains an insertion in the *TPR8* carboxylate clamp-tetratricopeptide repeat gene (At4g08320), has a significant effect on recombination (Figure 44). However, the observed effect was small (a 1.4 cM increase over Col-0/*420* F_1 replicates), and analysis of a second insertion line within the gene (SAIL 731HO4) did not confirm this effect. As no potential functions in recombination are apparent (Araport11, Krishnakumar *et al*. 2015), and insertion mutants in the gene have a small effect compared to the amount of variation CanQTL4 is responsible for in the Can-0/*420* F_2 population, this candidate seems unlikely to be the locus underlying CanQTL4. However, observation of the effect of different genotype groups at the peak allele for CanQTL4 in the Can-0/*420* F_2 population (Figure 23B) indicates that the Can-0 allele may act recessively. As measurements were made in T-DNA/*420* F_1 lines where the T-DNA insertion was in a hemizygous state, it is possible that this allele also functions recessively and therefore would not demonstrate an effect on recombination unless it was present as a homozygous insertion. Therefore, generation of a homozygous insertion line containing the *420* reporter system and measurement of recombination could provide confirmation of TPR8's effect on recombination.

Additionally, it is possible that the Can-0 allele has a strong effect on recombination through an alteration of function or expression level, while knockout of the gene has minimal effect. Transformation of individual accession alleles into the *420* reporter system is required to conclusively assess whether *TPR8* is the gene underlying CanQTL4. As CanQTL4 appears to act recessively, transformation of the Can-0 allele into the *420* line would not show an effect on recombination. Consequently, transformation of the Col-0 and Can-0 alleles would have to be performed as complementation of the GK-219GO7/*420* knockout line to quantify the effect of accession alleles and confirm *TPR8* as the modifier underlying CanQTL4. If *TPR8* was confirmed as the gene underlying CanQTL4, experiments could be performed to identify the causal variant underlying the difference in recombination observed between Can-0 and Col-0 alleles. An understanding of how this variant affects

the gene, potentially through modification of a specific domain, could help to clarify the mechanism by which *TPR8* affects recombination. Examination of the sequence of *TPR8* alleles in 25 accessions revealed that the Can-0 allele contains many unique non-synonymous polymorphisms in exons that are not seen in the other accessions (1001 Genomes Project, Alonso-Blanco *et al.* 2016). The large number of Can-0 specific non-synonymous polymorphisms that result in an amino acid change, and their distribution across five exons, makes it difficult to identify the likely causal variant.

The lack of a probable candidate for the causal variant also makes it difficult to determine whether the Can-0 TPR8-associated recombination phenotype is likely to be ancestral or more recently derived. For example, comparison of the protein sequence to orthologs identified in other *Brassicacea* species shows that, of the four Can-0 specific non-synonymous polymorphisms identified in exon 2, two are likely ancestral variants conserved within related species *A. lyrata*, *Capsella rubella*, *Capsella grandiflora* and *Boechera stricta*, while two are not (Supplementary Figure 2). Therefore, it is possible that the Can-0 variant underlying the difference in recombination could be ancestral, or it is equally possible that it could have been derived more recently, potentially as a consequence of adaptation. While overall the Can-0 *TPR8* allele differs considerably from the other accessions analysed and from related *Brassicacea*, identification of the causal variant is required before it can be definitively determined whether the Can-0 TPR8-associated recombination phenotype is more recently derived than that of the Col-0 allele.

The primary limitation of using T-DNA insertion lines to assess candidate genes is that the insertion alleles do not match those found in accessions (Alonso-Blanco and Koornneef 2000), meaning that the resulting effect on recombination may be entirely different or not apparent at all, depending on the consequences on protein expression and function. Additionally, assessment of multiple candidate genes limits the number of T-DNA lines that can be realistically assessed per gene, given the requirement of crossing to the *420* reporter system to enable measurement of recombination, and the measurement of multiple replicates per line. However, the alternative method of cloning and transformation of candidate alleles is also complex, making T-DNA insertion lines the more efficient system for screening multiple candidates. Ultimately, additional fine-mapping populations could be used to narrow the interval sufficiently that transformation of candidates was less challenging. While candidate genes have been identified for the QTLs, additional follow-up experiments are required to confirm these effects and determine their mechanism of action.

6.4 Characterisation of the HEI10 meiotic E3 ligase as a modifier of meiotic recombination

Analysis of a fourth QTL identified in a Cvi-0/Col-0 segregating accession cross population demonstrated that rQTL1^{420} overlapped the known recombination modifier *HEI10* (Chelysheva *et al.*

2012; Ziolkowski *et al*. 2017). *HEI10* was a strong candidate for the modifier on the basis of shared polymorphisms with the L*er* accession allele which is known to be associated with lower recombination compared to the Col-0 allele (Figures 45 and 46, Ziolkowski *et al*. 2017). Introgression of rQTL1^{420} into the Col-0 background revealed that individually, the Cvi-0 allele of rQTL1^{420} causes an average decrease in recombination of 3.1 cM in the *420* interval compared to Col-0/*420* controls. However, analysis of a cross to the *5.11* FTL interval suggests that rQTL1^{420} does not have a significant effect on recombination in centromeric intervals (Tables 42 and 43, Figure 50).

This effect is consistent with rQTL1^{420} corresponding to the *HEI10* gene, as analysis of *HEI10* overexpressor lines showed that increases in recombination were primarily observed in the euchromatic arms and sub-telomere, and not proximal to centromeres (Ziolkowski *et al*. 2017; Serra *et al*. bioRxiv). This feature of the *HEI10* effect on recombination could explain why rQTL1^{420} was not identified in segregating populations measuring crossovers in centromeric FTL intervals. It also suggests a possible role for HEI10 in mediating sex-specific differences in recombination – in Arabidopsis, male recombination is specifically increased in the sub-telomere (Giraut *et al*. 2011; Drouaud *et al*. 2007), indicating that HEI10 may affect male recombination rates. However, the effect on *420* recombination observed in the rQTL1^{420} introgression line is considerably reduced when compared to earlier fine mapping lines with larger introgression segments on chromosome 1 in the Cvi-0 genotype. This suggests that there may be more than one modifier on chromosome 1 affecting recombination, and that the latest introgression line has separated these loci and fixed one in the Col-0 genotype, thereby reducing the combined effect on recombination.

To test the hypothesis that *HEI10* was the gene underlying rQTL1^{420}, the Cvi-0 allele of *HEI10* (*HEI10*^{Cvi}) was transformed into the Col-0/*420* line, as was previously performed to demonstrate the differential effects of Col-0 and L*er*-0 alleles of *HEI10* (Ziolkowski *et al*. 2017). Measurement of T_1 lines demonstrated a significant difference between *HEI10*^{Col} and *HEI10*^{Cvi} overexpressor lines, consistent with differences in *HEI10* alleles underlying rQTL1^{420} and causing variation in recombination between Cvi-0 and Col-0 (Table 43 and Figure 51). However, due to low sample numbers, additional replicate T_1s need to be analysed to confirm this effect.

Analysis of common polymorphisms shared between accessions showing a *HEI10*-like QTL peak on chromosome 1 in segregating accession-cross populations, including L*er*-0 and Cvi-0, revealed the R264G non-synonymous amino-acid substitution as a strong candidate for the causal variant causing accession specific effects (Ziolkowski *et al*. 2017). Measurement of *420* recombination in T_1 plants containing a transgenic amino-acid swap construct with the 264G variant in the Col-0 allele of *HEI10* demonstrated that this variant had a significant effect when compared to *HEI10*^{Col} T_1 plants (Figure

51). This is consistent with this variant contributing to observed recombination differences between accessions. However, while this variant caused a significant difference from *HEI10*^{Col} T_1s, the difference was not as large as that observed between *HEI10*^{Col} and *HEI10*^{Ler} T_1 plants. While additional data is required to confirm this effect due to the low number of replicates examined, this suggests that additional variants in the L*er*-0 allele may contribute to the observed differences. As the *HEI10*^{Cvi} and *HEI10*^{R264G} T_1 lines show similar recombination rates, it is likely that R264G is the causal variant causing the difference between Col-0 and Cvi-0 alleles.

The R264G variant is an amino acid substitution converting arginine to glycine at amino acid 264 in the C-terminal region of the HEI10 protein, a region which is predicted by homology to be involved in recognition of substrates for HEI10 SUMOylation or ubiquitination functions (Deshaies and Joazeiro 2009). It is possible that this variant alters substrate recognition, thereby altering the targets of HEI10 or the level of SUMOylation/ubiquitination being performed. The additional synonymous polymorphism found in the C-terminal region of the L*er*-0 allele (Ziolkowski *et al*. 2017) could further exacerbate this difference from the Col-0 allele, thereby increasing the disparity in recombination effect. Alternatively, the disparity between *HEI10*^{Ler} and *HEI10*^{Cvi} allelic effects could be due to the additional Cvi-0 specific variants, found in introns and upstream of the gene in the predicted promoter sequence, antagonising the effect of the R264G variant in this background, although this is unlikely given the similarity of *HEI10*^{Cvi} and *HEI10*^{R264G} T_1 recombination estimates (Figure 46). While transformation of promoter swap constructs and analysis of allele expression patterns revealed that differences in expression level were unlikely to underlie the observed differences between *HEI10*^{Col} and *HEI10*^{Ler} effects on recombination (Ziolkowski *et al*. 2017), no corresponding analysis has been performed in Cvi-0. Additional work addressing this could aid understanding of the difference between *HEI10*^{Col}, *HEI10*^{Ler} and *HEI10*^{Cvi} alleles.

It is unclear whether the *HEI10* allele variation between Arabidopsis accessions is due to differential selection for modification of recombination or alternative features, or if it is a result of random variation that was not under strong selection and therefore moved to fixation in some accession populations by chance. Analysis of the frequency and distribution of the R264G variant amongst the accessions sequenced for the 1001 genomes project revealed that the glycine variant present in L*er*-0 and Cvi-0 was found in 95% of the accessions analysed (Alonso-Blanco *et al.* 2016; Ziolkowski *et al.* 2017). This glycine residue is also conserved in *HEI10* orthologs in other Brassicaceae species, suggesting that this variant may be ancestral, and that the arginine variant identified in Col-0 was derived more recently (Ziolkowski *et al.* 2017). This could potentially mean that the fixation of the arginine variant in Col-0 was the result of selection pressures encouraging increased recombination, which may suggest that specific environmental or genetic factors affecting Col-0 necessitated a change

in recombination patterns. While the benefit of altering recombination in Col-0 is unclear, it is possible that a *HEI10* variant associated with higher recombination may have been selected for to balance the level of recombination against the effect of an antagonistic variant in another modifier. However, while this concept appears to be consistent with observations made of two recombination modifiers that vary between Col-0 and L*er*-0 accessions, it is still speculative, since the potential selection pressures are unknown as this variant is found in geographically and genetically diverse accessions (Ziolkowski *et al.* 2017). Analysis of the surrounding sequence for signatures of selection in other variants may help to clarify whether the fixation of the R264 variant in Col-0 was likely to be the result of selection or random chance, thereby helping to determine the evolutionary significance of the variant.

Further characterisation of the effect of the HEI10 protein on recombination revealed that overexpression of the *HEI10*^{Col} allele caused an increase in crossover coincidence, which indicates a decrease in crossover interference (Tables 45 and 46, Figure 52). This was surprising given HEI10's involvement in the ZMM-dependent interfering crossover pathway (Chelysheva *et al*. 2012). Comparison with *recq4a recq4b* anti-crossover helicase mutants, which show an increase in non-interfering crossovers (Seguela-Arnaud *et al*. 2015), and have measured distances between DCO events that are not significantly different from random, thereby suggesting a loss of interference (Serra *et al*. bioRxiv), demonstrates that some level of interference is still acting in *HEI10* overexpression lines, although it is weaker than in wild-type. This suggests that increasing *HEI10* expression, which results in a significant increase in crossovers, could affect the sensitivity of recombination intermediates to the interference signal, or affect the propagation of the signal itself (Serra *et al*. bioRxiv).

As alteration to the C-terminal region of HEI10, in the form of the R264G variant, is known to affect recombination (Figure 51), it is possible that this also affects crossover interference. This C-terminal variation is predicted to affect substrate recognition and potentially SUMOylation or ubiquitination functions which, depending on the target proteins, could affect the influence HEI10 has on crossover interference (Deshaies and Joazeiro 2009; Ziolkowski *et al.* 2017). If HEI10 targets included the SC or chromosome axis proteins, which form the lateral elements of the SC and are believed to play a role in crossover interference (Kleckner *et al*. 2004), this could explain the link between *HEI10* overexpression and changes in interference (Figure 52). This would be consistent with evidence from the Sordaria fungus, where HEI10 integrates signals from the SC and associated recombination complexes via SUMOylation or ubiquitination functions to control crossover development (De Muyt *et al*. 2014). Modification of crossover interference through SUMOylation of chromosome axis components would also be consistent with yeast systems, where evidence from *S.cerevisiae* has

demonstrated that SUMOylation of the axis protein Red1 (homolog of Arabidopsis ASY3) and axis-associated topoisomerase II is required for wild-type levels of interference (Cheng *et al.* 2006; Zhang *et al*. 2014). The SUMOylation of Red1 is dependent on the Zip3 SUMO E3 ligase, which is both structurally and functionally related to the Arabidopsis HEI10 protein, supporting the theory that HEI10 targets may include axis proteins (Cheng *et al.* 2006; Eichinger and Jentsch 2010; Chelysheva *et al.* 2012). Ubiquitin-mediated removal of SUMOylated proteins is also required for wild-type interference in *S.cerevisiae*, suggesting that if HEI10 is shown to have a ubiquitination function in Arabidopsis, but not a SUMOylation function, this may still be relevant to the interference phenotype (Zhang *et al*. 2014).

As Arabidopsis HEI10 is related to both mammalian HEI10, which is involved in ubiquitin-mediated protein removal of recombination factors in meiosis, and yeast Zip3, which has SUMO E3 ligase activity, it is unclear whether it performs SUMOylation or ubiquitination functions during meiosis (Chelysheva *et al.* 2012; Qiao *et al.* 2014; Ziolkowski *et al.* 2017). It is also possible that HEI10 performs both functions in Arabidopsis - many proteins in meiosis are regulated by SUMOylation and ubiquitination, thereby demonstrating the requirement for both functions, yet no other HEI10 or Zip3 homolog has been identified in plants (Bhalla *et al*. 2008; Chelysheva *et al.* 2012; Wright *et al*. 2015; Ziolkowski *et al*. 2017). In mammals the SUMOylation and ubiquitination of recombination factors and the chromosome axis is performed by two separate proteins - the SUMO ligase RNF212 which designates intermediates for crossover formation by stabilising recombination proteins such as MSH4 and MSH5 through SUMOylation, and the ubiquitin ligase HEI10 which is involved in recombination protein and RNF212 turnover through ubiquitin-mediated removal of SUMO-conjugates to promote crossover progression (Qiao *et al.* 2014; Rao *et al*. 2017). The balance between these two processes affects the differential stabilisation of recombination factors associated with the axis, and thereby influences crossover designation at recombination sites (Qiao *et al.* 2014; Rao *et al*. 2017). Although a link to interference hasn't been shown in mammals as it has in yeast, the importance of SUMOylation in modulation of axis formation and stabilisation of axis-associated recombination factors suggests that it is possible that RNF212, and by extension HEI10, affect crossover interference in mammals (Zhang *et al.* 2014; Qiao *et al.* 2014; Rao *et al*. 2017). This raises the possibility that SUMOylation/ubiquitination of the axis and axis-associated recombination proteins is a conserved mechanism of crossover regulation throughout the eukaryotic domain, which increases the likelihood that HEI10 in Arabidopsis acts to target similar proteins in a SUMO-ubiquitin relay. It remains to be seen whether HEI10 is acting alone in this context or, as in mammals, if an antagonistic partner protein acts to maintain a balance in recombination.

Further QTL mapping populations or mutant screens may be able to identify a second protein, if it does exist, as a modifier of recombination frequency, as was performed for HEI10. Interestingly, the putative candidate underlying rQTL2^{420}, *ARI7*, is proposed to have a role in protein ubiquitination (Mladek *et al.* 2003). As variation within the *ARI7* gene appears to affect recombination, it is possible that this protein could act within the proposed SUMO-ubiquitin relay to modulate crossover formation, perhaps in conjunction with HEI10. Additional experiments to determine the targets of both proteins, and whether their effects interact to influence recombination, could provide more information about the roles of SUMOylation and ubiquitination in the regulation of recombination in plants.

The potential alteration of Arabidopsis HEI10 SUMOylation/ubiquitination of target proteins due to C-terminal polymorphisms could alter transmission of the interference signal if targets included the chromosome axis or SC proteins. This could occur through a direct effect on the formation or properties of the axis or SC itself, or it could occur due to an effect on the stabilisation of axis-associated recombination proteins– for example, reduction in SUMOylation activity could prevent intermediates being stabilised on the axis for crossover formation, as seen in mammals (Qiao *et al.* 2014; Rao *et al*. 2017). An effect like this could explain why different *HEI10* alleles with polymorphisms in the C-terminal region would have different effects on recombination. Analysis of interference using three-colour FTL systems in different *HEI10* allele overexpression systems could clarify whether these variants also affect crossover interference, and therefore potentially provide additional information about the nature of the interference signal and the role of SUMOylation and ubiquitination. As *HEI10* overexpression has a significant effect on recombination and interference regardless of C-terminal polymorphisms, it is possible that these variants do not cause an alteration in substrate choice, but rather in efficiency of substrate recognition. This could subsequently alter the amount of SUMOylation/ubiquitination performed by HEI10, which could affect the axis protein composition by modifying the relative stability of different protein components, thereby potentially affecting interference transmission by the axis. Alternatively, HEI10 could be targeting axis-associated recombination factors and affecting their relative stability and turnover, which could potentially influence the sensitivity of the recombination intermediate to the interference signal. Alteration of the efficiency of substrate recognition by HEI10 of either of these target possibilities could integrate the effects of overexpression and C-terminal alteration on recombination. However, these theories are speculative until additional experiments are performed to identify target proteins, as the targets of Arabidopsis HEI10 are currently unknown (Chelysheva *et al.* 2012; Ziolkowski *et al.* 2017).

It is interesting to speculate on why a system would evolve where HEI10 affects crossover interference, as this is a function that acts to prevent excess recombination and promote a more even

distribution of recombination events, which would seem advantageous for maintaining genome stability. It is possible that the effect of HEI10 on interference developed as a necessary requirement to allow the increased number of crossovers in a *HEI10* overexpressor line to form on individual chromosomes, rather than being an effect of selection acting specifically to affect crossover distribution. Alternatively, the decrease in interference could be due to random genetic variation linked to *HEI10* that, while not positively selected for, also wasn't strongly selected against, resulting in fixation by chance as a consequence of genetic drift. However, it is also possible that the HEI10-related reduction in interference could be specifically selected for, to enhance clustering of recombination events in specific regions, particularly in the gene-rich chromosome arms and sub-telomeres, thereby decreasing linkage drag of variants and improving the rate of adaptation (Barton and Charlesworth 1998). Additional work would be required to separate the effects of increased recombination and decreased interference to clarify some of these possibilities, but the interlinked nature of the two properties would make this highly difficult to perform.

6.5 Evolutionary considerations regarding variation in meiotic recombination

The variation in recombination frequency observed between accessions of *Arabidopsis thaliana*, and the apparent evolution of several modifiers that appear to cause significant increases in recombination in specific populations, is of particular interest when considered within the evolutionary context of mating systems. *Arabidopsis thaliana* is a predominantly self-fertilising species, with average outcrossing rates estimated at around 2%, although considerable variation is observed between natural populations (Bomblies *et al*. 2010). Recombination is often considered to have minimal use in self-fertilising species due to high levels of homozygosity making many recombination events largely ineffective in breaking up genetic associations (Marais *et al*. 2004). Therefore, it may seem unclear why *A.thaliana* accessions would have evolved such high levels of recombination and show such considerable variation in recombination rate (Lopez *et al.* 2012; Salome *et al*. 2012; Ziolkowski *et al*. 2015).

To understand the evolution of recombination rates in *A.thaliana* accessions, consideration needs to be made of why high levels of recombination might evolve in a predominantly self-fertilising species. While self-fertilisation offers advantages in both reproductive assurance, as there is no need to find a mating partner, and reduced reliance on pollinators thereby facilitating colonisation, it also has several disadvantages concerning the evolution of the species and is therefore sometimes considered an evolutionary dead-end (Takebayashi and Morrell 2001). Self-fertilisation increases the level of homozygosity present in the genome, which can reduce the amount of genetic variation available in the population to create new allele combinations and can also increase the probability of mildly

deleterious variants moving to fixation before they can be removed from the genome (Marais *et al.* 2004; Morrell *et al.* 2004). Many theories maintain that this can lead to evolutionary inflexibility as self-fertilisation and the subsequent high levels of homozygosity preclude adaptation to changing environments by reducing the effective population size (Gibbs *et al*. 1975; Takebayashi and Morrell 2001). Self-fertilising species such as *A.thaliana* are also predicted to have reduced genetic diversity due to local population extinction and recolonization removing alleles from the wider population, thereby further exacerbating the reduction in effective population size (Bomblies *et al.* 2010).

Despite these perceived disadvantages, self-fertilisation is a common mating system found in several important crop species, including wheat, barley and tomatoes (Morrell *et al.* 2004). Interestingly, it has been proposed that high levels of recombination may be selected for in self-fertilising lines to compensate for the reduced mixing of genomes resulting from minimal outcrossing (Stebbins 1950). Increasing the recombination frequency could maximise the benefits of rare outcrossing events by facilitating the rapid production of new allele combinations (Gibbs *et al.* 1975). This could be highly effective, allowing the accumulation of numerous recombination events in the genome over time, even with a low outcrossing rate of around 2% (Morrell *et al*. 2004). Additionally, low levels of recombination are known to reduce the efficacy of selection, as selection frequently acts in opposite directions on linked variants, both preventing the fixation of beneficial variants due to their association with deleterious variants and simultaneously limiting effective removal of deleterious variants from the genome due to their association with beneficial variants (Morrell *et al*. 2004; Marais *et al*. 2004). Increasing recombination could help to increase selection efficacy by separating linked variants, allowing selection to act on them differentially and effectively, which is important for evolution and adaptation in self-fertilising species as multi-locus associations are more frequent due to a tendency to persist in inbred populations after being generated by mutation or genetic drift (Marais and Charlesworth 2003; Morrell *et al.* 2004; Roze and Barton 2006; Webster and Hurst 2012). Low levels of recombination can also lead to reduced levels of genetic variation in the population by allowing background selection for elimination of deleterious mutations and selective sweeps of advantageous mutations to affect a much larger block of linked variants than they would in genomes with high recombination, resulting in a reduction in available allele combinations for a larger number of genes (Nordborg *et al*. 1996; Dvorak *et al*. 1997; Kraft *et al*. 1998; Mitchell-Olds and Schmitt 2006). Regions of the genome that do not undergo recombination are also believed to accumulate deleterious mutations (Takebayashi and Morrell 2001; Webster and Hurst 2012). This issue could potentially be more severe in self-fertilising species, as inbreeding depression makes recessive deleterious alleles more likely to move to fixation in the genome (Takebayashi and Morrell 2001).

Therefore, increased recombination could be encouraged in self-fertilisers to promote removal of deleterious variants by efficient selection before they can become fixed (Morrell *et al.* 2004).

Analysis of selection and genetic variation suggests that there are many possible reasons for self-fertilising species like *A.thaliana* to undergo significant amounts of recombination, which is consistent with several sources of evidence that recombination is higher in self-fertilisers than in close outcrossing relatives (Gibbs *et al*. 1975; Charlesworth *et al*. 1976; Hansson *et al.* 2006). Numerous cytological observations made in different species, including plants from the Elymus, Sitanion, Agropyron, Gilia, Lolium and Limnanthes genera, have shown that chiasma frequencies tend to be higher in self-fertilising plant species when compared to outbreeding relatives (Stebbins *et al.* 1946; Grant 1952; Rees and Thompson 1956; Jones and Rees 1966; Arroyo 1973). While chiasma counts have not yet been performed in *Arabidopsis lyrata,* an outcrossing relative of *A.thaliana*, to allow comparison of genome wide recombination between the two species, comparison of genetic map distances over a defined region of the genome indicated that recombination is higher in the self-fertilising *A.thaliana*, which is consistent with the evidence that chiasma number per chromosome arm is generally higher in inbreeding species (Morrell *et al*. 2004; Hansson *et al*. 2006).

This suggestion of *A.thaliana* having increased recombination compared to outcrossing relatives is consistent with the preliminary analysis of the four modifiers of meiotic recombination frequency identified in this project. Examination of the potential causal variants underlying these modifiers suggests that for three out of the four significant large-effect modifiers of recombination found in these QTL mapping populations (rQTL1^{420}, rQTL2^{420} and CanQTL4), the variants believed to be associated with an increase in recombination appear to be more recently derived in *A.thaliana* accession populations, whereas variants potentially associated with lower recombination appear to be more ancestral (Supplementary Figures 1 and 2; Ziolkowski *et al.* 2017). This may suggest that variants associated with increases in recombination are being selected for in *A.thaliana*, which may contribute to a possible overall increase in recombination in the self-fertilising species compared to outcrossing *A.lyrata*. Simulated models indicate that increases in self-fertilisation should often lead to increased selection for recombination, particularly in fluctuating environments or in the presence of genetic hitchhiking where variants change frequency in the population due to linkage to other variants under selection (Charlesworth *et al.* 1976). Under alternative conditions there may be selection for decreased recombination, although, as self-fertilising species frequently have higher chiasma frequencies, it seems likely that selection for increasing recombination must frequently outweigh selection against crossovers in the overall effect of selection across the whole genome (Charlesworth *et al.* 1976). However, it must also be considered that while it is possible that the modifiers increasing recombination have been selected for in the evolution of self-fertilising *A.thaliana* populations, it is

also possible that these phenotypes are just the result of genetic drift producing LD between variants associated with increasing recombination, due to minimal selection against increases in recombination.

Although a general increase in recombination in *A.thaliana* compared to close outcrossing relatives as a response to increased levels of self-fertilisation seems likely, it raises the question of why there is so much variation in recombination between *A.thaliana* accessions (Lopez *et al.* 2012; Salome *et al*. 2012; Ziolkowski *et al*. 2015). Given the possible reasons behind increased recombination in self-fertilising species, it seems probable that levels of recombination in *A.thaliana* accessions could be linked to the level of outcrossing occurring in each population, as this will affect heterozygosity and genetic variation (Gibbs *et al.* 1975; Bomblies *et al*. 2010). Several self-fertilising species show variation in genetic diversity and levels of outcrossing between populations, and there is evidence that wild populations of *A.thaliana* are also strongly differentiated and vary considerably in their effective outcrossing rates (Bomblies *et al.* 2010). This could cause variation in effective population sizes and levels of heterozygosity, and therefore affect the efficacy of selection acting on recombination in each accession population (Marais *et al.* 2004; Haudry *et al*. 2008). Simulation models of selection and recombination in fluctuating environments show that populations differ for recombination frequency depending on the level of self-fertilisation, supporting the idea that variation in recombination between *A.thaliana* accessions could be partially due to variation in the levels of outcrossing (Charlesworth *et al*. 1976).

The presence of different selection pressures in different environments could also result in variation in recombination between accessions – selection for traits in domesticated animals has led to indirect alteration of recombination frequency compared to wild progenitors, suggesting that environmental selection for traits unrelated to recombination that differ between accessions may have indirectly resulted in variation in recombination (Otto and Barton 2001; Smukowski and Noor 2011). Variation between accessions due to differential selection is a definite possibility, as recombination patterns are known to evolve rapidly, differing significantly even between closely related species such as humans and chimps (Hansson *et al.* 2006; Winckler *et al.* 2005). Alternatively, it is possible that this variation in recombination between accessions is not due to differential selection, but rather a relaxation of selective constraints modulating recombination levels. Selection is predicted to be weaker in self-fertilising species, due to inbreeding and genetic linkage limiting efficient removal of deleterious mutations and fixation of beneficial mutations, which could result in a reduction of the effect of selection on recombination modifiers (Morrell *et al*. 2004; Haudry *et al*. 2008; Webster and Hurst 2012). This could mean that variants in modifiers affecting recombination rates are instead subject to genetic drift, which is likely to vary between natural populations due to differences in outcrossing

levels and effective population sizes (Marais *et al.* 2004; Bomblies *et al*. 2010). This variation in genetic drift could affect the frequency of variants affecting recombination in each population, and could therefore potentially underlie the observed variation in recombination rate between *A.thaliana* accessions.

It is interesting to consider what the potential selection pressures affecting the evolution of recombination rate in *A.thaliana* could be, if the variation in recombination rate has evolved as a direct response to the distinct environmental conditions of each accession rather than as the consequence of genetic drift or some other process. *A.thaliana* accessions are found across a broad geographical distribution and range of climates, meaning that the abiotic and biotic selective pressures acting on them also vary considerably, resulting in adaptive phenotypic variation (Mitchell-Olds and Schmitt 2006; Brachi *et al*. 2010; Manzano-Piedras *et al*. 2014). Accessions will show phenotypic differences due to adaptation to diverse environmental conditions, for example considerable variation is found in flowering due to the alteration of genes involved in the regulation of flowering time to allow each population to flower during optimal growing conditions despite variation in photoperiod and temperature (Mitchell-Olds and Schmitt 2006; Brachi *et al.* 2010; El-Soda *et al*. 2014).

A variety of selection pressures could be responsible for the observed variation in recombination, providing that the environmental conditions create an advantage to recombination (Hadany *et al*. 2008). These pressures may include variation in the levels of outcrossing between populations which may alter the level of selection for recombination modifiers as discussed above (Charlesworth *et al*. 1976; Bomblies *et al*. 2010). Fluctuations in environmental conditions can also create an advantage to increased recombination, as the breaking of linkage between variants can improve rates of adaptation to new conditions (Charlesworth *et al.* 1976; Hadany *et al*. 2008; Webster and Hurst 2012). Simulations suggest that this could lead to increased selection for modifiers of recombination in unstable environments (Charlesworth *et al*. 1976).

In stable environments, plants are often well-adapted and exhibit low recombination rates, presumably to maintain beneficial allele combinations that are adapted to the current environment (Barton and Charlesworth 1998; Webster and Hurst 2012; Lloyd *et al*. 2018). This is also consistent with evidence that recombination is low under optimal conditions, but that it increases under certain stress conditions which may create a requirement for efficient adaptation to altered conditions (Bomblies *et al*. 2015; Lloyd *et al.* 2018). For example, strains of Sordaria from harsh sites with variable conditions show higher recombination rates than strains from lush stable sites when both sets are grown under the same laboratory conditions (Saleem *et al.* 2001; Bomblies *et al.* 2015). Although it is unknown whether this phenotype is also observed in the populations under their natural conditions,

it suggests that there may be selection for increased recombination under stressful conditions, or at least that the environmental sensitivity of the recombination phenotype is altered resulting in a change under lab conditions. A similar situation could be occurring in the Can-0 and Cvi-0 accessions of Arabidopsis which are known to experience high temperatures and arid climates in nature and show high recombination rates under lab conditions (Ziolkowski *et al.* 2015).

Environmental stress factors are known to vary between natural populations and affect a variety of traits in plants. For example, flowering time, a trait that shows considerable variation in nature, is known to change as environmental stress increases due to climate warming or increased aridity (Manzano-Piedras *et al.* 2014). This alteration of flowering in response to stress factors is likely to be linked to observations that stress response genes are differentially expressed between populations (Wada and Takeno 2010; Baduel *et al*. 2016). This differential expression is also likely to affect the recombination phenotype, as experiments performed in both animal and plant species have demonstrated links between the recombination process and regulators of stress response (De Storme and Geelen 2013; Stevison *et al.* 2017). There is also evidence of differential selection acting on stress response regulators in different populations, which could have pleiotropic effects on gene expression and ultimately affect traits like flowering time or recombination (Baduel *et al*. 2016). It is known that recombination can increase as a by-product of selection on unrelated traits, including selection for geotaxis and DDT resistance in flies, suggesting that organisms under strong selection due to harsh, stressful or fluctuating environments could show higher recombination rates as a general rule (Bourguet *et al.* 2003; Webster and Hurst 2012; Bomblies *et al*. 2015).

Alternatively, recombination rates could be altered by direct selection in response to specific stress factors. For example, stress in the form of pathogen infection is known to associate with variation in recombination rate within individuals (Stapley *et al.* 2017). This is believed to be linked to the Red Queen hypothesis of host-pathogen co-evolution, as higher levels of recombination could be selected for as a response to infection to produce more genetically diverse offspring which could prevent rapidly evolving parasites establishing themselves in a permissive host genotype (Salathe *et al.* 2009; Stapley *et al*. 2017). Increased recombination rates in infected organisms have been observed in many species, including Arabidopsis and Drosophila (Kovalchuk *et al.* 2003; Andronic 2012; Singh *et al.* 2015; Stapley *et al.* 2017; Dapper and Payseur 2017). Variation in rates of infection between populations and genotypes, combined with varying geographical distributions of pathogens, could therefore contribute to the evolution of variation in recombination rates observed in natural *A.thaliana* populations.

Another selection pressure that is likely to affect recombination rates in natural populations of *Arabidopsis thaliana* is temperature. Changes in ambient temperature are associated with altered recombination rates in many species, with extreme temperatures causing structural issues with meiotic proteins, often resulting in failure to complete meiosis (Morgan *et al.* 2017; Lloyd *et al.* 2018). Asynapsis and chromosome axis defects are frequently observed at high temperatures, which can often result in a concomitant reduction in recombination due to the dependence of recombination on chromosomal axis proteins and the SC in several species, including *A.thaliana* (Dowrick 1956; Loidl 1989; Morgan *et al.* 2017; Sanchez-Moran *et al*. 2007; Ferdous *et al*. 2012). Recombination in *A.thaliana* has been shown to be lowest at optimal growth temperatures, with increases in recombination observed at both high and low temperatures (Lloyd *et al*. 2018). Similar effects have been observed in other eukaryotic species, although the optimal temperature range does shift between populations (Hoffman and Parsons 1991; Francis *et al*. 2007; Wright *et al.* 2015; Morgan *et al.* 2017).

As it is unclear what the advantage of increased recombination would be at higher temperatures, beyond the aforementioned increased adaptation efficiency under new conditions, it is possible that this is not a directly adaptive response of recombination to temperature. The altered recombination rate may be an indirect consequence of the effect of temperature on meiotic proteins, as high temperatures are known to induce aggregation of SC proteins into polycomplexes, thereby altering the formation of the SC and affecting recombination (Morgan *et al.* 2017). Changes in chromosome axis structure in response to temperature may affect recombination by reducing the strength of crossover interference or slowing the process of crossover designation, resulting in reduced feedback suppression of additional CO events and thereby increasing recombination levels (Morgan *et al*. 2017). Selection for thermotolerance in populations experiencing higher ambient temperatures may alter the properties of SC or axis proteins to reduce their tendency to aggregate, which may then have a knock-on effect on recombination by affecting the development of the SC and associated recombination complexes (Morgan *et al.* 2017). This is supported by observations in *A.arenosa* where chromosome axis proteins and cohesins, which are believed to limit the aggregation of axis proteins, are under strong selection in a population which has adapted to a warmer climate (Wright *et al.* 2015; Morgan *et al.* 2017). The variation in recombination observed between many natural plant populations may therefore be a side-effect of the adaptation of meiosis to a new environment to maintain fertility.

As populations colonise new habitats, changes in temperature may increase selection on axis proteins and other SC components to stabilise the structures required for synapsis, recombination and completion of meiosis (Wright *et al.* 2015; Morgan *et al.* 2017). The Cvi-0 and Can-0 accessions grow

under warmer natural conditions than many other *A.thaliana* accessions, suggesting that it is possible that their observed increased recombination rates could be a by-product of selection for increased thermotolerance. While none of the candidate genes identified for the recombination QTLs found in this project (that vary between the Col-0, Can-0 and Cvi-0 accessions) encode axis proteins or cohesins, some of them do encode regulatory proteins, suggesting that they could be involved in responses to environmental factors. These genes could be involved in responses to temperature or alternative environmental factors, resulting in pleiotropic effects that may ultimately influence recombination.

While *HEI10*, the gene underlying rQTL1^{420}, has a known function in recombination and *ARI7*, the candidate gene underlying rQTL2^{420}, appears to have a potential role in a process known to influence recombination, the situation for the candidate genes underlying rQTL5^{420} and CanQTL4 is less clear (Mladek *et al*. 2003; Chelysheva *et al.* 2012; Ziolkowski *et al*. 2017). The candidate gene underlying rQTL5^{420}, a PP2C protein phosphatase (At5g53140), is predicted to be involved in protein dephosphorylation, which is important for the modulation of a variety of cellular processes (Kerk *et al.* 2002). Although there is a lack of functional information about PP2C proteins in Arabidopsis specifically, plant PP2C proteins are generally known to function in regulation of stress signalling pathways that affect the cell cycle, and PP2C proteins in yeast are known to regulate heat shock signalling (Kerk *et al*. 2002; Schweighofer *et al*. 2004). This suggests that *PP2C* (At5g53140) is likely to be indirectly affecting recombination through a pleiotropic stress signalling cascade rather than specifically acting to modulate recombination levels. However, there is some evidence to suggest that PP2C physically interacts with proteins involved in SUMOylation and ubiquitination, which could possibly link PP2C to the recombination-associated SUMO-ubiquitin relay, although this link is tenuous (Araport11, Krishnakumar *et al*. 2015).

While *PP2C* (At5g53140) could conceivably be affecting recombination directly by dephosphorylating a recombination protein, it seems more likely, given the conserved role of PP2C proteins in stress signalling cascades, that it affects recombination indirectly through the regulation of protein signalling. For example, given the link between plant PP2C proteins and cell cycle signalling, PP2C could influence recombination by regulating the signalling for progression out of the prophase I stage of meiosis, as delayed progression can affect recombination rates (Zhang *et al*. 2014; Joyce and McKim 2010; Joyce and McKim 2011). Alternatively, PP2C could modulate the signalling linking DSBs to the recombination machinery, as this is known to involve the phosphorylation of γH2AX (Chowdhury *et al*. 2005; Baudat *et al.* 2013). Either of these possibilities could explain the apparent link between PP2C and recombination while also providing a mechanism for coordination of recombination with other cellular events. Therefore, variation in the *PP2C* gene could conceivably underlie some of the variation in recombination rates between *A.thaliana* accessions. Variation in *PP2C* alleles between the Col-0

and Cvi-0 accessions could potentially alter the activity of the protein and thereby alter the efficiency of the signalling cascade, having a knock-on effect on the rate of recombination. While it is possible that this variation is under selection to provide a mechanism for the modulation of recombination as a response to stress, possibly to make recombination rates more robust and less subject to environmental variation, it is more likely that this variation was selected for to regulate cellular responses to environmental signals, with the effect on recombination being a by-product of the pleiotropic effects of signalling proteins.

The TPR8 carboxylate clamp-tetratricopeptide repeat protein that is believed to underlie CanQTL4 similarly does not appear to have any obvious link to recombination, including the recombination-associated SUMO-ubiquitin relay. While no function or cellular location has yet been ascribed to the protein, based on protein structure it is believed to have the potential to interact with the heat shock chaperones Hsp90 and Hsp70 as a co-chaperone (Prasad *et al.* 2010). Hsp90 has an important role in signal transduction as it is involved in the regulation of the stability and interactions of several signalling proteins. Co-chaperones act to regulate Hsp90 activity and help to recruit interacting proteins, providing specificity to the interaction (Prasad *et al*. 2010). *TPR8* in Arabidopsis has been shown to be heat responsive, and it is likely that the protein acts as a co-chaperone during heat stress (Prasad *et al*. 2010). It is therefore unlikely to have a direct effect on recombination that was adapted due to selection pressures for increased recombination, but instead is more likely to have pleiotropic effects on the stability and interactions of meiotic proteins as a response to heat stress.

However, there is a small possibility that TPR8 could be involved in a mechanism specifically implemented for the regulation of recombination in response to heat shock, by facilitating the interaction of Hsp90 with specific recombination related proteins. The variation in *A.thaliana* accession alleles of *TPR8* could reflect adaptation of thermotolerance to different temperature ranges – as the Can-0 allele probably evolved at a higher ambient temperature than the Col-0 allele, it is likely that the heat shock response is induced at a higher temperature in Can-0. The variation in *TPR8* alleles could therefore allow temperature range-appropriate modulation of recombination in response to heat stress, by regulating the interaction of Hsp90 with recombination related proteins. This could also include regulating the interaction of chaperones with axis proteins, thereby modulating the aggregation of SC components in response to temperature stress to prevent subsequent chromosome segregation issues and meiotic failure. The modification of a regulatory protein or its expression to adapt to changing temperatures instead of modification of structural components like axis proteins could help to integrate cellular responses to heat stress, and also provide flexibility in response to fluctuating temperatures. It is possible that if the ambient temperature is consistently high, selection will act to adapt structural components like axis proteins to reduce the probability of protein

aggregation and asynapsis, but if the ambient temperature is variable in an unstable environment, selection may act on regulatory proteins, providing a mechanism to maintain meiotic processes despite environmental fluctuations (Morgan *et al*. 2017). However, it is still unclear whether the increase in recombination associated with the Can-0 allele of *TPR8* was directly selected for, or if it developed as a by-product of selection for alteration of the heat-shock response to maintain meiotic stability at higher ambient temperatures.

An alternative possibility is that some of the potential recombination modifiers identified by this project do not act as modifiers underlying variation in recombination between accession populations in their natural habitats. While the genes underlying the identified QTLs clearly have some link to recombination that varies between accessions, it is possible that this difference in recombination is caused by differential responses to growth conditions in the laboratory and therefore may not manifest as recombination variation in populations under their natural conditions. Each *A.thaliana* accession is adapted to a different natural habitat, and laboratory conditions may be closer to the natural habitat of some accessions than others, thereby creating differences in the perception of environmental conditions between plants (Manzano-Piedras *et al.* 2014; Bomblies *et al.* 2015). For example, while accessions are known to have adapted to different ambient temperatures, they were grown here under the same conditions, which may be more optimal for some accessions than others, leading some plants to exhibit phenotypes associated with temperature stress under laboratory growth conditions (Manzano-Piedras *et al*. 2014; Bomblies *et al*. 2015; Morgan *et al.* 2017). Although the accessions used here for QTL mapping may have similar recombination rates when growing at their respective optimum temperatures in their natural habitats, the responses of different genotypes to stress factors caused by controlled laboratory conditions may result in considerable variation in recombination rates. While this doesn't preclude the possibility that the variation in recombination between accessions observed here is the result of differential adaptation of recombination patterns, it could mean that some of the modifiers identified by QTL mapping hold little relevance to variation in recombination in natural populations, but rather reflect the differential expression of the gene or activity of the protein between accessions in response to the laboratory environment. This theory is particularly relevant considering that the candidate genes underlying two of the identified QTLs appear to be regulatory genes involved in stress response pathways that may have pleiotropic effects.

While no obvious signs of plant stress were observed in the QTL mapping populations, such as growth reduction, chlorosis or wilting and plants appeared healthy, it is still possible that responses to stress could have altered the recombination phenotype – a possibility which must be taken into consideration when discussing the evolution of recombination modifiers in natural accessions (Pasternak *et al*. 2005). It has been demonstrated that QTL mapping performed in plants grown under

controlled laboratory conditions reveals different QTLs underlying traits such as flowering time to experiments performed under natural field conditions (Koornneef *et al*. 2004; Brachi *et al.* 2010). Experiments involving the growth of *A.thaliana* RILs under several different conditions revealed that QTLs detected in natural environments are not always detected under controlled laboratory conditions (Weinig *et al*. 2002; Brachi *et al.* 2010). This is probably due to the interaction of the genome with the increased range of environmental signals that plants are exposed to in their natural environment in comparison to the laboratory conditions, as natural conditions are less predictable and factors such as temperature and light intensity are likely to fluctuate from day to day (Brachi *et al.* 2010). The combination of multiple environmental factors experienced in the wild produces unique phenotypic responses that cannot be precisely reproduced under controlled conditions, where several environmental cues are usually absent (Suzuki *et al*. 2014; El-Soda *et al*. 2014). These circumstances demonstrate the importance of performing experiments under ecologically realistic conditions wherever possible when studying adaptive variation (Brachi *et al.* 2010; El-Soda *et al.* 2014).

Unfortunately, *A.thaliana* accessions adapted to different environments were crossed to create the QTL mapping populations used here, which prevented the use of growth conditions matching the natural environments of the plants. While growth conditions were selected that should not cause any of the accessions used any significant stress, adaptive differences between the genotypes mean that QTL mapping in these populations is always going to be influenced by differential Gene x Environment (GxE) interactions, which may lead to the identification of recombination modifiers that, while they have a considerable effect on the phenotype under the laboratory growth conditions, do not have a significant effect on recombination under natural conditions (Koornneef *et al.* 2004; El-Soda *et al*. 2014; Bomblies *et al*. 2015). This could be the case for some of the modifiers identified by this project – for example, *TPR8* may not be responsible for differences in recombination between natural populations of Col-0 and Can-0 accessions, but variation between *TPR8* alleles may affect the response of recombination to temperature under laboratory conditions, resulting in its detection as a QTL underlying recombination variation.

When considering the possibility that some of the factors modifying recombination in these QTL mapping populations may be responses to stress rather than adaptations to natural habitats, it becomes apparent that careful consideration must be made of potential candidate genes. If the genes underlying recombination variation under controlled laboratory conditions are factors involved in stress responses, or genes with pleiotropic effects on processes that may include recombination, they may have been prematurely removed from the candidate list due to their function not having an obvious link to recombination. Although some pleiotropic regulators were included in the candidate lists, if the candidates identified as the most likely genes underlying each QTL are not confirmed as

recombination modifiers by future experiments, other regulatory genes that were removed from the lists may need to be considered as potential candidates in future as they may have indirect effects on recombination that could not be predicted from their perceived functions.

6.6 Future work

The aim of this project was to identify significant modifiers of meiotic recombination frequency in populations generated from crosses between natural Arabidopsis accessions that differed in recombination frequency. Multiple significant modifiers in different genetic backgrounds were identified, and fine-mapping identified candidate genes for each QTL. Further characterisation of one candidate, *HEI10*, confirmed that this gene had accession specific effects on recombination, that a variant in the C-terminal region affected recombination, and that the gene affected recombination by reducing crossover interference. However, many questions about natural variation in recombination remain, and additional questions have been raised by this project – primarily whether the candidate genes identified for $rQTL2^{420}$, $rQTL5^{420}$ and CanQTL4 will be confirmed to effect recombination and what their possible recombination-related functions could be.

Summation of measurements made in F_1 accession/FTL cross lines demonstrated that the Cvi-0 and Can-0 accessions had significantly increased recombination compared to Col-0 homozygous lines. However, at least for Cvi-0, this is in disagreement with chiasma counts made in PMCs of homozygous accession lines (Lopez *et al*. 2012). Measurement of chiasma counts in PMCs of F_1 lines in addition to homozygous Can-0 and Cvi-0 lines would clarify whether this disparity was due to the interaction of modifiers with Col-0 factors in the F_1, or due to a technical difference in methods. Cytogenetic analysis of QTL-NIL PMCs could also be performed, as it was for Col/L*er* recombinant lines showing transgressive recombination phenotypes (Ziolkowski *et al*. 2017). Staining of MLH1 foci could measure the total number of class I crossovers per cell in these lines, allowing estimation of the genome-wide effect of the QTL on recombination. Additional analysis of crossovers in QTL-NILs containing *zmm* knockout mutations could also be performed to determine whether these QTL affect class I or class II crossover pathways. However, to perform these analyses on CanQTL4, additional introgression of the Can-0 allele of CanQTL4 into the Col-0 background is required to generate a QTL-NIL and confirm the effect of this individual QTL on recombination.

Cvi-0 and Can-0 accessions showed similar recombination patterns when estimated by F_1 FTL measurements, however QTL mapping revealed that these patterns were the result of different modifiers. This suggests that each accession may have evolved a similar phenotype through a different mechanism, potentially suggesting a link through convergent adaptation to shared environmental factors. It is possible that the overall increases in recombination observed in these accessions are the

result of selection acting to maintain fertility despite changing conditions (Bomblies *et al.* 2015; Kong *et al*. 2004). Evidence from humans shows that recombination frequency is positively correlated with reproductive success in women, which is consistent with indications that low CO numbers can cause mis-segregation of chromosomes in the first meiotic division and subsequent aneuploidy, which is a leading cause of miscarriage (Kong *et al.* 2004; Chowdhury *et al.* 2009; Fledel-Alon *et al*. 2011; Baudat *et al*. 2013). Increased COs are proposed to improve links between homologous chromosomes, thereby promoting stable bivalent formation and successful chromosome orientation on the metaphase plate, allowing accurate chromosome segregation (Kong *et al.* 2004; Coop *et al*. 2008; Baudat *et al*. 2013). This is believed to be of particular importance in older women as other associations between chromosomes, such as cohesin complexes, are believed to degrade with age, resulting in segregation defects and reduced fertility (Kong *et al.* 2004; Coop *et al*. 2008).

Modifiers of recombination in human populations are thought to be affected by selective forces acting to promote increased recombination and thereby increase survival of oocytes in women by reducing chromosome mis-segregation (Kong *et al*. 2004; Coop *et al*. 2008; Fledel-Alon *et al.* 2011). While increases in segregation defects are not known to occur with age in Arabidopsis, environmental factors such as temperature are known to influence homologous chromosome connections and segregation and may therefore produce similar selection pressures for increased recombination to maintain fertility (Wright *et al*. 2015; Morgan *et al*. 2017; Lloyd *et al.* 2018). The Cvi-0 and Can-0 accessions may therefore be subject to factors that reduce fertility, either through segregation defects or alternative processes, leading to selection for increased recombination rates in these accessions to compensate for this and increase the likelihood of gamete survival.

Preliminary analysis of the potential causal variants underlying the identified QTLs suggests that the Cvi-0 variants underlying rQTL2^{420} and rQTL5^{420} are likely to be more recently derived than the Col-0 variants, and that the same may potentially be true for the Can-0 variant underlying CanQTL4, although additional work is required to confirm this (Supplementary Figures 1 and 2). If the Cvi-0 and Can-0 variants underlying the differences in recombination pattern are indeed more recently derived and are not conserved with the ancestral variants found in other *Brassicacea* species (Supplementary Figures 1 and 2), it is possible that they may have become fixed in these island populations as a consequence of selection for changes in recombination. Confirmation of the candidate genes and causal variants, and an examination of the sequence for signatures of selection, could clarify whether this is the case and aid understanding of the evolution of variation in recombination. Comparison of recombination in accessions from different habitats could also provide information to assess whether the alteration of recombination pattern is a commonly adapted strategy, and whether any specific environmental features correlate with recombination rate. This analysis could suggest common

selection pressures involved in natural variation in recombination, potentially providing evidence to link evolution of higher recombination rates to selection acting on fertility or alternative traits.

It is also possible that the observed increases in recombination in Cvi-0 and Can-0 accessions are not the result of selection acting on recombination specifically, but rather an indirect effect of selection acting on an alternative trait. The Can-0 and Cvi-0 accessions come from the Canary Islands and the Cape Verde islands, off the west coast of Africa, respectively. Both habitats have higher average ambient temperatures than those common for the habitats of other accessions, including that of North-American Col-0, although temperatures do fluctuate considerably throughout the year. There is considerable evidence that Cvi-0 in particular has adapted its physiology to a hot dry environment, resulting in clear phenotypic differences when compared to many other accessions (Bouchabke *et al*. 2008; Monda *et al*. 2011; Huang *et al*. 2014). Studies of recombination in *A.thaliana* using FTL measurements over temperature gradients have shown that increases in growth temperature correlate with increases in recombination, in both subtelomeric and pericentromeric regions of the genome (Francis *et al.* 2007). This increase in recombination is consistent with observations made in other species, such as *Vicia faba* and *D.melanogaster*, and is likely to be a consequence of the evolution of thermotolerance having an indirect effect on recombination, either through adaptation of chromosome axis proteins or cohesins, or of regulatory proteins involved in heat shock response (Berkemeier and Linnert 1987; Stern 1926; Francis *et al*. 2007; Morgan *et al.* 2017). Therefore, it is possible that the modifiers causing the general increase in recombination in Cvi-0/FTL and Can-0/FTL crosses could be the result of adaptation to higher ambient temperatures and the effect they can have on molecular mechanisms involved in recombination and meiosis. Consistent with this, some of the variants in Cvi-0 and Can-0 that are believed to underlie the QTLs found in this project, and are associated with higher recombination, appear to be more recently derived among accessions (Supplementary Figure 1), suggesting that they could potentially be the result of adaptation to the Cape Verde or Canary Island environments.

As temperatures on Cape Verde and the Canary islands vary considerably throughout the year, it is possible that any adaptation to temperature may involve regulatory proteins that can be modulated to buffer responses to changing temperatures, rather than adaptation of structural meiotic proteins like axis proteins as is suggested for *A.arenosa* populations adapted to warmer climates (Wright *et al*. 2015; Morgan *et al.* 2017). It could be interesting to analyse recombination in Col-0/FTL, Cvi-0/FTL and Can-0/FTL lines over a range of temperatures to determine whether the Cvi-0 and Can-0 recombination rates are more robust to temperature change, which could support the theory of these higher rates being the result of a regulatory adaptation to the local environment. Tomato cultivars adapted to different temperatures respond differently to experimental growth temperatures, with

heat resistant lines showing more moderate changes in recombination at higher temperatures than lines adapted to colder conditions (Zhuchenko *et al*. 1986; Rybnikov *et al*. 2017). While increases in temperature are likely to cause changes in recombination in most tomato lines, the size of the change is modulated by the genotype and a similar situation might be found in *A.thaliana* accessions (Rybnikov *et al*. 2017). An analysis of *A.thaliana* recombination over a range of temperatures could also help to clarify whether the high recombination rates compared to Col-0 are consistent in Cvi-0 and Can-0 lines, which would suggest adaptation, or if they are the result of temperature stress caused by laboratory conditions.

Another point of interest is that centromeric recombination is significantly higher in the Cvi-0/FTL F_1 lines than in comparative Col-0/FTL homozygous lines. This could be related to reduced heterochromatin and DNA methylation levels being identified in Cvi-0 (Tessadori *et al*. 2009; Pignatta *et al*. 2014) - factors known to suppress recombination around the centromere (Yelina *et al*. 2012; Copenhaver *et al*. 1999; Choulet *et al*. 2014; Yelina *et al*. 2015). Given this information, and the variability in recombination rates observed in the Cvi-0/*CEN3* and Cvi-0/*5.11* F_2 populations measuring recombination around the centromeres, analysis of individual methylation profiles in a segregating F_2 population is an attractive option to determine whether these levels correlate with recombination frequency. This could provide information about the factors underlying differences observed in recombination patterns between Cvi-0 and Col-0 accessions, and it could also help to determine whether stochastic variation in methylation levels correlates with the background variation in recombination observed in these populations that cannot be attributed to significant *trans*-acting modifiers.

Further work also needs to be performed to corroborate some of the findings made by this project. While fine-mapping populations did identify candidate genes for $rQTL2^{420}$, $rQTL5^{420}$ and CanQTL4, these need to be confirmed as the genes underlying the observed QTL associations through establishment of accession-specific allelic effects on recombination. This would be performed through transformation of the alleles into the Col-0/*420* background, as was performed for confirmation of the HEI10 modifier (Ziolkowki *et al*. 2017). As $rQTL2^{420}$and $rQTL5^{420}$ are semi-dominant modifiers, the effects of transforming the Cvi-0 allele could be compared to those of transformation of the Col-0 allele into Col-0/*420* to determine the difference in allele effects. However, the recessive effect of the CanQTL4 modifier means that this strategy is unlikely to be informative – transformation of CanQTL4 accession alleles must be performed into knockout lines to observe an effect.

If these candidates are confirmed as natural recombination modifiers, additional work can be performed in lines containing mutant alleles, and in QTL-NILs, to determine their effect on

recombination. This could involve looking at meiotic progression, genome wide recombination frequencies and recombination protein localisation during meiosis. If a clear function in recombination is determined, it could also be interesting to look at expression differences between the natural candidate alleles, or differences in protein structure or cellular localisation of the candidate protein that could affect the influence of the gene on recombination.

Another interesting direction for future research, if these candidates were confirmed as recombination modifiers, could involve investigating whether these genes influence crossover interference. This could be performed using three-colour FTL lines in conjunction with QTL-NILs, homozygous T-DNA insertion lines, or transformants containing a transgenic copy of the Cvi-0 or Can-0 allele of the relevant gene. While this analysis has not yet been performed for the three novel modifiers identified here (rQTL2^{420}, rQTL5^{420} and CanQTL4), it has been used in the characterisation of the effect of *HEI10* overexpression on recombination (rQTL1^{420}, Figure 52). This revealed that increasing *HEI10* expression results in both an increase in Class I interference-sensitive crossovers and a decrease in interference (Figure 52; Ziolkowski et al. 2017; Serra *et al*. bioRxiv), which is surprising because a reduction in interference is also a common feature of *zmm* mutants in Arabidopsis, which show reduced Class I crossovers and a random distribution of residual crossovers (Higgins *et al.* 2004; Mercier *et al.* 2015). Interestingly, experiments performed in mutants of anti-CO proteins that are known to limit the formation of Class II crossovers also demonstrate reduced interference. Measurements made in *figl1, flip, recq4ab, fancm* and *rmi1* mutants that have lost anti-CO activity show an increase in non-interfering crossovers and a concomitant decrease in interference (Crismani *et al.* 2012; Girard *et al.* 2015; Seguela-Arnaud *et al*. 2015; Seguela-Arnaud *et al.* 2016; Fernandes *et al*. 2017).

Although it is currently unclear whether the apparent effects of *ARI7, PP2C* and *TPR8* on recombination are due to effects on the Class I or Class II crossover pathway, or both, the fact that modification of interference appears to be a common feature in lines with altered recombination frequencies suggests that it is feasible that they could affect interference. It could also be informative to perform a QTL scan in a segregating accession-cross population for genes affecting interference, as was performed in mammals (Sandor *et al.* 2012). However, this would be problematic in Arabidopsis due either to the difficulty in extracting sufficient pollen from each individual plant to perform three-colour flow cytometry or, alternatively, the time it would take to measure hundreds of individuals through tetrad scoring or alternative techniques (Yelina *et al.* 2013).

If the candidate genes are not confirmed by additional experiments, further work would need to be performed to identify new candidate genes. Measurement of recombination in QTL-NILs for

rQTL2^{420}and rQTL5^{420} clearly demonstrates the presence of large-effect modifiers of recombination in these regions of the genome, so testing of additional T-DNA lines for alternative candidate genes in this interval, such as those eliminated from the initial candidate list on the basis of function that could have an indirect effect on recombination, could identify a significant modifier. Alternatively, it is possible that measurement of recombination in F_1 lines that contain the T-DNA insertion in a hemizygous state would not show a significant difference from wild-type, even if the insertion was in the gene affecting recombination, as many mutations only show an effect in total knockout lines (such as several *zmm* mutants; Ziolkowski *et al*. 2017). Therefore, measurement of recombination in T-DNA insertion lines with homozygous insertions could also be used to identify lines with significant effects on recombination, particularly for CanQTL4 candidates as the Can-0 allele of this modifier appears to act recessively. If these methods did not identify a significant modifier, screening of additional large fine-mapping populations to identify recombinants could be performed. This could narrow the interval to an area containing only a few genes that it would then be feasible to assess through transformation of alleles.

HEI10 was confirmed as a recombination modifier varying between Col-0 and Cvi-0 accessions, but additional *HEI10* allele T_1 replicates are required to substantiate this and allow quantification of the difference in effect between alleles. Furthermore, while it is known that HEI10 affects recombination, and that the R264G variant is involved in this effect, its mechanism of action is unclear. HEI10 was also shown to affect crossover interference, and it is unclear if this is due to the overall effect on recombination, or if this effect can be separated from upregulation of crossover frequency. Assessment of crossover interference in different *HEI10* allele overexpressor lines could be an interesting experiment, to see if the extent of the increase in recombination correlates directly to levels of crossover interference. Analysis of interference in the *HEI10*^{R264G} overexpressor line and comparison to interference estimates in the *HEI10*^{Col} overexpressor line could clarify if the effect on interference is also linked to this variant, which would suggest an influence of HEI10 substrate recognition by the C-terminal region on interference. Identification of the targets of HEI10, potentially through analysis of cytological co-localisation and yeast two-hybrid experiments, or protein co-immunoprecipitation, would also be useful to determine how HEI10 affects recombination and crossover interference. As it is possible that HEI10 mediates its effect on crossover interference through an effect on the chromosome axis or SC, cytological analysis of meiotic progression in *HEI10* overexpression lines, along with detailed immunostaining of axis proteins and measurement of axis length, could be performed to determine if there is a significant difference from wild-type cells. Previous work has established that meiotic prophase progression appears normal in a *HEI10* overexpression line, and that ASY1 axis protein localisation seems to match wild-type observations,

although bivalents are more compact (Ziolkowski *et al.* 2017; Serra *et al.* 2017). Further analysis of additional proteins and synapsis progression could provide more information, and potentially aid understanding of the effect of *HEI10* on interference.

6.7 Future considerations

Identification of natural recombination modifiers in Arabidopsis provides considerable information about the recombination process and the potential proteins involved, in addition to providing potential candidates that may be responsible for variation in recombination between natural populations. These candidate genes may therefore provide insight into the natural evolution of recombination and the ways that plants under different conditions may have adapted their recombination phenotypes to deal with different selection forces (Alonso-Blanco and Koornneef 2000; Koornneef *et al*. 2004; Mitchell-Olds and Schmitt 2006; Bomblies *et al*. 2015). Analysis of the natural habitats of accessions and correlation of environmental factors with recombination phenotypes could allow identification of specific selection pressures that may be responsible for variation in recombination, thereby providing further information about factors that may influence recombination rates. However, while this data was produced from experiments using accessions that are adapted to different environments and may well exhibit differential recombination rates and patterns in nature, it is important to consider the fact that they were grown under laboratory conditions that do not match their natural conditions (Koornneef *et al*. 2004; Brachi *et al*. 2010; El-Soda *et al*. 2014; Bomblies *et al*. 2015). This may have affected the conclusions drawn by this project, meaning that the modifiers identified may not be responsible for recombination variation under natural conditions (Koornneef *et al*. 2004; Brachi *et al*. 2010; El-Soda *et al*. 2014). If this is indeed the case, identification of these modifiers may still aid understanding of recombination, and also potentially provide insight into factors influencing recombination under stress conditions, but this information may not be relevant to the evolution of natural variation in recombination. Additional experiments performed under field conditions may be required to determine the relevance of these modifiers to natural variation, and therefore their evolutionary significance (Brachi *et al*. 2010; El-Soda *et al*. 2014; Bomblies *et al*. 2015).

In addition to providing interesting information about recombination and meiosis, identification of recombination modifiers also serves a more practical purpose, in providing avenues of experimental manipulation to promote recombination in crop species. Genetic mapping and introgression of agricultural traits in crop species requires high levels of recombination to reduce genetic linkage of loci and prevent linkage drag of undesirable loci during introgression of beneficial variants into an optimised line (Bauer *et al*. 2013; Kunzel *et al*. 2000; Mercier *et al*. 2015). To improve this, modifiers capable of increasing genome-wide recombination need to be identified that could conceivably be

manipulated to alter recombination levels in crops. Recombination is a highly conserved process in eukaryotes (Mercier *et al.* 2015; Villeneuve and Hillers 2001), making it possible that modifiers identified in Arabidopsis could perform the same function in other plant species. Generation of hyper-recombinogenic lines is a goal for crop improvement (Henderson 2012; Serra *et al*. bioRxiv; Bauer *et al*. 2013; Mercier *et al*. 2015), and studying combinations of recombination modifiers in Arabidopsis can help to achieve this. For example, combination of the *HEI10* overexpressor line identified through a natural variation screen (Ziolkowski *et al*. 2017) and the *recq4a recq4b* mutants identified through a mutagenic screen (Seguela-Arnaud *et al*. 2015; Seguela-Arnaud *et al*. 2016) results in a massive increase in recombination, to 3.7-fold higher than wild type levels (Serra *et al*. bioRxiv), which could be extremely beneficial if replicated in crops.

While the use of *HEI10* in combination with other recombination modifiers is simplified by its dosage-dependent effects ensuring that additional transgenic copies have a significant effect on recombination, this would not necessarily be true for all of the modifiers identified in this project (Ziolkowski *et al.* 2017; Figure 51). However, as some of the modifiers identified in this project show semi-dominance, it is possible that the addition of transgenic copies of the alleles associated with higher recombination, such as the Cvi-0 allele of *ARI7*, could produce lines with higher recombination rates (Abiola *et al*. 2003; Maloof 2003). Similarly, if any of the other modifiers are shown to have dosage-dependent effects, it is also possible that transformation of multiple transgenic copies of these modifiers could create lines with even more drastically increased recombination rates, as is observed with *HEI10* overexpression (Ziolkowski *et al*. 2017; Serra *et al*. bioRxiv). Combination of these modifiers with mutations known to affect recombination rate could be used to create hyper-recombinogenic lines that exceed even the 3.7-fold increase observed in *HEI10*-overexpressor *recq4a recq4b* lines (Serra *et al*. bioRxiv).

However, use of these modifiers to create hyper-recombinogenic lines, particularly in crop species, is dependent on a number of caveats. One of the first issues with using hyper-recombination for trait introgression is that beneficial linkage-groups elsewhere in the genome are also likely to be broken up, producing temporarily sub-optimal lines which may complicate the breeding process. In this respect, region-specific modifiers may be of more use, allowing recombination in areas that traditionally experience little shuffling or, conversely, promoting recombination in gene-rich areas without compromising the stability of repeat arrays and other features through ectopic crossovers. Use of any of these modifiers in plant breeding would therefore require thorough characterisation of their effects across the genome following confirmation of the loci as recombination modifiers. Further examination of plants would also be required to ensure that these modifiers were not having pleiotropic effects on the cell beyond effects on recombination, as these would limit the potential uses

of modifiers in experimental manipulation. Identification of the causal variant responsible for the altered recombination phenotype may also be useful here, as it could allow modification of an allele by site-directed mutagenesis to promote changes in recombination while potentially minimising any off-target effects that may be attributed to other variants within the allele.

Even if these modifiers are confirmed to have a significant effect on recombination, either through transformation of transgenic copies or introgression of the allele associated with high recombination into another genetic background, this effect may not necessarily be of use in crop breeding populations. The modifier may have a very small effect on recombination, meaning that its use in experimental manipulation may not be worth the effort of constructing the lines. It is also possible that the effect may be dependent on other factors, such as the chromatin state, that may vary between plants, which may make the modifier ineffective in crop species with vastly different levels of heterochromatin and other features (Henderson 2012). Alternatively, the effect of the modifier may vary depending on environmental conditions, as these effects have only been observed thus far under laboratory conditions (Koornneef *et al*. 2004; El-Soda *et al*. 2014; Bomblies *et al*. 2015). Any of these possibilities could reduce the usefulness of a modifier in a crop breeding programme, where both the genome and environmental conditions are likely to deviate considerably from the conditions tested here, which could mean that these modifiers may only offer limited use in a subset of possible situations (El-Soda *et al.* 2014).

Replication of these effects on recombination in crop species could be attempted in several ways, including transformation of species with a transgenic copy of the Arabidopsis modifier, or introgression or transformation of a species-specific homolog of the modifier from a related strain (Alonso-Blanco and Koornneef 2000; Kearsey and Farquhar 1998; Abiola *et al.* 2003; Wijnker and de Jong 2008; Crismani *et al.* 2013). Each of these approaches has unique advantages and drawbacks that would affect the experimental use of recombination modifiers in crop species. Transformation of the Arabidopsis gene into a crop species to alter recombination does offer some advantages over other strategies, given that the gene has already been identified and characterised and many species have pre-established transformation protocols, making the experiment relatively simple (Hellens *et al.* 2000; Mitchell-Olds and Schmitt 2006). However, this strategy is unlikely to work with some modifiers, as protein interactions and regulatory networks are likely to differ between species, which would prevent the modifier acting as it would within Arabidopsis meiosis (Mitchell-Olds and Schmitt 2006). While this would depend on the gene, as many components of the recombination pathway are highly conserved between species, the probability of the Arabidopsis modifier having variable effects on recombination in addition to potential pleiotropic effects means that this strategy is not without risk.

Alternatively, identification and use of a species-specific homolog of the modifier could reduce the likelihood of off-target effects and allow better integration with other cellular processes, as regulators and protein-interactions are likely to exhibit variation between species and component proteins may co-evolve, ensuring that combinations of species-specific elements produce the most optimal results (Goh and Cohen 2002; Bomblies *et al*. 2015; Mercier *et al*. 2015). Once a homolog was identified, wild strains containing considerable genetic diversity could be examined to find natural alleles that may be responsible for variation in recombination which could then be introgressed into breeding lines or inserted via transformation (Wijnker and de Jong 2008; Prakash and Chopra 1988; Paran and Zamir 2003; Mitchell-Olds 2001). This strategy could be particularly promising in *Brassicacea* crops which are closely related to Arabidopsis, such as oil seed rape, as they are more likely to contain homologs with similar functions and variants due to their genetic similarities (Kearsey and Farquhar 1998; Mitchell-Olds and Schmitt 2006).

Although the addition of transgenic copies of the homolog would appear to be the more efficient strategy, as introgression takes time and linkage drag could result in the inclusion of undesirable traits, current GMO regulations in many nations may limit the use of lines constructed using transgenic methods, whereas lines containing introgressed segments would not be subject to the same restrictions (Jones 2015; Huang *et al*. 2016). However, even disregarding these problems, it is possible that a homolog may not be identifiable in crop species due to differences in the genome and recombination system, making this strategy untenable for some modifiers. For example, while orthologs of *HEI10*, *ARI7* and *TPR8* can be found in other *Brassicacea* species, no ortholog of *A.thaliana PP2C* was found in the preliminary analysis of sequence variants (Ziolkowski *et al*. 2017; JGI Phytozome BLAST, www.phytozome.jgi.doe.gov, 7th March 2018; Supplementary Figures 1 and 2). This could potentially be due to At5g53140 (*PP2C*) being the result of a possible recent gene duplication event in *A.thaliana*, where the function may be the product of subsequent diversification, meaning that it is unlikely to be conserved in related species (Kerk *et al*. 2002). All of these possibilities must be taken into consideration when attempting to use the information gained from a model organism like *A.thaliana* to modify processes in crop species. Ultimately, while these modifiers may offer some practical applications for crop breeding, significant work would have to be performed first to confirm their effects and study homologous systems in crop species before the plausibility of these applications could be assessed.

However, even if the modifiers identified here do not prove useful for crop modification, they could still be valuable for use in Arabidopsis for experimental purposes. Generation of hyper-recombinogenic lines can also be useful for genetic mapping experiments in laboratory plant systems, as it promotes separation of linked variants (Bauer *et al*. 2013; Mercier *et al*. 2015). Increases in

recombination could also speed up the introgression of mutations into different backgrounds which could be extremely beneficial for experiments under time constraints. Therefore, the results of this project are useful both from a perspective of scientific understanding and additional real-world applications. Although there is still a considerable amount of work to be completed before the results of this project could be used for practical applications, this information provides an exciting direction for future research.

Acknowledgements

This work was supported by a Doctoral Training Partnership from the Biotechnology and Biological Sciences Research Council (BBSRC).

I would like to thank Avraham Levy (The Weizmann Institute), Gregory Copenhaver (University of North Carolina Chapel Hill) and Scott Poethig (University of Pennsylvania) for FTL lines, Andy Riddell for assistance with three-colour flow cytometry and the Wellcome-MRC Cambridge Stem Cell Institute for use of their flow cytometer.

I would like to thank my thesis examiners, Phil Howell and Kirsten Bomblies, for their valuable advice.

Thank you to Piotr Ziolkowski, Natasha Yelina and Kyuha Choi for all your help when I first arrived setting up the FTLs and the QTL mapping project, and for your advice and help in the lab.

Thank you to Charlie Underwood (and Piotr) and Emma Lawrence for seed stocks, experimental constructs and useful project discussions.

Thank you to Heidi Serra, Christophe Lambing, Alexander Blackwell and Sabrina Gonzalez-Jorge for helpful lab advice, protocols and a good laugh in the lab.

Thank you to Mel Steer for all your help with the plants and good company at the PGF, and to James Barlow and Pawel Baster for all your help in the lab.

Thank you to Ombretta Orsini for always making me laugh! And thanks to Alex, Luke, Claire, Patrick, Sasha, Emma, Quentin and Divya for all the fun times and company in and out of the lab (when I wasn't being a grouch).

Thank you to all my colleagues and friends in Ian and David's labs for good scientific advice, possibly less useful (and sometimes bizarre) life advice and endless good company at tea time.

Thank you to the members of the Department of Plant Sciences and Christ's college for all your help and good conversation.

To my supervisor, Ian Henderson – thank you for all your advice, useful discussions and help in planning out this project, even when it hasn't always gone to plan!

And last, but not least, thank you to my family – Mum, Dad, Joe, Alfie, Nan, Granddad and Holly for sitting through the tears and the laughter and supporting me every step of the way. And thanks to Tilly, Bonnie and Molly for being there whether it's a good day or not.

Table 1: Fluorescent reporter intervals used for measurement of recombination. Interval is listed with chromosome, location on the chromosome, method of scoring, size of the interval in Mb, and position of the flanking T-DNA fluorescence genes (Melamed-Bessudo *et al*. 2005; Berchowitz and Copenhaver 2008; Wu *et al*. 2015).

Interval	Chromosome	Method	T-DNA 1	T-DNA 2	Interval size (Mb)	Location
1.19	1	Seed	519,052-*dsRed*	6,804,499-*GFP*	6.3	Sub-telomeric
I2f	2	Pollen	18,286,716-*dsRed2*	18,957,093-*YFP*	0.67	Sub-telomeric
420	3	Seed	256,516-*GFP*	5,361,637-*dsRed2*	5.11	Sub-telomeric
3.9	3	Seed	9,741,508-*GFP*	15,980,483-*dsRed*	6.2	Centromeric
CEN3	3	Pollen	11,115,724-*YFP*	16,520,560-*dsRed2*	5.40	Centromeric
5.11	5	Seed	6,501,045-*dsRed*	13,470,052-*GFP*	7	Centromeric

Table 2. List of T-DNA insertion lines ordered for analysis of rQTL2^{420} candidate loci. Salk lines (Alonso *et al*. 2003), SAIL lines (Sessions *et al*. 2002) and GABI-Kat lines (Kleinboelting *et al*. 2012) were obtained from the Nottingham Arabidopsis Stock Centre (NASC). NASC stock ID is given for each seed line.

T-DNA insertion line	NASC ID
Salk 082541	N675962
Salk 027620	N665648
Salk 145153	N667547
GK 380E06	N436438
Salk 024459	N685479
Salk 059528	N681156
Salk 054181	N662501
Salk 109522	N684414
Salk 050436	N679649
GK 345C06	N433054
Salk 014797	N682028
Salk 064633	N669790

Table 3. List of T-DNA insertion lines ordered for analysis of rQTL5^{420} candidate loci. Salk lines (Alonso *et al*. 2003), SAIL lines (Sessions *et al*. 2002) and GABI-Kat lines (Kleinboelting *et al*. 2012) were obtained from the Nottingham Arabidopsis Stock Centre (NASC). NASC stock ID is given for each seed line.

T-DNA insertion line	NASC ID
Salk 010368	N673798
Salk 090163	N653356
Salk 048952	N655507
SAIL 369C12	N860756

Table 4. List of T-DNA insertion lines ordered for analysis of CanQTL4 candidate loci. Salk lines (Alonso *et al*. 2003), SAIL lines (Sessions *et al*. 2002) and GABI-Kat lines (Kleinboelting *et al*. 2012) were obtained from the Nottingham Arabidopsis Stock Centre (NASC). NASC stock ID is given for each seed line.

T-DNA insertion line	NASC ID
Salk 146231	N658456
Salk 047663	N682317
GK 219G07	N421007
SAIL 731H04	N861157

Table 5. Fluorescent seed count data for the *420* FTL interval in Col-0/*420*, Cvi-0/*420* and Can-0/*420* F_1 lines. Genetic distance of the *420* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015). The aggregate number of recombinant (G+R) and non-recombinant (D+N) seeds for Col-0/*420* and Cvi-0/*420,* and Col-0/*420* and Can-0/*420* lines were used to construct 2x2 contingency tables and perform two-tailed χ^2 tests to test for significant differences.

Genotype	Green (G)	Red (R)	Double colour (D)	Non-colour (N)	Total (T)	cM
Col-0/*420*	143	134	1035	257	1569	19.57
Col-0/*420*	107	112	861	230	1310	18.41
Col-0/*420*	113	101	986	256	1456	15.97
Col-0/*420*	127	119	993	249	1488	18.19
Col-0/*420*	127	138	1041	247	1553	18.84
Col-0/*420*	118	112	928	232	1390	18.20
Col-0/*420*	119	131	948	243	1441	19.19
Col-0/*420*	108	108	1022	278	1516	15.44
Col-0/*420*	133	128	973	244	1478	19.57
Col-0/*420*	116	126	1069	274	1585	16.66
Cvi-0/*420*	89	95	1039	244	1467	13.45
Cvi-0/*420*	106	91	1092	266	1555	13.59
Cvi-0/*420*	86	88	1088	282	1544	11.99
Cvi-0/*420*	92	70	1345	285	1792	9.49
Cvi-0/*420*	61	81	869	230	1241	12.18
Cvi-0/*420*	91	82	1094	294	1561	11.78
Cvi-0/*420*	76	79	1105	290	1550	10.56
Cvi-0/*420*	71	98	1117	287	1573	11.39
Cvi-0/*420*	79	87	1177	305	1648	10.64
Cvi-0/*420*	70	95	955	223	1343	13.15
Can-0/*420*	55	64	832	246	1197	10.49
Can-0/*420*	75	65	1080	307	1527	9.63
Can-0/*420*	72	87	1081	314	1554	10.82
Can-0/*420*	86	92	1053	306	1537	12.34
Can-0/*420*	49	51	665	193	958	11.05
Can-0/*420*	72	61	775	222	1130	12.56

Can-0/*420*	72	83	1072	284	1511	10.85
Can-0/*420*	69	66	951	296	1382	10.30
Can-0/*420*	66	70	970	267	1373	10.45
Can-0/*420*	86	85	993	300	1464	12.46

Cvi-0/*420* $\chi^2(1) = 180.3919$, $p=7.96 \times 10^{-41}$

Can-0/*420* $\chi^2(1) = 211.4977$, $p=1.29 \times 10^{-47}$

Table 6. Fluorescent seed count data for the *5.11* CTL interval in Col-0/*5.11*, Cvi-0/*5.11* and Can-0/*5.11* F_1 lines. Genetic distance of the *5.11* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015). The aggregate number of recombinant (G+R) and non-recombinant (D+N) seeds for Col-0/*5.11* and Cvi-0/*5.11,* and Col-0/*5.11* and Can-0/*5.11* lines were used to construct 2x2 contingency tables and perform two-tailed χ^2 tests to test for significant differences.

Genotype	Green (G)	Red (R)	Double colour (D)	Non-colour (N)	Total (T)	cM
Col-0/*5.11*	156	168	1193	286	1803	19.96
Col-0/*5.11*	173	176	1195	281	1825	21.42
Col-0/*5.11*	145	156	1066	290	1657	20.21
Col-0/*5.11*	162	172	1177	282	1793	20.79
Col-0/*5.11*	153	139	1104	281	1677	19.27
Col-0/*5.11*	157	141	1123	286	1707	19.32
Col-0/*5.11*	153	152	1235	297	1837	18.27
Col-0/*5.11*	174	145	1142	287	1748	20.31
Col-0/*5.11*	137	153	1169	297	1756	18.16
Col-0/*5.11*	163	165	1285	292	1905	19.03
Cvi-0/*5.11*	199	208	1191	240	1838	25.36
Cvi-0/*5.11*	184	167	1029	234	1614	24.83
Cvi-0/*5.11*	201	213	1159	253	1826	26.07
Cvi-0/*5.11*	217	256	1291	263	2027	26.97
Cvi-0/*5.11*	209	223	1168	243	1843	27.12
Cvi-0/*5.11*	207	201	1107	252	1767	26.64
Cvi-0/*5.11*	182	197	1086	236	1701	25.54
Cvi-0/*5.11*	232	194	1162	263	1851	26.54
Cvi-0/*5.11*	234	247	1289	270	2040	27.31
Cvi-0/*5.11*	184	213	1190	247	1834	24.70
Can-0/*5.11*	113	117	926	207	1363	18.61
Can-0/*5.11*	150	116	985	233	1484	19.91
Can-0/*5.11*	130	124	973	235	1462	19.22
Can-0/*5.11*	157	136	1026	264	1583	20.64
Can-0/*5.11*	159	140	1030	263	1592	20.98
Can-0/*5.11*	113	127	1050	247	1537	17.07

Can-0/*5.11*	153	147	1095	262	1657	20.13
Can-0/*5.11*	159	144	1062	241	1606	21.09
Can-0/*5.11*	146	141	981	254	1522	21.08
Can-0/*5.11*	129	134	950	243	1456	20.08

Cvi-0/*5.11* $\chi^2(1) = 138.9628$, $p=8.98 \times 10^{-32}$

Can-0/*5.11* $\chi^2(1) = 0.198319$, $p=0.6561$

Table 7. Fluorescent seed count data for the *3.9* CTL interval in Col-0/*3.9* and Cvi-0/*3.9* F_1 lines. Genetic distance of the *3.9* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015). The aggregate number of recombinant (G+R) and non-recombinant (D+N) seeds for Col-0/*3.9* and Cvi-0/*3.9* lines were used to construct a 2x2 contingency table and perform a two-tailed χ^2 test to test for significant differences.

Genotype	Green (G)	Red (R)	Double colour (D)	Non-colour (N)	Total (T)	cM
Col-0/*3.9*	105	105	941	242	1393	16.42
Col-0/*3.9*	126	101	1026	273	1526	16.19
Col-0/*3.9*	117	110	1031	274	1532	16.12
Col-0/*3.9*	98	100	1040	285	1523	13.98
Col-0/*3.9*	137	135	1060	260	1592	18.86
Col-0/*3.9*	106	121	1088	277	1592	15.45
Col-0/*3.9*	105	130	1117	303	1655	15.38
Col-0/*3.9*	116	114	1098	284	1612	15.46
Col-0/*3.9*	112	98	1038	276	1524	14.89
Col-0/*3.9*	114	113	976	258	1461	16.98
Cvi-0/*3.9*	248	155	1058	189	1650	28.48
Cvi-0/*3.9*	199	150	1007	167	1523	26.40
Cvi-0/*3.9*	215	170	1100	191	1676	26.48
Cvi-0/*3.9*	186	141	976	170	1473	25.43
Cvi-0/*3.9*	176	164	987	218	1545	25.18
Cvi-0/*3.9*	188	156	1104	219	1667	23.37
Cvi-0/*3.9*	162	164	968	191	1485	25.10
Cvi-0/*3.9*	225	154	1018	199	1596	27.54
Cvi-0/*3.9*	212	170	1074	198	1654	26.65
Cvi-0/*3.9*	202	185	1065	151	1603	28.09

$\chi^2(1) = 338.760478$, $p=1.19 \times 10^{-75}$

Table 8. Fluorescent seed count data for the *1.19* CTL interval in Col-0/*1.19* and Cvi-0/*1.19* F_1 lines. Genetic distance of the *1.19* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015). The aggregate number of recombinant (G+R) and non-recombinant (D+N) seeds for Col-0/*1.19* and Cvi-0/*1.19* lines were used to construct a 2x2 contingency table and perform a two-tailed χ^2 test to test for significant differences.

Genotype	Green (G)	Red (R)	Double colour (D)	Non-colour (N)	Total (T)	cM
Col-0/*1.19*	120	128	861	208	1317	21.05
Col-0/*1.19*	114	116	1052	273	1555	16.08
Col-0/*1.19*	139	146	974	235	1494	21.36
Col-0/*1.19*	159	151	1010	250	1570	22.21
Col-0/*1.19*	121	141	1031	254	1547	18.68
Col-0/*1.19*	114	120	1020	257	1511	16.92
Col-0/*1.19*	140	128	1022	258	1548	19.15
Col-0/*1.19*	123	123	989	256	1491	18.15
Col-0/*1.19*	122	137	997	247	1503	19.05
Col-0/*1.19*	118	112	946	231	1407	17.96
Cvi-0/*1.19*	119	122	985	316	1542	17.09
Cvi-0/*1.19*	97	100	1058	299	1554	13.60
Cvi-0/*1.19*	105	108	1065	328	1606	14.28
Cvi-0/*1.19*	119	113	883	242	1357	18.88
Cvi-0/*1.19*	91	70	633	171	965	18.37
Cvi-0/*1.19*	114	130	835	220	1299	20.99
Cvi-0/*1.19*	100	118	864	258	1340	17.86

$\chi^2(1) = 11.23291$, p=0.000804

Table 9. Fluorescent pollen flow cytometry count data for the *CEN3* FTL interval in Col-0/*CEN3* and Cvi-0/*CEN3* F_1 lines. Genetic distance of the *CEN3* interval is calculated as cM= 100 x (2 x R_6)/(R_2 - (R_5 - R_4)). R_2 is the total number of pollen analysed, R_4 is the number of pollen expressing both eYFP and RFP, R_5 is the number of pollen not expressing eYFP or RFP and R_6 is the number of pollen expressing eYFP only (Yelina *et al*. 2013). The aggregate number of recombinant (R_6) and non-recombinant (R_4) pollen for Col-0/*CEN3* lines and Cvi-0/*CEN3* lines were used to construct 2x2 contingency tables and perform a two-tailed χ^2 test to test for significant differences.

Genotype	Total (R_2)	eYFP/RFP (R_4)	Non-colour (R_5)	eYFP (R_6)	cM
Col-0/*CEN3*	27254	8934	14926	1219	11.47
Col-0/*CEN3*	26708	9642	13615	1288	11.33
Col-0/*CEN3*	23078	8566	11564	1121	11.17
Col-0/*CEN3*	17171	5644	9397	811	12.09
Col-0/*CEN3*	9373	3005	5325	342	9.70
Col-0/*CEN3*	13572	5308	6673	671	10.99
Col-0/*CEN3*	8673	3027	4665	356	10.12
Col-0/*CEN3*	6084	2067	3311	219	9.05
Col-0/*CEN3*	6486	2109	3676	277	11.26
Col-0/*CEN3*	14059	5256	7010	719	11.69
Cvi-0/*CEN3*	10866	2988	5761	938	23.18
Cvi-0/*CEN3*	31107	7154	18490	2340	23.67
Cvi-0/*CEN3*	50111	10373	30240	2544	16.82
Cvi-0/*CEN3*	28914	8228	15854	2118	19.90
Cvi-0/*CEN3*	25971	8068	11989	2884	26.16
Cvi-0/*CEN3*	30587	9327	15223	2969	24.05
Cvi-0/*CEN3*	27678	8206	14660	2296	21.64
Cvi-0/*CEN3*	45698	13288	24253	3902	22.47
Cvi-0/*CEN3*	27571	8508	14159	2432	22.19

$\chi^2(1) = 3097.34$, p=0.00

Table 10. Fluorescent pollen flow cytometry count data for the *I2f* FTL interval in Col-0/*I2f* and Cvi-0/*I2f* F_1 lines. Genetic distance of the *I2f* interval is calculated as cM= 100 x (2 x R_6)/(R_2 - (R_5 - R_4)). R_2 is the total number of pollen analysed, R_4 is the number of pollen expressing both eYFP and RFP, R_5 is the number of pollen not expressing eYFP or RFP and R_6 is the number of pollen expressing eYFP only (Yelina *et al*. 2013). The aggregate number of recombinant (R_6) and non-recombinant (R_4) pollen for Col-0/*I2f* lines and Cvi-0/*I2f* lines were used to construct 2x2 contingency tables and perform a two-tailed χ^2 test to test for significant differences.

Genotype	Total (R2)	eYFP/RFP (R4)	Non-colour (R5)	eYFP (R6)	cM
Col-0/*I2f*	10009	4085	4992	337	7.40
Col-0/*I2f*	7438	3037	3786	236	7.06
Col-0/*I2f*	8419	3385	4276	278	7.39
Col-0/*I2f*	10414	4097	5331	354	7.71
Col-0/*I2f*	12053	4053	6585	341	7.16
Col-0/*I2f*	7283	2559	4015	211	7.24
Col-0/*I2f*	6486	2422	3395	181	6.57
Col-0/*I2f*	9625	3380	5152	246	6.27
Cvi-0/*I2f*	11287	4076	6168	500	10.88
Cvi-0/*I2f*	12009	3313	7484	527	13.45
Cvi-0/*I2f*	11961	3355	7104	491	11.96
Cvi-0/*I2f*	17310	4750	10664	849	14.90
Cvi-0/*I2f*	8012	1874	5224	319	13.69
Cvi-0/*I2f*	9120	2764	5382	408	12.55
Cvi-0/*I2f*	10687	2831	6794	515	15.32
Cvi-0/*I2f*	6616	2044	4038	204	8.83
Cvi-0/*I2f*	12144	4027	6849	590	12.66
Cvi-0/*I2f*	15453	4481	9074	928	17.09

$\chi^2(1) = 661.68$, $p=6.44 \times 10^{-146}$

Table 11: Fluorescent pollen flow cytometry count data for the *CEN3* FTL interval in Cvi-0/*CEN3* F_2 population. Genetic distance of the *CEN3* interval is calculated as cM = 100*(2*Y)/(T-(N-D)). Y is eYFP positive pollen, T is the total number of pollen, N is the number of eYFP and RFP negative pollen and D is the number of eYFP and RFP positive pollen (Yelina *et al*. 2013).

Individual	Total (T)	Double colour (D)	Non-colour (N)	Yellow (Y)	cM
1-2	5792	1788	2965	414	17.94
2-4	6902	2041	3590	422	15.77
3-2	17156	5686	7785	1584	21.04
3-6	12712	2844	7347	970	23.63
4-2	12368	4593	6453	504	9.59
4-3	11349	3651	5824	756	16.48
4-5	11429	3152	5761	1086	24.63
6-5	15528	4419	8505	916	16.01
6-6	16798	4870	8894	970	15.19
7-2	6223	2208	2981	427	15.67
7-6	21891	6104	11997	1309	16.36
8-6	6835	2156	3575	383	14.14
10-2	7206	1917	3825	514	19.40
10-6	10069	3126	5463	542	14.02
11-2	22748	4435	14123	1629	24.95
12-6	7819	2310	4108	529	17.57
13-1	12922	3647	7418	566	12.37
16-2	9056	1992	5046	473	15.76
18-4	7949	1941	4440	718	26.35
18-5	10371	2559	5411	718	19.10
18-6	9815	2364	5831	451	14.21
19-1	20852	4207	10803	1623	22.77
19-4	8103	1665	4919	395	16.29
20-2	5653	1260	3344	286	16.03
21-4	7172	2015	3906	332	12.57
22-6	4002	1050	2154	364	25.12
23-3	6478	2226	3112	467	16.70
23-4	15543	3874	8279	1144	20.54

24-2	6294	1908	2988	499	19.14
24-6	9385	1679	5400	797	28.14
25-3	8256	2297	4298	504	16.12
25-4	7148	1221	4118	723	34.02
25-5	7333	2005	3749	433	15.49
26-1	10995	3379	5525	791	17.88
26-4	7701	2014	4156	624	22.45
27-4	8564	1668	4711	562	20.36
27-6	9909	1003	6571	901	41.51
28-6	6340	1764	3226	513	21.03
32-2	6051	2020	3183	307	12.56
33-4	4744	1293	2501	372	21.04
34-1	6683	1359	4139	249	12.76
34-3	6877	1425	4529	341	18.08
34-5	8549	2046	4981	455	16.21
35-4	7681	2199	3744	490	15.97
35-5	7795	1468	4463	471	19.63
36-5	8828	2056	5045	522	17.88
36-6	14758	3561	8788	435	9.13
38-5	13764	2914	8352	714	17.15
39-1	10811	3047	6208	572	14.95
39-5	18591	4714	10055	999	15.08
39-6	6956	2641	3539	320	10.56
40-2	8992	3010	4707	453	12.42
40-4	21863	2298	18518	449	15.91
41-5	15906	4046	9782	663	13.04
42-2	10546	3373	5205	815	18.71
43-1	7698	2290	3833	581	18.88
43-M	8792	2652	5091	347	10.92
43-4	12280	2972	7217	790	19.66
44-5	10717	3015	5776	737	18.53
45-2	14117	5156	6485	1084	16.95
46-2	7744	2534	4082	552	17.82

46-4	10474	2423	5608	803	22.03
46-6	7976	2016	4335	744	26.30
47-3	14509	5181	6585	1247	19.03
47-4	7106	2042	3518	619	21.99
48-1	20629	5149	12137	1097	16.08
48-4	16007	4940	8876	823	13.64
49-4	11984	3258	5990	1017	21.98
49-5	23723	7647	12953	1109	12.04
50-4	14219	4460	7462	981	17.49
50-5	19057	6134	9980	1146	15.07
52-6	23248	7283	10676	2084	20.99
53-4	41591	10899	24002	2269	15.93
54-6	13890	4272	7351	862	15.95
55-1	16073	4641	7922	1202	18.79
55-2	14074	4901	6636	991	16.06
55-5	17408	4520	9727	864	14.16
56-1	22024	5981	12103	1515	19.05
58-1	8642	2044	4961	789	27.56
58-3	9526	2842	5072	640	17.54
60-1	10976	2516	5975	1052	27.99
60-2	15600	4675	8047	990	16.19
60-5	18826	3260	12278	785	16.01
62-4	9261	2589	4579	708	19.47
63-2	19052	7001	9720	722	8.84
63-4	29302	10025	14945	1651	13.54
64-6	8986	3063	4738	468	12.80
66-6	13858	3548	7181	1202	23.51
67-2	17248	5160	9107	1224	18.40
68-2	25869	9013	12734	1753	15.83
68-3	13141	3381	7039	995	20.98
70-2	10524	2774	5569	599	15.50
70-4	13694	2845	8957	691	18.23
70-5	7065	2204	3317	695	23.35

71-1	22797	7319	12173	1155	12.87
72-3	15172	5630	7790	664	10.21
72-5	18835	5267	9384	1648	22.39
73-2	12807	3005	7026	1077	24.52
73-6	28901	8601	15830	1862	17.18
74-3	8185	2251	4141	638	20.27
74-4	14418	3821	7328	1289	23.63
75-1	19767	6175	10284	1132	14.46
75-3	24335	6045	13162	1888	21.93
75-4	17307	3556	10957	658	13.28
76-1	11452	3964	5981	649	13.76
78-2	10588	3026	5713	626	15.85
78-4	14118	4027	8098	623	12.40
78-5	18195	5566	9741	950	13.55
80-3	16593	4531	9254	916	15.43
83-1	43964	14487	21760	3169	17.27
84-6	19194	6085	9970	1123	14.67
85-1	8605	2670	4164	737	20.73
85-5	28892	7887	15886	2085	19.96
86-2	15299	4246	8805	768	14.30
87-5	24684	7257	13374	1764	19.00
88-1	29636	7593	17198	1672	16.69
88-3	10924	2574	7315	283	9.15
90-3	27244	7218	16118	1109	12.09
91-3	20696	7016	9781	1609	17.95
91-5	20417	5018	11169	1697	23.79
95-6	27106	7325	14805	1542	15.71
97-4	7406	1959	4053	375	14.12
97-5	9508	2509	4911	772	21.73
97-6	34225	11957	18400	1608	11.58
98-1	13495	3601	7642	1067	22.57
98-2	22189	7398	10856	1690	18.04
99-5	6475	1762	3688	423	18.60

99-6	11784	4173	6080	671	13.59
100-1	13339	3520	7165	961	19.83
100-2	14039	4299	6702	1266	21.76
100-4	21328	6074	11401	1838	22.97
101-4	19432	5733	11037	854	12.09
101-5	25958	7900	14371	1258	12.91
105-1	28812	10137	15188	1356	11.41
105-3	11357	2757	6722	615	16.64
106-1	9310	3144	4284	886	21.69
106-3	16220	5259	8601	917	14.24
106-5	9327	3065	5092	451	12.36
106-6	25542	6314	15646	1743	21.51
108-5	27434	9654	13647	1768	15.08
109-4	11285	3082	6835	385	10.22
109-5	20774	6308	12486	614	8.41
110-2	11746	3956	5773	982	19.78
110-6	6818	2080	3610	481	18.19
111-1	9368	2707	5458	555	16.77
114-1	23217	8021	10833	2067	20.26
115-4	15325	5275	7695	982	15.22
116-1	30469	9963	14727	2454	19.09
116-4	21303	7325	10932	1218	13.77
116-6	13766	4900	6587	1075	17.80
117-4	14478	4851	6537	1489	23.28
118-1	18052	5255	9850	956	14.21
118-2	6465	1804	3768	338	15.02
118-4	17802	5537	9379	1103	15.80
122-5	20381	6811	9910	1636	18.93
127-6	9425	2141	5971	523	18.70
129-1	6768	1812	3756	446	18.49
131-4	16011	5053	8232	1113	17.35
132-3	16387	5068	8261	1298	19.68
133-1	11169	3434	6066	691	16.19

133-3	23821	8935	11094	1753	16.19
134-1	15463	5057	6945	1693	24.94

Table 12. Chromosome 3 marker genotype counts from hot and cold quartile Cvi-0/*CEN3* F_2 individuals. The number of individuals with homozygosity (HOM) or heterozygosity (HET) for the chromosome 3 genotyping markers, in the F_2 population quartiles with highest (Hot) and lowest (Cold) recombination. The p value was calculated by constructing a 2 x 2 contingency table for homozygous and heterozygous genotype counts in the Hot and Cold quartiles and performing a two-tailed χ^2 test for significant differences. FDR correction for multiple testing was then applied (Ziolkowski *et al.* 2015).

Marker coordinates (bp)	Hot quartile HET	Hot quartile HOM	Cold quartile HET	Cold quartile HOM	FDR p value
1031000	20	19	21	19	1.14096375
1746000	21	17	25	14	0.953766667
2718000	21	19	25	15	1.045045714
3621000	19	17	21	16	1.331474545
4126000	21	16	22	18	1.252692857
4715000	23	16	23	17	1.191030667
7638000	23	17	32	8	0.149675
8140000	25	14	33	7	0.256864
8935000	25	13	37	3	0.06998
9404000	21	10	28	3	0.19318
10695000	33	5	36	4	1.325834
11649000	37	0	39	0	1.176470588
12356000	38	0	39	1	1.088943333
15949000	39	0	40	0	1.111111111
16679000	37	0	40	0	1.052631579
17223000	31	9	38	2	0.23046
18459000	27	11	31	8	0.97656
19064000	13	5	7	2	1.26114
19165000	26	11	26	11	1
21008000	23	17	24	16	1.260952308

Table 13: Summary of findings from Cvi-0/*CEN3* F_2 QTL scan. QTL interactions tested for by Hayley-Knott regression, according to the full model y= $rQTL1^{CEN3}$ + $rQTL2^{CEN3}$ + $rQTL3^{CEN3}$ + $rQTL5^{CEN3}$ + $rQTL1^{CEN3}$: $rQTL2^{CEN3}$ + $rQTL1^{CEN3}$: $rQTL3^{CEN3}$ + $rQTL1^{CEN3}$: $rQTL5^{CEN3}$ + $rQTL2^{CEN3}$: $rQTL3^{CEN3}$ + $rQTL2^{CEN3}$: $rQTL5^{CEN3}$ + $rQTL3^{CEN3}$: $rQTL5^{CEN3}$, where : designates an interaction. Analysis of significance performed using a drop-one-term ANOVA table. QTL, or QTL interaction, is listed with chromosome, LOD score, percentage of variance explained and F-test p-value.

Chromosome	QTL	LOD	% Variance	p value (F)
1	$rQTL1^{CEN3}$	8.64	10.9	0.002752
2	$rQTL2^{CEN3}$	14.6	20.1	6.14E-07
3	$rQTL3^{CEN3}$	10.2	13.1	0.000367
5	$rQTL5^{CEN3}$	9.4	12	0.001036
1:2	$rQTL1^{CEN3}$: $rQTL2^{CEN3}$	1.42	1.61	0.263483
1:3	$rQTL1^{CEN3}$: $rQTL3^{CEN3}$	1.38	1.57	0.276434
1:5	$rQTL1^{CEN3}$: $rQTL5^{CEN3}$	0.48	0.54	0.777599
2:3	$rQTL2^{CEN3}$: $rQTL3^{CEN3}$	1.28	1.45	0.316584
2:5	$rQTL2^{CEN3}$: $rQTL5^{CEN3}$	0.35	0.39	0.865316
3:5	$rQTL3^{CEN3}$: $rQTL5^{CEN3}$	0.93	1.05	0.48872

Table 14: Fluorescent pollen count data for the *CEN3* FTL interval in Can-0/*CEN3* F_2 population. Genetic distance of the *CEN3* interval is calculated as cM = 100*(2*Y)/(T-(N-D)). Y is eYFP positive pollen, T is the total number of pollen, N is the number of eYFP and RFP negative pollen and D is the number of eYFP and RFP positive pollen (Yelina *et al*. 2013).

Individual	Total (T)	Double colour (D)	Non-colour (N)	Yellow (Y)	cM
1-1	8471	1701	5125	796	31.54
3-2	6243	1605	3599	444	20.90
4-5	16109	4496	9277	999	17.64
7-6	18070	6188	9641	871	11.92
8-2	12920	4250	7157	556	11.11
9-2	5243	1502	3005	242	12.94
9-6	10521	3402	5307	683	15.85
12-2	13587	5083	7001	656	11.24
17-3	4760	1221	2784	296	18.52
18-2	9387	3379	5028	350	9.05
18-6	16986	3721	10365	1349	26.09
19-1	15539	5297	8301	955	15.24
19-6	14499	3569	8544	976	20.50
24-4	7997	2596	4498	352	11.55
25-1	7414	2145	4377	270	10.42
25-5	10580	4012	5511	437	9.62
32-6	11706	5034	5256	633	11.02
38-2	8391	2900	4128	626	17.48
39-6	16575	4502	9781	862	15.26
40-6	16036	5321	8513	768	11.96
48-1	12562	3669	7492	538	12.31
50-2	14185	4307	8023	682	13.03
50-5	6562	2107	3618	238	9.42
51-5	12528	2085	8055	1163	35.47
53-4	14098	5288	7009	688	11.12
55-4	12389	3608	7122	629	14.17
60-2	15157	4458	8604	856	15.55
65-1	5339	1475	3230	269	15.01

69-1	11313	4067	6080	522	11.23
69-5	9026	2593	5148	524	16.20

Table 15: Fluorescent pollen count data for the *I2f* FTL interval in Cvi-0/*I2f* F_2 population. Genetic distance of the *I2f* interval is calculated as cM = 100*(2*Y)/(T-(N-D)). Y is eYFP positive pollen, T is the total number of pollen, N is the number of eYFP and RFP negative pollen and D is the number of eYFP and RFP positive pollen (Yelina *et al*. 2013).

Individual	Total (T)	Double colour (D)	Non-colour (N)	Yellow (Y)	cM
2-5	6059	1748	3330	140	6.25
4-1	6575	2392	3593	116	4.32
6-2	11108	3289	6164	428	10.40
7-5	6074	2011	3356	206	8.71
8-5	6682	2275	3695	332	12.62
8-6	6011	2386	3170	118	4.52
9-5	8537	2063	5553	408	16.17
9-6	10961	3921	6041	402	9.09
11-1	5431	1353	3283	348	19.88
11-5	10827	4309	5680	275	5.82
12-4	3776	1316	2217	50	3.48
13-1	10239	3522	5851	340	8.60
13-5	3825	1027	2257	85	6.55
14-4	10843	3814	5944	328	7.53
16-2	5251	2031	2783	108	4.80
16-5	2940	1165	1487	70	5.35
18-3	13050	5074	6631	456	7.94
18-4	5514	2127	2951	116	4.95
18-6	11832	5027	6239	171	3.22
20-3	14079	5313	7481	577	9.69
20-4	12853	5067	6467	605	10.56
21-1	6808	2445	3573	149	5.25
21-3	5999	1615	3327	164	7.65
21-4	5355	2037	2831	158	6.93
22-1	13817	2946	8374	206	4.91
22-5	14896	4923	8251	338	5.84
23-3	11566	3880	6466	189	4.21
26-5	12486	4785	4836	1045	16.81

27-3	12754	4057	7198	382	7.95
28-2	11436	4003	6291	223	4.88
28-5	13526	4589	7683	204	3.91
29-4	17285	6018	9599	732	10.68
30-1	12240	4051	7057	263	5.70
31-1	20375	7301	11792	196	2.47
31-5	14872	4422	9056	323	6.31
32-3	14373	5138	7450	742	12.30
34-3	13716	4212	7739	831	16.31
36-1	12223	2290	7746	303	8.96
36-2	20030	3555	12209	706	12.41
37-2	12661	3827	5179	1405	24.85
37-6	17205	3364	10739	462	9.40
38-1	8864	2475	5033	141	4.47
39-3	10099	2950	4000	607	13.42
40-5	8195	2552	4700	434	14.35
40-6	12573	3573	7125	447	9.91
41-2	5983	1570	3608	341	17.29
41-3	13611	3588	8102	526	11.56

Table 16. Fluorescent seed count data for the *420* FTL interval in the Cvi-0/*420* F_2 population. Genetic distance of the *420* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015).

Individual	Green (G)	Red (R)	Double colour (D)	Non-colour (N)	Total (T)	cM
1-1	71	90	1195	249	1605	10.59
1-2	58	52	1193	256	1559	7.32
2-2	79	101	1032	192	1404	13.77
2-4	81	77	1481	343	1982	8.32
2-5	114	153	1042	217	1526	19.37
2-6	75	80	1460	376	1991	8.11
3-4	171	197	1035	153	1556	27.41
5-1	161	208	1298	219	1886	21.98
5-5	57	41	884	215	1197	8.55
6-1	41	51	1337	286	1715	5.52
6-2	112	120	912	158	1302	19.77
6-4	101	111	845	197	1254	18.64
7-2	106	129	914	185	1334	19.52
7-3	123	156	1141	212	1632	18.88
7-4	107	107	1134	220	1568	14.73
7-5	108	122	1454	349	2033	12.04
8-1	102	151	1416	295	1964	13.84
8-2	113	119	1359	287	1878	13.23
8-5	138	170	1289	275	1872	18.09
8-6	68	71	912	171	1222	12.11
9-2	120	120	1348	248	1836	14.06
9-3	93	145	1004	183	1425	18.39
9-4	92	93	1062	223	1470	13.50
9-5	161	185	1295	246	1887	20.42
9-6	111	113	1406	259	1889	12.66
10-2	102	79	951	203	1335	14.63
10-3	74	77	1424	277	1852	8.52
10-5	81	103	985	201	1370	14.48

10-6	142	163	1000	170	1475	23.42
11-1	142	132	1195	219	1688	17.82
11-5	145	117	1558	298	2118	13.25
12-1	104	139	899	170	1312	20.65
12-2	43	46	1439	416	1944	4.69
12-5	57	36	1129	268	1490	6.45
12-6	72	63	1316	350	1801	7.80
13-1	55	80	1608	392	2135	6.54
13-5	185	194	1344	210	1933	22.03
13-6	189	182	1419	291	2081	19.79
14-4	92	89	1215	240	1636	11.75
14-5	90	73	1268	303	1734	9.89
15-1	53	77	972	214	1316	10.42
15-2	76	72	824	191	1163	13.66
15-5	48	47	766	167	1028	9.71
15-6	73	60	928	205	1266	11.12
16-3	55	65	1043	259	1422	8.83
16-4	121	112	1025	249	1507	16.89
16-6	97	101	1108	229	1535	13.86
17-1	41	36	921	249	1247	6.38
17-2	57	66	957	218	1298	9.97
17-4	110	125	1011	215	1461	17.64
17-6	52	63	878	182	1175	10.32
18-2	80	81	1062	248	1471	11.62
18-4	71	72	1294	294	1731	8.63
19-1	58	93	1049	270	1470	10.86
19-2	104	96	909	209	1318	16.54
19-4	47	44	983	234	1308	7.22
19-5	59	64	1227	310	1660	7.71
20-2	44	50	1144	251	1489	6.53
20-5	37	37	944	213	1231	6.20
20-6	121	163	1099	213	1596	19.74
21-1	100	118	989	238	1445	16.44

21-2	71	67	928	198	1264	11.59
21-5	113	124	1114	209	1560	16.56
22-1	56	58	869	207	1190	10.09
22-6	52	49	879	209	1189	8.89
24-2	67	78	1222	312	1679	9.05
24-4	48	29	1062	247	1386	5.72
25-1	77	91	983	229	1380	13.02
26-1	91	77	1383	390	1941	9.07
26-4	89	91	878	213	1271	15.34
26-5	93	121	825	149	1188	20.02
26-6	119	150	1018	178	1465	20.45
27-2	93	106	868	158	1225	17.84
27-3	32	43	937	241	1253	6.18
28-1	104	96	1268	252	1720	12.40
29-1	151	169	1321	228	1869	18.91
29-2	104	99	1179	267	1649	13.18
29-5	62	52	1427	434	1975	5.95
30-1	107	116	986	163	1372	17.85
30-2	60	60	1162	230	1512	8.28
30-3	58	43	1068	259	1428	7.34
30-5	56	62	915	201	1234	10.07
31-1	117	88	1294	328	1827	11.93
31-2	98	96	901	264	1359	15.47
31-4	96	117	1131	228	1572	14.62
32-1	148	147	1296	253	1844	17.54
32-4	114	96	1209	263	1682	13.38
32-6	75	89	1309	267	1740	9.92
33-1	50	57	1292	272	1671	6.62
33-5	26	31	881	203	1141	5.13
33-6	80	73	940	259	1352	12.04
34-2	107	93	1191	295	1686	12.66
34-3	78	78	1183	257	1596	10.31
34-4	153	170	1259	183	1765	20.38

34-5	83	96	884	225	1288	15.03
34-6	84	102	1266	238	1690	11.69
35-2	59	58	906	183	1206	10.22
35-5	102	113	1109	237	1561	14.88
35-6	49	60	1013	238	1360	8.36
36-1	124	128	1298	281	1831	14.87
36-2	105	111	996	181	1393	16.94
36-5	172	171	1051	187	1581	24.76
38-4	58	57	1103	223	1441	8.33
38-5	160	166	1047	191	1564	23.64
39-2	132	136	936	186	1390	21.62
39-5	147	164	1220	207	1738	19.87
40-3	70	72	1254	330	1726	8.60
40-4	59	50	1277	350	1736	6.49
40-5	83	85	918	181	1267	14.28
41-2	120	152	1214	294	1780	16.67
41-3	56	76	1283	277	1692	8.13
41-4	152	189	692	142	1175	35.23
41-6	99	119	1084	257	1559	15.13
42-2	54	58	1126	281	1519	7.67
42-4	197	197	1108	255	1757	25.74
43-1	53	78	1125	332	1588	8.62
44-1	101	123	985	215	1424	17.21
44-2	66	58	1087	229	1440	9.02
44-4	64	77	1105	236	1482	10.02
44-6	49	43	1069	260	1421	6.70
45-1	49	50	944	188	1231	8.39
46-2	150	124	1047	211	1532	19.86
46-4	106	120	975	192	1393	17.81
47-1	101	141	1047	199	1488	17.86
47-6	103	78	1036	207	1424	13.64
48-1	143	164	970	199	1476	23.58
48-4	54	56	1051	224	1385	8.29

48-5	51	69	1082	288	1490	8.41
49-1	89	105	875	193	1262	16.78
49-2	37	37	992	206	1272	6.00
49-5	63	67	1188	280	1598	8.50
50-3	133	153	1216	265	1767	17.76
50-6	72	53	1069	259	1453	9.01
51-2	105	107	884	169	1265	18.46
51-4	138	138	991	188	1455	21.22
51-5	68	83	904	236	1291	12.47
51-6	58	62	921	228	1269	9.95
52-1	50	84	903	222	1259	11.28
52-2	96	90	1079	260	1525	13.05
53-1	83	105	913	196	1297	15.73
53-2	87	96	995	197	1375	14.34
53-4	124	130	1084	162	1500	18.68
53-5	124	136	1272	219	1751	16.15
54-1	206	184	1162	205	1757	25.43
54-2	117	128	1054	234	1533	17.52
54-6	102	105	1068	228	1503	14.88
55-2	54	61	916	222	1253	9.64
55-4	88	110	992	191	1381	15.55
55-6	63	61	1206	305	1635	7.90
56-1	129	126	1235	262	1752	15.80
56-5	132	143	1001	163	1439	21.40
57-2	67	65	1075	207	1414	9.82
57-5	76	102	1121	220	1519	12.50
57-6	70	70	1030	199	1369	10.81
58-1	75	47	864	205	1191	10.83
58-5	94	94	1232	315	1735	11.50
Q3	40	55	1817	413	2325	4.17
K1	74	85	1808	450	2417	6.81
J9	69	85	1753	402	2309	6.91
S1	70	89	1624	403	2186	7.56

AF4	85	95	1847	429	2456	7.62
AA3	101	102	2103	451	2757	7.66
V1	94	101	1917	511	2623	7.73
AD5	88	89	1715	399	2291	8.05
L6	103	115	1932	441	2591	8.80
AI1	101	119	1868	427	2515	9.17
T6	103	107	1707	467	2384	9.24
D9	106	110	1794	442	2452	9.24
AM1	125	108	1861	468	2562	9.55
S3	107	123	1845	417	2492	9.70
AM7	119	128	1951	440	2638	9.85
K7	118	124	1796	480	2518	10.12
R7	125	99	1729	377	2330	10.13
AA1	110	132	1843	421	2506	10.17
V7	115	107	1697	357	2276	10.28
B2	109	106	1533	354	2102	10.81
A4	80	93	1259	237	1669	10.97
D4	136	136	1898	445	2615	11.01
V2	118	127	1727	352	2324	11.17
L1	155	143	2124	400	2822	11.19
D8	112	142	1754	385	2393	11.25
AR8	121	138	1696	383	2338	11.77
R1	123	125	1617	344	2209	11.94
AB1	138	130	1750	353	2371	12.03
R9	96	129	1427	319	1971	12.15
W9	97	114	1350	268	1829	12.29
J4	143	157	1878	386	2564	12.48
Q9	164	161	1899	529	2753	12.60
J1	158	149	1896	359	2562	12.80
AG1	154	184	2035	436	2809	12.86
P1	129	168	1809	360	2466	12.87
AF6	173	143	1895	400	2611	12.94
S2	168	143	1847	409	2567	12.95

AI9	146	151	1801	344	2442	13.01
AN2	156	160	1912	335	2563	13.20
B3	135	148	1646	366	2295	13.20
K4	170	176	1997	359	2702	13.75
AM4	205	206	2251	467	3129	14.13
AN6	179	163	1742	439	2523	14.62
G9	141	157	1521	319	2138	15.07
AP2	197	180	1933	372	2682	15.21
AD1	180	190	1913	316	2599	15.43
AU7	188	206	1886	395	2675	16.01
R2	200	201	1848	455	2704	16.13
U8	182	164	1639	307	2292	16.45
C3	207	203	1812	476	2698	16.57
Q4	205	179	1625	439	2448	17.16
F7	180	220	1708	366	2474	17.74
AR6	193	212	1756	295	2456	18.13
A2	224	221	1904	292	2641	18.57
H2	214	248	1859	345	2666	19.17
F2	189	223	1681	272	2365	19.28
AJ1	229	207	1717	340	2493	19.36
I9	228	225	1773	356	2582	19.43
V9	215	233	1731	373	2552	19.45
R6	189	214	1593	284	2280	19.60
J3	214	241	1730	343	2528	20.00
D7	119	160	970	240	1489	20.93
D3	177	175	1114	248	1714	23.24
C9	243	230	1594	236	2303	23.24

Table 17: Summary of findings from Cvi-0/*420* F_2 QTL scan. QTL interactions tested for by Hayley-Knott regression, according to the full model y= $rQTL1^{420}$ + $rQTL2^{420}$ + $rQTL5^{420}$ + $rQTL1^{420}$: $rQTL2^{420}$ + $rQTL1^{420}$: $rQTL5^{420}$ + $rQTL2^{420}$: $rQTL5^{420}$, where : designates an interaction. Analysis of significance performed using a drop-one-term ANOVA table. QTL, or QTL interaction, is listed with chromosome, LOD score, percentage of variance explained and F-test p-value.

Chromosome	QTL	LOD	% Variance	p value (F)
1	$rQTL1^{420}$	44.02	44.44	2.00E-16
2	$rQTL2^{420}$	16.5	12.11	3.18E-11
5	$rQTL5^{420}$	19.72	15.01	6.33E-14
1:2	$rQTL1^{420}$: $rQTL2^{420}$	0.66	0.41	0.594
1:5	$rQTL1^{420}$: $rQTL5^{420}$	0.28	0.17	0.88
2:5	$rQTL2^{420}$: $rQTL5^{420}$	0.81	0.5	0.491

Table 18. Fluorescent seed count data for the *420* FTL interval in the Can-0/*420* F_2 population. Genetic distance of the *420* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015).

Individual	Green (G)	Red (R)	Double colour (D)	Non-colour (N)	Total (T)	cM
1-1	93	69	1061	319	1542	11.12
2-2	83	94	992	258	1427	13.29
7-6	81	92	1162	297	1632	11.23
8-2	98	143	1352	406	1999	12.89
12-2	100	95	1262	394	1851	11.16
18-1	115	95	1232	362	1804	12.41
18-3	158	148	897	207	1410	24.77
18-4	125	112	1023	278	1538	16.83
20-6	150	131	1226	333	1840	16.66
21-1	84	55	767	216	1122	13.27
22-4	110	130	1394	376	2010	12.75
23-1	81	67	865	241	1254	12.60
23-3	106	64	842	233	1245	14.74
37-2	71	62	830	263	1226	11.51
38-6	98	88	1084	289	1559	12.74
39-6	67	70	823	253	1213	12.02
40-6	79	65	957	253	1354	11.27
43-2	72	89	1024	321	1506	11.33
48-1	86	72	916	245	1319	12.80
48-6	96	97	841	207	1241	17.00
49-4	79	82	1243	400	1804	9.36
50-2	140	130	1171	315	1756	16.78
50-5	116	68	1285	390	1859	10.44
53-4	114	89	1032	304	1539	14.20
54-1	160	121	1066	282	1629	19.07
54-6	98	88	1134	337	1657	11.94
55-6	85	73	949	285	1392	12.08
60-2	79	68	891	224	1262	12.42

67-3	169	165	1330	342	2006	18.33
69-1	82	76	893	269	1320	12.79
84-1	101	98	1181	315	1695	12.52
84-6	122	145	1257	332	1856	15.60
85-1	123	135	1278	343	1879	14.83
85-2	119	99	1035	260	1513	15.63
85-3	93	84	1270	356	1803	10.35
85-4	156	133	1289	315	1893	16.65
85-6	145	151	1225	287	1808	17.99
86-2	84	73	1257	344	1758	9.37
86-4	100	114	1160	332	1706	13.45
87-1	141	152	1362	351	2006	15.86
87-2	110	90	1188	316	1704	12.52
87-3	151	159	1156	277	1743	19.73
88-1	88	89	1223	347	1747	10.70
88-4	103	124	1264	338	1829	13.29
88-5	98	92	1196	325	1711	11.80
88-6	75	83	1118	303	1579	10.56
89-2	91	83	1014	265	1453	12.79
89-3	113	131	1279	358	1881	13.94
89-6	129	110	1225	370	1834	14.01
90-1	105	78	974	268	1425	13.79
90-3	103	126	1183	287	1699	14.53
91-6	137	144	1152	292	1725	17.89
93-5	123	130	1287	360	1900	14.34
93-6	99	93	1215	349	1756	11.61
94-3	116	131	1056	245	1548	17.48
94-4	120	109	1072	262	1563	15.92
94-5	99	113	1087	295	1594	14.33
95-1	92	93	805	210	1200	16.83
95-2	93	112	1232	348	1785	12.23
95-4	133	118	1029	242	1522	18.14
96-2	81	80	1187	369	1717	9.86

96-4	85	97	1319	358	1859	10.32
96-6	159	156	1066	262	1643	21.48
97-1	115	130	1054	236	1535	17.49
97-4	143	140	1061	242	1586	19.80
97-6	150	155	947	187	1439	24.10
98-3	111	113	1035	291	1550	15.68
98-6	99	111	1132	297	1639	13.76
100-5	117	127	1037	281	1562	17.08
100-6	108	104	988	291	1491	15.41
101-2	139	142	977	250	1508	20.80
101-3	91	88	970	283	1432	13.40
101-5	87	100	1040	316	1543	12.96
102-3	92	85	932	251	1360	13.99
102-4	122	105	1157	291	1675	14.62
102-5	124	115	991	233	1463	17.95
102-6	95	82	1009	255	1441	13.15
103-4	81	93	1057	331	1562	11.84
103-5	130	124	963	224	1441	19.53
103-6	105	89	1029	260	1483	14.07
104-1	86	92	964	283	1425	13.39
104-3	78	70	958	279	1385	11.33
104-5	84	98	915	237	1334	14.73
105-2	99	125	920	232	1376	17.88
105-5	101	74	732	185	1092	17.57
105-6	64	92	827	221	1204	13.93
106-2	117	112	1002	287	1518	16.44
106-4	73	88	743	211	1115	15.67
107-2	107	121	993	247	1468	16.97
107-3	74	86	815	216	1191	14.48
108-2	80	89	853	256	1278	14.24
108-3	82	101	879	225	1287	15.41
108-5	66	67	952	242	1327	10.58
108-6	87	97	805	229	1218	16.46

109-1	105	122	863	204	1294	19.43
109-2	74	76	1031	266	1447	10.97
109-3	75	82	1060	332	1549	10.71
109-4	87	120	917	254	1378	16.36
110-2	101	100	757	209	1167	19.04
110-5	92	83	737	202	1114	17.19
110-6	83	75	779	218	1155	14.77
111-1	81	106	894	220	1301	15.59
111-2	96	106	997	281	1480	14.73
111-3	133	126	977	246	1482	19.35
111-4	118	115	1081	285	1599	15.82
111-5	80	76	911	222	1289	12.94
111-6	78	83	927	263	1351	12.73
112-1	70	71	1148	362	1651	8.94
113-2	120	123	1063	237	1543	17.23
113-3	149	143	1101	260	1653	19.58
114-1	95	80	893	242	1310	14.39
115-1	96	86	1119	318	1619	11.96
115-3	106	84	896	223	1309	15.76
115-4	123	104	1040	253	1520	16.26

Table 19: Summary of findings from Can-0/*420*F_2 QTL scan. QTL interactions tested for by Hayley-Knott regression, according to the full model y= CanQTL3 + CanQTL4 + CanQTL3:CanQTL4, where : designates an interaction. Analysis of significance performed using a drop-one-term ANOVA table. QTL, or QTL interaction, is listed with chromosome, LOD score, percentage of variance explained and F-test p-value.

Chromosome	QTL	LOD	% Variance	p value (F)
3	CanQTL3	8.98	19.12	7.86E-07
4	CanQTL4	13.6	32.03	8.10E-11
3:4	CanQTL3:CanQTL4	0.6	1.08	0.631

Table 20. Chromosome 3 marker genotype counts from hot and cold quartile Can-0/*420* F_2 individuals. The number of individuals with homozygosity (HOM) or heterozygosity (HET) for the chromosome 3 genotyping markers, in the F_2 population quartiles with highest (Hot) and lowest (Cold) recombination. The p value was calculated by constructing a 2 x 2 contingency table for homozygous and heterozygous genotype counts in the Hot and Cold quartiles and performing a two-tailed χ^2 test for significant differences. FDR correction for multiple testing was then applied (Ziolkowski *et al.* 2015).

Marker coordinates (bp)	Hot quartile HET	Hot quartile HOM	Cold quartile HET	Cold quartile HOM	FDR p value
1031000	22	1	17	0	0.383924
5419000	18	0	11	1	0.283846667
9194000	14	13	26	0	0.000368
10695000	14	12	27	1	0.00102
11649000	12	11	18	3	0.027264
12356000	13	12	21	4	0.030586
16008000	13	14	25	3	0.002573333
19165000	15	12	19	9	0.397538286

Table 21. Fluorescent seed count data for the *5.11* CTL interval in the Cvi-0/*5.11* F_2 population. Genetic distance of the *5.11* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015).

Individual	Green(G)	Red (R)	Double colour (D)	Non-colour (N)	Total (T)	cM
1-1	187	197	950	215	1549	28.99
1-2	193	196	1084	232	1705	26.26
1-3	203	241	1099	196	1739	30.05
1-4	173	189	1068	248	1678	24.60
1-5	171	188	970	179	1508	27.62
2-1	186	184	1128	245	1743	24.14
2-4	182	189	1052	204	1627	26.25
3-2	140	158	949	202	1449	23.27
3-3	175	162	1016	201	1554	24.75
3-4	163	158	976	224	1521	23.98
3-5	181	201	1118	249	1749	24.95
3-6	186	184	1010	198	1578	27.13
4-1	188	168	1026	204	1586	25.77
4-2	189	212	1064	214	1679	27.73
4-4	152	137	898	181	1368	24.01
5-2	194	189	1044	202	1629	27.21
5-3	193	198	1042	223	1656	27.35
5-4	211	192	1006	187	1596	29.64
5-5	137	150	886	187	1360	23.98
5-6	179	172	996	203	1550	26.03
6-1	181	200	936	192	1509	29.64
6-2	195	184	1037	186	1602	27.42
6-3	170	184	928	185	1467	28.07
6-4	168	178	885	166	1397	28.96
6-5	180	195	1061	254	1690	25.42
6-6	190	174	990	188	1542	27.34
7-1	184	177	960	169	1490	28.21
7-2	165	160	868	174	1367	27.58

7-3	190	192	1061	242	1685	26.07
7-4	181	181	910	183	1455	29.12
7-6	190	207	1086	222	1705	26.90
8-1	158	162	962	238	1520	23.91
8-2	157	167	968	216	1508	24.48
8-5	167	154	1048	213	1582	22.92
9-2	187	200	970	212	1569	28.82
9-4	199	195	1074	227	1695	26.85
9-5	168	172	956	182	1478	26.52
9-6	153	171	970	209	1503	24.58
10-1	158	202	1028	215	1603	25.78
10-2	203	239	1026	224	1692	30.90
10-3	189	195	1150	243	1777	24.65
10-5	240	212	1177	241	1870	28.13
10-6	172	189	989	237	1587	26.17
11-1	180	188	1096	234	1698	24.73
11-3	147	164	887	178	1376	25.98
11-4	201	218	943	152	1514	33.18
11-5	236	225	1032	204	1697	32.42
12-2	254	237	1025	184	1700	35.01
12-4	183	191	1017	195	1586	27.31
12-5	176	195	1108	215	1694	25.03
12-6	180	217	1057	227	1681	27.36
13-3	212	191	1154	234	1791	25.84
13-4	188	202	1095	232	1717	26.13
13-5	187	215	1060	222	1684	27.71
13-6	199	200	1128	246	1773	25.84
14-2	228	227	1162	219	1836	28.98
14-3	174	148	849	183	1354	27.59
14-5	176	230	1040	204	1650	28.73
15-2	185	166	875	191	1417	28.97
15-3	134	127	834	194	1289	22.86
15-4	136	127	725	153	1141	26.58

15-5	194	177	1066	227	1664	25.56
15-6	127	137	944	200	1408	20.94
16-2	222	235	1072	210	1739	31.12
16-4	190	181	1000	203	1574	27.30
16-5	202	205	1097	230	1734	27.16
16-6	216	186	972	215	1589	29.71
17-3	142	183	965	209	1499	24.74
17-4	196	167	945	182	1490	28.39
18-2	205	235	1168	256	1864	27.34
18-3	180	174	1069	249	1672	24.07
18-5	215	195	1010	175	1691	30.29
18-6	184	191	1153	225	1753	24.36
19-2	219	235	1075	227	1691	30.51
19-4	194	164	1007	206	1571	26.23
19-6	225	205	1087	211	1691	29.13
20-2	168	157	898	205	1428	26.19
20-4	130	129	808	163	1230	23.92
21-1	202	200	1042	195	1691	28.62
21-2	191	170	1017	201	1579	26.33
21-4	156	179	1029	204	1568	24.32
21-5	190	185	1068	217	1660	25.96
21-6	195	168	995	220	1578	26.52
22-1	203	197	1093	198	1691	27.41
22-3	187	203	1154	225	1769	25.23
22-4	194	178	964	185	1691	28.53
22-5	168	181	937	184	1691	27.53
23-2	202	186	1035	210	1691	27.56
23-3	190	178	1047	196	1611	26.30
24-1	245	197	1129	236	1691	28.53
24-2	223	242	1170	242	1691	28.97
24-3	197	222	973	201	1691	31.16
25-1	226	212	1132	216	1691	28.62
25-2	177	191	1140	231	1739	24.05

25-3	184	206	1126	236	1752	25.52
25-4	198	200	1037	193	1691	28.51
25-6	157	199	922	206	1691	27.87
26-1	220	219	1043	217	1691	30.49
26-2	224	230	1147	209	1691	29.41
26-4	170	187	1137	232	1726	23.43
26-5	192	233	1039	207	1691	29.91
27-2	182	153	896	177	1691	27.60
27-3	169	195	1010	207	1581	26.55
27-5	229	221	959	174	1691	34.31
27-6	171	149	1078	231	1629	22.08
28-6	177	190	1046	213	1626	25.93
29-1	184	198	964	211	1691	28.63
29-2	186	212	1010	190	1691	29.16
29-4	199	176	1051	234	1660	25.96
29-5	151	147	1048	224	1570	21.24
29-6	192	217	1087	208	1691	27.89

Table 22. Fluorescent seed count data for the *420* FTL interval in the rQTL2^{420} BC$_1$F$_2$ fine mapping population (referred to as BP). Population was generated from self-fertilisation of the Cvi*420*F$_3$ (F$_2$ 10-6)/Col 1-3 line (BC$_1$F$_1$ line, see Figures 26 and 27). Genetic distance of the *420* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015).

Individual	Green (G)	Red (R)	Double colour (D)	Non-colour (N)	Total (T)	cM
1-2	159	170	898	183	1410	26.97
1-5	144	134	999	264	1541	20.05
1-6	173	164	887	188	1412	27.70
2-2	151	154	991	257	1553	22.08
2-3	131	125	1077	251	1584	17.73
2-5	104	119	925	246	1394	17.53
2-6	125	107	706	153	1091	24.19
4-1	162	173	1045	253	1633	23.21
4-3	104	83	677	166	1030	20.19
4-4	111	145	813	184	1253	23.10
5-3	126	116	874	233	1349	19.92
5-4	140	126	978	229	1473	20.07
5-5	123	145	895	219	1382	21.76
6-1	147	135	1012	246	1540	20.39
6-2	169	143	907	216	1435	24.82
8-5	118	136	929	215	1398	20.21
9-1	168	153	960	229	1510	24.18
9-2	178	182	952	207	1519	27.47
9-3	130	132	973	224	1459	19.95
9-5	147	135	983	216	1481	21.31
12-3	133	154	743	164	1194	27.94
12-6	208	184	999	183	1574	29.15
13-1	148	152	1030	255	1585	21.17
14-2	165	159	953	231	1508	24.48
14-4	145	160	1183	287	1775	18.99
15-4	141	152	980	248	1521	21.60
15-5	115	109	805	194	1223	20.40
16-1	121	136	852	188	1297	22.30
16-3	134	132	804	194	1264	23.90
16-5	130	140	1103	313	1686	17.56
18-6	148	167	1215	281	1811	19.25
19-1	154	171	924	216	1465	25.41
19-2	160	148	793	147	1248	28.84
19-3	194	169	919	182	1464	29.00
19-5	182	178	1032	226	1618	25.50

20-1	194	192	897	190	1473	31.01
20-2	155	168	964	184	1471	25.11
21-1	148	137	978	244	1507	21.15
21-2	119	121	711	164	1115	24.53
21-3	187	193	1019	233	1632	26.90
21-4	155	150	869	185	1359	25.76
21-6	147	169	1060	257	1633	21.71
22-4	150	141	1045	280	1616	20.01
22-5	123	112	821	217	1273	20.58
23-4	143	144	979	213	1479	21.78
24-2	178	191	1028	221	1618	26.25
25-3	113	145	779	182	1219	24.06
26-2	136	140	866	187	1329	23.54
26-3	137	115	771	165	1188	24.12
27-5	141	149	909	227	1426	22.98
28-2	136	158	906	217	1417	23.51
29-3	144	123	765	169	1201	25.48
30-5	142	124	981	265	1512	19.49
32-4	124	140	838	203	1305	22.84
32-5	111	110	865	205	1291	18.91
33-5	146	153	831	188	1318	26.09
33-6	141	155	902	219	1417	23.70
34-2	141	140	949	204	1434	22.02
34-4	134	132	692	145	1103	28.05
34-5	161	165	839	190	1355	27.97
34-6	145	154	957	226	1482	22.77
35-1	132	133	984	230	1479	19.90
35-3	114	95	779	207	1195	19.36
36-2	138	141	985	216	1480	21.07
36-5	148	146	902	220	1416	23.53
36-6	144	154	955	203	1456	23.15
37-4	125	148	910	213	1396	21.97
38-5	139	127	825	184	1275	23.66
38-6	161	158	863	173	1355	27.26
39-1	136	144	824	179	1283	24.93
40-1	131	116	959	242	1448	18.83
40-4	158	139	1000	258	1555	21.39
40-5	149	148	956	253	1506	22.18
41-4	150	135	772	179	1236	26.59
41-5	117	103	729	166	1115	22.19
41-6	132	115	905	228	1380	19.87
42-4	131	109	861	218	1319	20.24
42-5	118	101	947	218	1384	17.32

43-5	125	124	867	238	1354	20.49
44-1	148	126	867	191	1332	23.28
44-5	105	114	689	166	1074	23.05
45-5	173	150	788	167	1278	29.68
45-6	120	124	845	191	1280	21.34
46-2	151	138	711	140	1140	29.79
46-5	146	153	698	141	1138	31.11
47-3	185	158	771	151	1265	32.35
47-5	157	177	924	188	1446	26.65
48-5	114	136	813	174	1237	22.81
48-6	132	158	791	162	1243	26.97
49-1	183	170	910	160	1423	29.02
49-2	159	188	680	117	1144	37.28
50-2	128	166	896	206	1396	23.92
50-3	122	139	647	128	1036	29.56
51-1	190	190	962	207	1549	28.63
51-2	188	190	857	169	1404	32.06
51-3	137	139	906	214	1396	22.24
52-1	111	135	843	204	1293	21.29
52-4	140	163	871	194	1368	25.37
52-5	133	113	738	179	1163	24.04
53-2	160	152	780	140	1232	29.75
53-4	200	162	901	169	1432	29.69
54-2	151	147	749	129	1176	29.77
55-6	149	140	817	207	1313	25.18
56-4	146	148	875	183	1352	24.83
57-6	124	147	801	169	1241	24.95
58-1	112	113	858	221	1304	19.07
58-2	171	148	757	166	1242	30.26
59-2	108	124	677	139	1048	25.35
59-5	127	132	709	142	1110	26.97
60-1	167	163	993	234	1557	24.10
60-3	153	142	874	214	1383	24.28
60-4	125	142	725	155	1147	26.89
60-5	141	150	788	173	1252	26.85
61-2	135	119	788	191	1233	23.32
61-4	127	125	957	231	1440	19.38
62-1	139	119	817	169	1244	23.50
63-5	121	144	868	186	1319	22.66
64-1	93	91	694	179	1057	19.26
64-2	159	179	1019	246	1603	23.95
64-4	173	171	1057	237	1638	23.84
65-1	194	208	871	139	1412	34.38

65-2	111	131	732	159	1133	24.32
65-4	171	158	841	159	1329	28.94
65-6	176	178	823	159	1336	31.44
66-1	174	216	902	184	1476	31.33
66-2	151	147	834	207	1339	25.51
66-5	149	152	903	201	1405	24.40
66-6	144	145	672	136	1097	31.22
67-1	119	157	763	155	1194	26.67
67-3	169	164	867	180	1380	28.07
67-5	163	186	926	198	1473	27.46
68-1	122	114	759	189	1184	22.45
68-2	134	129	888	228	1379	21.35
69-2	148	147	880	183	1358	24.80
70-2	141	161	857	187	1346	25.75
70-6	181	164	822	181	1348	30.13
71-4	143	130	901	194	1368	22.48
71-6	160	151	714	121	1146	32.38
72-1	135	157	800	178	1270	26.50
72-2	136	141	948	226	1451	21.37
72-5	144	136	710	169	1159	28.11
73-6	123	126	725	160	1134	25.11
74-1	162	175	864	182	1383	28.40
74-5	139	127	977	234	1477	20.01
75-6	125	131	928	205	1389	20.54
77-4	135	124	790	177	1226	24.01
78-1	166	152	911	205	1434	25.40
79-6	148	112	877	205	1342	21.74
80-1	180	181	838	181	1380	30.95
80-6	124	135	849	219	1327	21.92
81-1	133	140	860	212	1345	22.93
81-2	163	164	814	159	1300	29.51
81-4	114	114	926	212	1366	18.38
81-6	140	147	1006	258	1551	20.63
83-1	122	106	904	219	1351	18.61
83-2	149	126	673	138	1086	29.75
83-6	134	126	769	149	1178	25.26
84-4	100	130	747	165	1142	22.72
85-1	110	121	808	202	1241	20.77
85-2	156	140	814	193	1303	26.13
87-1	110	108	653	150	1021	24.31
87-2	116	123	714	139	1092	25.02
87-5	134	118	666	159	1077	27.06
88-2	120	114	887	201	1322	19.63

Table 23. Fluorescent seed count data for the *420* FTL interval in the rQTL2^{420} BP progeny BC_1F_3 mapping populations. Populations were generated from self-fertilisation of lines BP 6-2, BP 14-2 and BP 66-6 (BC_1F_2 lines, see Figures 26 and 27). Genetic distance of the *420* interval is calculated as cM = 100 x $(1-(1-2(G+R)/T)^{1/2})$. G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015).

Parent	Individual	Green (G)	Red (R)	Double-colour (D)	Non-colour (N)	Total (T)	cM
BP 6-2	2-6	167	164	970	221	1522	24.83
BP 6-2	3-3	144	164	930	196	1434	24.47
BP 6-2	4-5	159	148	908	202	1417	24.72
BP 6-2	5-1	161	140	914	197	1412	24.26
BP 6-2	5-4	157	164	953	204	1478	24.79
BP 6-2	6-3	151	185	1039	236	1611	23.65
BP 6-2	8-1	137	139	892	199	1367	22.79
BP 6-2	8-3	161	150	784	177	1272	28.52
BP 6-2	10-2	147	147	893	203	1390	24.04
BP 6-2	12-1	137	151	812	201	1301	25.35
BP 6-2	12-5	162	174	921	196	1453	26.69
BP 6-2	13-4	135	143	848	198	1324	23.84
BP 6-2	15-1	135	148	1026	217	1526	20.68
BP 6-2	16-6	135	132	907	242	1416	21.08
BP 6-2	17-2	172	172	956	168	1468	27.11
BP 6-2	17-4	150	158	922	229	1459	23.99
BP 6-2	18-1	119	134	942	226	1421	19.76
BP 6-2	20-6	176	194	1009	193	1572	27.25
BP 6-2	21-2	155	136	946	223	1460	22.45
BP 6-2	21-4	180	147	801	170	1298	29.56
BP 6-2	23-6	143	139	872	177	1331	24.09
BP 6-2	26-6	140	161	937	223	1461	23.32
BP 6-2	27-1	183	187	867	203	1440	30.28
BP 6-2	27-2	154	161	1016	246	1577	22.51
BP 6-2	27-5	94	111	897	240	1342	16.66
BP 14-2	3-2	142	126	855	194	1317	22.99
BP 14-2	3-5	160	157	853	201	1371	26.68
BP 14-2	4-1	153	177	939	202	1471	25.75
BP 14-2	5-6	174	182	927	216	1499	27.54
BP 14-2	8-2	183	185	1110	253	1731	24.18
BP 14-2	8-6	138	155	1001	233	1527	21.50
BP 14-2	12-5	170	166	979	180	1495	25.80
BP 14-2	13-4	174	174	898	172	1418	28.64
BP 14-2	15-4	196	164	978	220	1558	26.66
BP 14-2	16-4	160	155	915	208	1438	25.04
BP 14-2	18-2	164	145	924	209	1442	24.41

BP 14-2	18-4	172	173	952	202	1499	26.54
BP 14-2	20-6	176	181	1015	224	1596	25.66
BP 14-2	21-4	127	122	773	178	1200	23.51
BP 14-2	22-1	150	150	875	189	1364	25.16
BP 14-2	22-4	124	104	768	190	1186	21.55
BP 14-2	23-3	160	129	979	251	1519	21.29
BP 14-2	23-5	168	167	1077	256	1668	22.65
BP 66-6	1-4	140	152	736	163	1191	28.61
BP 66-6	3-2	165	152	782	164	1263	29.43
BP 66-6	3-3	218	188	989	188	1583	30.21
BP 66-6	4-6	209	190	1055	226	1680	27.54
BP 66-6	5-5	197	200	988	186	1571	29.67
BP 66-6	6-6	168	157	983	201	1509	24.55
BP 66-6	7-1	183	165	972	222	1542	25.93
BP 66-6	11-2	151	157	933	201	1442	24.32
BP 66-6	12-1	161	160	977	232	1530	23.82
BP 66-6	14-1	161	187	1068	244	1660	23.79
BP 66-6	17-3	164	163	853	199	1379	27.49
BP 66-6	19-1	172	185	959	230	1546	26.64
BP 66-6	21-3	189	210	955	184	1538	30.64
BP 66-6	23-1	136	138	778	172	1224	25.68
BP 66-6	23-4	182	183	976	205	1546	27.35
BP 66-6	26-1	154	168	913	208	1443	25.59
BP 66-6	26-4	179	225	990	190	1584	30.01
BP 66-6	26-5	156	129	853	201	1339	24.22
BP 66-6	28-1	167	179	1008	226	1580	25.03
BP 66-6	29-4	141	141	735	142	1159	28.35
BP 66-6	29-5	175	168	866	184	1393	28.76
BP 66-6	30-3	196	219	993	221	1629	29.97
BP 66-6	30-4	157	182	920	204	1463	26.75
BP 66-6	33-3	175	170	912	217	1474	27.07
BP 66-6	34-1	174	184	1035	243	1636	25.01
BP 66-6	36-3	202	170	956	215	1543	28.04
BP 66-6	38-1	181	169	1117	234	1701	23.29

Table 24. Fluorescent seed count data for the *420* FTL interval in the rQTL2^{420} BP progeny second-generation BC_1F_4 mapping populations. Populations were generated from self-fertilisation of lines BP 14-2 13-4 and BP 14-2 15-4 (BC_1F_3 lines, see Figures 26 and 27). Genetic distance of the *420* interval is calculated as cM = 100 x $(1-(1-2(G+R)/T)^{1/2})$. G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015).

Parent	Individual	Green (G)	Red (R)	Double-colour (D)	Non-colour (N)	Total (T)	cM
BP 14-2 13-4	1-2	194	208	989	190	1581	29.90
BP 14-2 13-4	1-4	151	158	909	174	1392	25.43
BP 14-2 13-4	3-4	193	191	1036	214	1634	27.20
BP 14-2 13-4	4-2	194	199	1031	192	1616	28.33
BP 14-2 13-4	5-1	157	189	1067	224	1637	24.02
BP 14-2 13-4	5-3	202	187	1078	245	1712	26.14
BP 14-2 13-4	5-6	174	194	1027	222	1617	26.19
BP 14-2 13-4	6-1	244	233	1162	215	1854	30.33
BP 14-2 13-4	6-2	179	181	1105	242	1707	23.96
BP 14-2 13-4	6-5	174	169	996	218	1557	25.21
BP 14-2 13-4	7-1	178	148	999	229	1554	23.81
BP 14-2 13-4	8-1	215	201	1062	210	1688	28.79
BP 14-2 13-4	8-4	182	194	978	217	1571	27.80
BP 14-2 13-4	8-5	177	194	995	225	1591	26.95
BP 14-2 13-4	9-1	196	205	985	192	1578	29.87
BP 14-2 13-4	9-5	218	218	1137	238	1811	27.99
BP 14-2 13-4	10-5	184	197	1045	221	1647	26.70
BP 14-2 13-4	11-2	232	201	1124	211	1768	28.57
BP 14-2 13-4	11-5	201	201	1109	222	1733	26.78
BP 14-2 13-4	12-1	237	215	1131	244	1827	28.92
BP 14-2 13-4	13-2	199	208	1132	237	1776	26.40
BP 14-2 13-4	14-2	180	185	1018	231	1614	25.99
BP 14-2 13-4	14-3	192	213	1068	244	1717	27.32
BP 14-2 13-4	14-6	201	195	1037	236	1669	27.51
BP 14-2 13-4	15-1	164	163	922	206	1455	25.80
BP 14-2 13-4	15-4	126	142	906	196	1370	21.98
BP 14-2 13-4	15-5	192	180	1040	236	1648	25.94
BP 14-2 13-4	16-1	185	172	1064	236	1657	24.56
BP 14-2 13-4	16-6	187	189	1012	235	1623	26.74
BP 14-2 13-4	17-3	187	182	945	228	1542	27.79
BP 14-2 13-4	17-4	199	174	1054	226	1653	25.93
BP 14-2 13-4	17-5	221	171	1098	243	1733	26.00
BP 14-2 13-4	18-5	178	191	1081	268	1718	24.47
BP 14-2 13-4	19-4	207	193	1140	243	1783	25.75
BP 14-2 13-4	20-1	186	184	1094	248	1712	24.65

BP 14-2 13-4	20-3	201	235	1113	251	1800	28.20
BP 14-2 13-4	21-3	214	207	1111	216	1748	28.01
BP 14-2 13-4	22-5	176	171	1056	248	1651	23.87
BP 14-2 13-4	24-1	170	189	1100	255	1714	23.77
BP 14-2 13-4	24-3	179	162	1054	241	1636	23.64
BP 14-2 13-4	25-3	170	171	1139	278	1758	21.77
BP 14-2 13-4	25-4	175	189	1221	283	1868	21.88
BP 14-2 13-4	26-4	165	159	1018	233	1575	23.28
BP 14-2 13-4	26-5	179	176	972	226	1553	26.32
BP 14-2 13-4	26-6	187	197	1052	246	1682	26.28
BP 14-2 13-4	27-2	176	168	1035	227	1606	24.40
BP 14-2 13-4	27-4	157	155	1057	273	1642	21.26
BP 14-2 13-4	27-5	167	158	996	234	1555	23.71
BP 14-2 13-4	28-1	188	169	1016	219	1592	25.74
BP 14-2 13-4	28-2	169	176	1111	262	1718	22.65
BP 14-2 13-4	28-5	167	193	1084	268	1712	23.88
BP 14-2 13-4	29-3	169	217	1081	227	1694	26.23
BP 14-2 13-4	29-6	188	198	1107	236	1729	25.60
BP 14-2 13-4	30-2	157	192	1097	267	1713	23.02
BP 14-2 13-4	30-4	197	195	1067	199	1658	27.40
BP 14-2 13-4	30-5	162	147	1044	263	1616	21.41
BP 14-2 13-4	30-6	208	216	1131	234	1789	27.47
BP 14-2 15-4	4-1	160	151	1066	236	1613	21.62
BP 14-2 15-4	4-5	158	151	1045	252	1606	21.57
BP 14-2 15-4	8-1	170	195	1033	247	1645	25.42
BP 14-2 15-4	8-4	194	184	1069	252	1699	25.50
BP 14-2 15-4	8-5	156	167	1003	237	1563	23.40
BP 14-2 15-4	11-1	180	146	1089	245	1660	22.08
BP 14-2 15-4	11-6	181	168	1025	201	1575	25.38
BP 14-2 15-4	13-1	178	187	1124	234	1723	24.08
BP 14-2 15-4	14-1	188	168	1047	211	1614	25.24
BP 14-2 15-4	14-3	187	208	1174	256	1825	24.69

Table 25. Logarithm of odds (LOD) scores for association between rQTL2^{420} peak genotyping markers on chromosome 2 and *420* recombination frequency in the Cvi-0/*420* fine mapping population. Genotyping markers are indicated by a number representing their location on chromosome 2 (chromosome-kilobase) e.g. marker 2-12121 is found on chromosome 2 at position 12121783bp.

Marker	LOD
2-12121	5.57
2-12399	4.74
2-12618	4.45
2-12768	4.56
2-13157	4.38
2-13188	4.38
2-13195	4.07
2-13235	3.86
2-13427	4.2
2-13521	4.06

Table 26. Fluorescent seed count data for the *420* FTL interval in Col-0/*420* F_1 lines, rQTL2^{420}NILs and rQTL5^{420}NILs. Genetic distance of the 420 interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015). The aggregate number of recombinant (G+R) and non-recombinant (D+N) seeds for Col-0/*420* F_1 replicates and rQTL2^{420}NIL replicates, and Col-0/*420* F_1 replicates and rQTL5^{420}NIL replicates, were used to construct 2x2 contingency tables and perform two-tailed χ^2 tests to test for significant differences.

Line	Replicate	Green (G)	Red (R)	Double-colour (D)	Non-colour (N)	Total (T)	cM
rQTL2(420)NIL	1	154	168	934	214	1470	25.04
rQTL2(420)NIL	2	189	189	1028	225	1631	26.76
rQTL2(420)NIL	3	162	171	941	189	1463	26.19
rQTL2(420)NIL	4	211	214	1096	235	1756	28.17
rQTL2(420)NIL	5	197	173	1035	214	1619	26.32
rQTL2(420)NIL	6	196	198	1018	240	1652	27.68
rQTL2(420)NIL	7	147	136	796	185	1264	25.69
rQTL2(420)NIL	8	127	114	694	151	1086	25.42
rQTL2(420)NIL	9	145	161	859	187	1352	26.02
rQTL2(420)NIL	10	141	157	891	178	1367	24.90
rQTL5(420)NIL	1	100	105	1198	333	1736	12.60
rQTL5(420)NIL	2	96	92	1075	301	1564	12.85
rQTL5(420)NIL	3	107	81	1106	323	1617	12.39
rQTL5(420)NIL	4	108	94	1161	300	1663	12.99
rQTL5(420)NIL	5	119	90	1141	326	1676	13.36
rQTL5(420)NIL	6	112	101	1166	304	1683	13.58
rQTL5(420)NIL	7	91	117	1113	314	1635	13.65
rQTL5(420)NIL	8	122	95	1223	323	1763	13.18
rQTL5(420)NIL	9	91	96	1092	310	1589	12.56
rQTL5(420)NIL	10	99	108	1070	299	1576	14.13
Col-0/420 F1	1	140	128	1081	275	1624	18.15
Col-0/420 F1	2	124	133	1001	276	1534	18.46
Col-0/420 F1	3	134	99	942	239	1414	18.12
Col-0/420 F1	4	117	133	903	235	1388	20.01
Col-0/420 F1	5	120	116	872	207	1315	19.93
Col-0/420 F1	6	126	136	957	219	1438	20.28
Col-0/420 F1	7	139	150	1035	247	1571	20.50
Col-0/420 F1	8	142	136	987	252	1517	20.41
Col-0/420 F1	9	135	133	1003	236	1507	19.73
Col-0/420 F1	10	110	120	899	184	1313	19.40

rQTL2^{420}NIL $\chi^2(1)$ =119.1, p=2 x 10^{-27} , rQTL5^{420}NIL $\chi^2(1)$ =167.18, p=6.11 x 10^{-38}

Table 27. Fluorescent seed count data for the *5.11* FTL interval in rQTL2^{420}NIL/*5.11* replicates and rQTL5^{420}NIL/*5.11* replicates. For comparable Col-0/*5.11* F_1 line data see Table 6. Genetic distance of the *5.11* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015). The aggregate number of recombinant (G+R) and non-recombinant (D+N) seeds for Col-0/*5.11* F_1 replicates and rQTL2^{420}NIL/*5.11* replicates, and Col-0/*5.11* F_1 replicates and rQTL5^{420}NIL/*5.11* replicates, were used to construct 2x2 contingency tables and perform two-tailed χ^2 tests to test for significant differences.

Line	Replicate	Green (G)	Red (R)	Double-colour (D)	Non-colour (N)	Total (T)	cM
rQTL2(420)NIL/*5.11*	1	95	114	799	185	1193	19.40
rQTL2(420)NIL/*5.11*	2	105	78	795	204	1182	16.91
rQTL2(420)NIL/*5.11*	3	126	100	845	217	1288	19.44
rQTL2(420)NIL/*5.11*	4	118	111	899	197	1325	19.11
rQTL2(420)NIL/*5.11*	5	105	112	881	223	1321	18.06
rQTL2(420)NIL/*5.11*	6	88	94	691	149	1022	19.76
rQTL2(420)NIL/*5.11*	7	98	96	884	218	1296	16.30
rQTL2(420)NIL/*5.11*	8	121	98	817	188	1224	19.87
rQTL2(420)NIL/*5.11*	9	106	103	848	204	1261	18.24
rQTL5(420)NIL/*5.11*	1	93	102	756	183	1134	19.00
rQTL5(420)NIL/*5.11*	2	97	96	804	210	1207	17.53
rQTL5(420)NIL/*5.11*	3	90	89	793	205	1177	16.58
rQTL5(420)NIL/*5.11*	4	111	113	807	219	1250	19.90
rQTL5(420)NIL/*5.11*	5	71	86	712	166	1035	16.54
rQTL5(420)NIL/*5.11*	6	111	106	857	222	1296	18.44
rQTL5(420)NIL/*5.11*	7	67	90	716	205	1078	15.81
rQTL5(420)NIL/*5.11*	8	81	79	703	214	1077	16.16

rQTL2^{420}NIL/*5.11* $\chi^2(1) = 4.038$, $p = 0.089$, rQTL5^{420}NIL/*5.11* $\chi^2(1) = 12.62$, $p = 7.63 \times 10^{-4}$

Table 28. Candidate loci identified for rQTL2[420]. Position of the locus on chromosome 2 is indicated in base pairs, and the predicted gene function based on locus information from TAIR (Berardini *et al*. 2015) and Araport11 (Krishnakumar *et al*. 2015) is given for each locus. Polymorphisms between Col-0 and Cvi-0 accessions that are not shared between Cvi-0 and Can-0 accessions (Salk Genome Express Browser, Alonso-Blanco *et al*. 2016) are given for each locus. T-DNA insertion lines used for locus analysis are listed with insertion position (Salk T-DNA Express tool (The Salk Institute Genomic Analysis Laboratory)).

Position (bp)	Locus	Gene	Function	Col/Cvi polymorphism	T-DNA line	T-DNA line insertion position (bp)
13194185-13200128	*AT2G31010*	Protein kinase superfamily protein	Phosphorylation	Non-synonymous SNP in exon 3; SNPs in introns 3, 6 and 8	Salk 109522	13194245-13194581; Promoter
13220478-13222609	*AT2G31070*	tcp10: TCP family protein	Cell differentiation; leaf development	One non-synonymous SNP and one synonymous SNP in exon 1	Salk 050436	13221810-13222045; Exon 1
13263579-13266560	*AT2G31130*	Unknown protein	Hypothetical protein	Non-synonymous SNP in exon 2	Salk 064633	13264169-13264483; Exon 3
13296143-13298277	*AT2G31210*	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein	Transcription regulation; Anther development	Non-synonymous SNP in exon 1	GK-345C06	13296655-13297380; Exon 1
13329009-13331830	*AT2G31270*	CDT1A: Encodes a cyclin-dependent protein kinase	DNA replication; chromosome organisation; gametophyte development	One non-synonymous SNP and one synonymous SNP in exon 4; Three SNPs in promoter	Salk 014797	13328441-13328831; Promoter
13353821-13359728	*AT2G31320*	PARP1: Encodes a poly(ADP-ribose) polymerase	DNA ligation involved in DNA repair	Three SNPs in promoter; Synonymous SNP exon 1; Two non-synonymous SNPs exon 10; Synonymous SNP exon 17	GK-380E06 Salk 145153	13355957-13356361; Exon 11 13359817-13359974; Promoter
13379033-13381523	*AT2G31370*	POSF21: Basic-leucine zipper (bZIP) transcription factor family protein	Transcription regulation	Two non-synonymous SNPs exon 1; two synonymous SNPs exon 4	Salk 024459	13379318-13379574; Exon 1
13392022-13392851	*AT2G31410*	Unknown protein	Coiled-coil protein	Non-synonymous SNP exon 1	Salk 059528	13392447-13392714; Exon 1
13401132-13404177	*AT2G31450*	ATNTH1: Helix-hairpin-helix DNA-binding motif	Base-excision repair	SNP intron 6	Salk 054181	13401921-13402042; Exon 8

13416646-13421170	*AT2G31510*	ARI7: ARIADNE 7, RING/U-box superfamily protein	Protein ubiquitination; nucleic acid binding	Two non-synonymous SNPs and one synonymous SNP intron 1; non-synonymous SNP intron 14	Salk 082541 Salk 027620	13419074-13419426; Exon 7 13416879-13417350; Intron 14/Exon 15

Table 29. Fluorescent seed count data for the *420* FTL interval in Col-0/*420* F_1 and rQTL2^{420} candidate T-DNA/*420* F_1 lines. Genetic distance of the *420* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015). The aggregate number of recombinant (G+R) and non-recombinant (D+N) seeds for the replicates of each T-DNA/*420* F_1 line were compared to the Col-0/*420* F_1 replicates and used to construct 2x2 contingency tables and perform two-tailed χ^2 tests to test for significant differences. P-values for each T-DNA/*420* F_1 are given in the final column.

Line	Replicate	Green (G)	Red(R)	Double -colour (D)	Non-colour (N)	Total (T)	cM	p-value
Salk 082541/*420* F_1	1	110	118	958	265	1451	17.19	
Salk 082541/*420* F_1	2	100	78	986	289	1453	13.11	
Salk 082541/*420* F_1	3	68	92	898	240	1298	13.20	
Salk 082541/*420* F_1	4	122	88	1085	284	1579	14.33	
Salk 082541/*420* F_1	5	120	88	1081	306	1595	14.02	
Salk 082541/*420* F_1	6	91	92	875	239	1297	15.28	
Salk 082541/*420* F_1	7	104	101	1014	261	1480	14.97	
Salk 082541/*420* F_1	8	121	115	1094	292	1622	15.80	
Salk 082541/*420* F_1	9	127	114	1092	262	1595	16.47	
Salk 082541/*420* F_1	10	115	130	1113	310	1668	15.96	
Salk 082541/*420* F_1	11	95	109	1034	261	1499	14.69	
Salk 082541/*420* F_1	12	122	147	1132	293	1694	17.39	
Salk 082541/*420* F_1	13	84	93	759	205	1141	16.95	
Salk 082541/*420* F_1	14	118	118	1078	257	1571	16.36	
Salk 082541/*420* F_1	15	122	122	1076	274	1594	16.70	
Salk 082541/*420* F_1	16	131	129	1218	283	1761	16.05	
Salk 082541/*420* F_1	17	114	98	1118	301	1631	13.97	6.35E-11
Salk 027620/*420* F_1	1	142	157	1204	295	1798	18.30	
Salk 027620/*420* F_1	2	142	163	1262	329	1896	17.64	
Salk 027620/*420* F_1	3	117	125	971	263	1476	18.02	
Salk 027620/*420* F_1	4	158	149	1214	308	1829	18.50	
Salk 027620/*420* F_1	5	142	148	1120	250	1660	19.34	
Salk 027620/*420* F_1	6	134	146	1086	256	1622	19.08	
Salk 027620/*420* F_1	7	141	126	1022	254	1543	19.13	
Salk 027620/*420* F_1	8	115	129	963	263	1470	18.27	
Salk 027620/*420* F_1	9	142	125	991	253	1511	19.59	
Salk 027620/*420* F_1	10	147	132	1147	276	1702	18.02	
Salk 027620/*420* F_1	11	156	156	1200	319	1831	18.81	
Salk 027620/*420* F_1	12	153	122	949	240	1464	20.99	
Salk 027620/*420* F_1	13	139	158	1107	279	1683	19.56	
Salk 027620/*420* F_1	14	160	148	1198	281	1787	19.05	
Salk 027620/*420* F_1	15	125	112	1007	241	1485	17.49	

Salk 027620/*420* F_1	16	138	141	1209	289	1777	17.18	
Salk 027620/*420* F_1	17	139	134	962	259	1494	20.34	
Salk 027620/*420* F_1	18	130	156	1184	325	1795	17.46	
Salk 027620/*420* F_1	19	108	131	978	240	1457	18.03	
Salk 027620/*420* F_1	20	143	120	1192	329	1784	16.03	
Salk 027620/*420* F_1	21	147	119	1153	297	1716	16.94	
Salk 027620/*420* F_1	22	137	130	1166	305	1738	16.77	0.6069
GK 380E06/*420* F_1	1	127	136	987	254	1504	19.36	
GK 380E06/*420* F_1	2	139	143	1140	269	1691	18.36	
GK 380E06/*420* F_1	3	111	131	1070	270	1582	16.69	
GK 380E06/*420* F_1	4	118	142	1074	289	1623	17.56	
GK 380E06/*420* F_1	5	130	157	1107	282	1676	18.91	
GK 380E06/*420* F_1	6	158	159	1189	258	1764	19.96	
GK 380E06/*420* F_1	7	114	120	974	253	1461	17.56	
GK 380E06/*420* F_1	8	123	101	861	212	1297	19.09	
GK 380E06/*420* F_1	9	104	107	973	235	1419	16.18	
GK 380E06/*420* F_1	10	111	126	882	239	1358	19.32	
GK 380E06/*420* F_1	11	107	120	1046	259	1532	16.12	
GK 380E06/*420* F_1	12	149	137	1109	250	1645	19.24	
GK 380E06/*420* F_1	13	135	140	1065	304	1644	18.42	0.9862
Salk 145153/*420* F_1	1	148	131	1081	292	1652	18.62	
Salk 145153/*420* F_1	2	145	138	1123	286	1692	18.42	
Salk 145153/*420* F_1	3	147	133	980	265	1525	20.45	
Salk 145153/*420* F_1	4	144	146	980	214	1484	21.95	
Salk 145153/*420* F_1	5	152	153	1166	257	1728	19.56	
Salk 145153/*420* F_1	6	161	142	1045	261	1609	21.05	
Salk 145153/*420* F_1	7	129	148	1032	254	1563	19.65	
Salk 145153/*420* F_1	8	143	129	1055	240	1567	19.20	
Salk 145153/*420* F_1	9	145	138	1084	277	1644	19.02	
Salk 145153/*420* F_1	10	152	172	1199	300	1823	19.72	
Salk 145153/*420* F_1	11	134	134	1049	237	1554	19.06	
Salk 145153/*420* F_1	12	135	135	1058	264	1592	18.71	
Salk 145153/*420* F_1	13	137	124	1049	277	1587	18.08	
Salk 145153/*420* F_1	14	113	148	1086	271	1618	17.70	
Salk 145153/*420* F_1	15	110	123	1011	260	1504	16.92	
Salk 145153/*420* F_1	16	136	132	1130	299	1697	17.29	
Salk 145153/*420* F_1	17	114	107	894	214	1329	18.30	
Salk 145153/*420* F_1	18	127	133	988	264	1512	19.00	
Salk 145153/*420* F_1	19	145	130	1049	272	1596	19.04	
Salk 145153/*420* F_1	20	119	127	1024	273	1543	17.47	
Salk 145153/*420* F_1	21	123	122	967	250	1462	18.46	
Salk 145153/*420* F_1	22	128	112	952	236	1428	18.52	
Salk 145153/*420* F_1	23	142	136	1225	307	1810	16.76	0.6401

Salk 024459/*420* F_1	1	148	147	1087	293	1675	19.52	
Salk 024459/*420* F_1	2	141	148	1130	290	1709	18.65	
Salk 024459/*420* F_1	3	128	114	1067	272	1581	16.70	
Salk 024459/*420* F_1	4	135	147	1105	250	1637	19.04	
Salk 024459/*420* F_1	5	134	139	1184	300	1757	16.98	
Salk 024459/*420* F_1	6	89	92	873	204	1258	15.61	
Salk 024459/*420* F_1	7	103	107	908	239	1357	16.90	
Salk 024459/*420* F_1	8	121	127	1149	273	1670	16.16	
Salk 024459/*420* F_1	9	132	115	1014	263	1524	17.79	
Salk 024459/*420* F_1	10	138	152	1214	295	1799	17.68	0.11654
Salk 059528/*420* F_1	1	126	135	1051	258	1570	18.30	
Salk 059528/*420* F_1	2	124	142	931	242	1439	20.61	
Salk 059528/*420* F_1	3	116	127	1017	261	1521	17.51	
Salk 059528/*420* F_1	4	141	154	1056	277	1628	20.15	
Salk 059528/*420* F_1	5	77	83	571	140	871	20.46	
Salk 059528/*420* F_1	6	123	128	941	258	1450	19.14	
Salk 059528/*420* F_1	7	149	162	1129	268	1708	20.26	
Salk 059528/*420* F_1	8	105	132	895	218	1350	19.45	
Salk 059528/*420* F_1	9	142	131	1001	272	1546	19.57	
Salk 059528/*420* F_1	10	123	114	1054	265	1556	16.61	
Salk 059528/*420* F_1	11	106	97	970	256	1429	15.39	
Salk 059528/*420* F_1	12	124	120	1130	284	1658	16.00	
Salk 059528/*420* F_1	13	139	128	1035	277	1579	18.65	
Salk 059528/*420* F_1	14	128	110	1069	303	1610	16.07	
Salk 059528/*420* F_1	15	121	106	893	240	1360	18.38	
Salk 059528/*420* F_1	16	91	110	995	288	1484	14.61	0.713
GK 345C06/*420* F_1	1	121	125	1013	229	1488	18.19	
GK 345C06/*420* F_1	2	111	103	931	226	1371	17.07	
GK 345C06/*420* F_1	3	108	106	1047	262	1523	15.21	
GK 345C06/*420* F_1	4	126	113	961	236	1436	18.32	
GK 345C06/*420* F_1	5	106	109	927	223	1365	17.24	
GK 345C06/*420* F_1	6	145	154	1158	308	1765	18.69	
GK 345C06/*420* F_1	7	122	109	983	240	1454	17.40	
GK 345C06/*420* F_1	8	97	79	804	202	1182	16.20	
GK 345C06/*420* F_1	9	160	146	1126	288	1720	19.74	
GK 345C06/*420* F_1	10	146	125	1073	277	1621	18.41	
GK 345C06/*420* F_1	11	135	129	970	233	1467	19.99	
GK 345C06/*420* F_1	12	112	120	928	253	1413	18.05	
GK 345C06/*420* F_1	13	122	125	1021	242	1510	17.97	
GK 345C06/*420* F_1	14	102	140	1010	270	1522	17.42	
GK 345C06/*420* F_1	15	116	113	941	253	1423	17.65	
GK 345C06/*420* F_1	16	123	118	981	241	1463	18.11	
GK 345C06/*420* F_1	17	146	129	1126	309	1710	17.64	0.319

Salk 014797/*420* F_1	1	76	76	672	163	987	16.81	
Salk 014797/*420* F_1	2	102	126	934	232	1394	17.97	
Salk 014797/*420* F_1	3	134	120	1140	280	1674	16.54	
Salk 014797/*420* F_1	4	140	133	1023	235	1531	19.79	
Salk 014797/*420* F_1	5	134	123	1159	302	1718	16.29	
Salk 014797/*420* F_1	6	119	131	1148	277	1675	16.24	
Salk 014797/*420* F_1	7	123	119	971	265	1478	17.99	
Salk 014797/*420* F_1	8	83	65	835	217	1200	13.21	
Salk 014797/*420* F_1	9	125	120	1048	256	1549	17.32	
Salk 014797/*420* F_1	10	117	127	1078	267	1589	16.76	
Salk 014797/*420* F_1	11	136	147	1139	266	1688	18.47	
Salk 014797/*420* F_1	12	158	160	1121	279	1718	20.64	
Salk 014797/*420* F_1	13	126	148	1078	272	1624	18.60	
Salk 014797/*420* F_1	14	145	117	1147	288	1697	16.86	
Salk 014797/*420* F_1	15	125	142	1013	251	1531	19.30	
Salk 014797/*420* F_1	16	88	79	714	183	1064	17.17	
Salk 014797/*420* F_1	17	154	144	1088	302	1688	19.57	0.14622
Salk 064633/*420* F_1	1	117	146	1176	325	1764	16.23	
Salk 064633/*420* F_1	2	165	134	1231	331	1861	17.62	
Salk 064633/*420* F_1	3	136	170	1207	287	1800	18.76	
Salk 064633/*420* F_1	4	140	129	1027	241	1537	19.38	
Salk 064633/*420* F_1	5	103	84	842	222	1251	16.27	
Salk 064633/*420* F_1	6	85	71	639	161	956	17.92	
Salk 064633/*420* F_1	7	127	121	1022	269	1539	17.68	
Salk 064633/*420* F_1	8	138	160	1220	336	1854	17.63	
Salk 064633/*420* F_1	9	143	153	1133	295	1724	18.97	
Salk 064633/*420* F_1	10	135	131	1163	302	1731	16.77	
Salk 064633/*420* F_1	11	160	140	1143	263	1706	19.48	
Salk 064633/*420* F_1	12	139	121	1139	297	1696	16.73	
Salk 064633/*420* F_1	13	132	162	1178	304	1776	18.21	
Salk 064633/*420* F_1	14	140	127	1138	279	1684	17.36	
Salk 064633/*420* F_1	15	121	134	1104	306	1665	16.71	
Salk 064633/*420* F_1	16	124	112	947	239	1422	18.26	0.18662
Salk 054181/*420* F_1	1	103	108	839	222	1272	18.25	
Salk 054181/*420* F_1	2	92	117	819	208	1236	18.65	
Salk 054181/*420* F_1	3	85	114	790	204	1193	18.37	
Salk 054181/*420* F_1	4	99	102	890	220	1311	16.73	
Salk 054181/*420* F_1	5	94	92	824	185	1195	17.01	
Salk 054181/*420* F_1	6	83	122	791	190	1186	19.11	0.774
Salk 109522/*420* F_1	1	85	112	730	187	1114	19.61	
Salk 109522/*420* F_1	2	109	94	796	211	1210	18.49	
Salk 109522/*420* F_1	3	88	100	742	166	1096	18.95	
Salk 109522/*420* F_1	4	92	92	770	183	1137	17.76	

Salk 109522/*420* F_1	5	97	98	763	195	1153	18.65	
Salk 109522/*420* F_1	6	92	99	738	181	1110	19.02	
Salk 109522/*420* F_1	7	110	103	836	185	1234	19.08	
Salk 109522/*420* F_1	8	102	98	748	185	1133	19.57	
Salk 109522/*420* F_1	9	107	105	754	194	1160	20.35	
Salk 109522/*420* F_1	10	88	82	711	178	1059	17.60	
Salk 109522/*420* F_1	11	80	84	650	176	990	18.23	
Salk 109522/*420* F_1	12	100	96	778	221	1195	18.03	
Salk 109522/*420* F_1	13	84	73	638	143	938	18.44	0.7511
Salk 050436/*420* F_1	1	89	89	723	178	1079	18.14	
Salk 050436/*420* F_1	2	89	87	671	165	1012	19.24	
Salk 050436/*420* F_1	3	105	88	760	192	1145	18.58	
Salk 050436/*420* F_1	4	75	69	710	165	1019	15.30	
Salk 050436/*420* F_1	5	98	113	830	199	1240	18.78	
Salk 050436/*420* F_1	6	117	86	750	174	1127	20.02	
Salk 050436/*420* F_1	7	90	87	773	189	1139	16.98	
Salk 050436/*420* F_1	8	65	79	723	178	1045	14.89	0.4052
Col-0/*420* F_1	1	151	154	1148	282	1735	19.48	
Col-0/*420* F_1	2	123	143	1086	252	1604	18.25	
Col-0/*420* F_1	3	134	139	1040	298	1611	18.69	
Col-0/*420* F_1	4	130	152	1122	262	1666	18.67	
Col-0/*420* F_1	5	122	99	816	203	1240	19.78	
Col-0/*420* F_1	6	77	89	680	189	1035	17.58	
Col-0/*420* F_1	7	90	95	761	184	1130	17.99	
Col-0/*420* F_1	8	95	81	790	194	1160	16.54	
Col-0/*420* F_1	9	92	110	738	196	1136	19.73	

Table 30. Fluorescent seed count data for the *420* FTL interval in an rQTL5^{420} BC$_1$F$_2$ fine mapping population. Population was generated from self-fertilisation of the Cvi*420*F$_3$(F$_2$ 24-4)/Col 1-6 line (BC$_1$F$_1$ line, see Figures 33 and 34). Genetic distance of the *420* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015).

Individual	Green (G)	Red (R)	Double-colour (D)	Non-colour (N)	Total (T)	cM
1-4	100	103	1004	275	1482	14.79
1-6	111	91	1127	306	1635	13.23
2-5	91	101	993	276	1461	14.14
3-5	105	102	783	196	1186	19.32
5-1	93	98	1023	289	1503	13.64
5-3	88	81	1046	282	1497	12.01
5-4	116	108	919	261	1404	17.48
5-5	89	77	864	239	1269	14.07
8-2	83	104	914	226	1327	15.26
8-5	98	81	917	247	1343	14.36
8-6	79	82	887	246	1294	13.33
10-6	67	69	793	226	1155	12.56
11-1	85	81	1034	258	1458	12.12
11-3	73	85	1034	309	1501	11.15
12-2	73	74	928	250	1325	11.79
12-4	96	102	1133	304	1635	12.95
14-1	83	87	845	241	1256	14.60
15-4	65	77	1006	283	1431	10.47

Table 31. Fluorescent seed count data for the *420* FTL interval in an rQTL5^{420} BC_1F_2 fine mapping population. Population was generated from self-fertilisation of the Cvi*420*F_3(F_2 2-3)/Col 1-5 line (BC_1F_1 line, see Figures 33 and 34). Genetic distance of the *420* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015).

Individual	Green (G)	Red (R)	Double-colour (D)	Non-colour (N)	Total (T)	cM
1-3	116	123	934	259	1432	18.38
1-5	114	126	1024	271	1535	17.10
2-1	130	130	1019	267	1546	18.54
2-2	136	162	1098	285	1681	19.66
2-4	134	121	984	241	1480	19.04
4-1	111	108	940	267	1426	16.76
4-2	116	114	797	200	1227	20.94
4-3	105	103	910	259	1377	16.46
5-5	124	135	1005	258	1522	18.78
5-6	141	149	1099	279	1668	19.24
6-1	154	117	1086	279	1636	18.23
6-5	154	144	1067	269	1634	20.30
8-3	151	116	1068	275	1610	18.25
9-4	123	142	1107	301	1673	17.34
10-2	130	146	1092	293	1661	18.29
10-5	112	126	1025	259	1522	17.10
10-6	141	129	1089	289	1648	18.00
11-2	97	105	1084	261	1547	14.04
11-3	136	143	1137	310	1726	17.74
12-4	131	123	1073	272	1599	17.40
13-5	103	112	1027	293	1535	15.15
14-3	104	123	1140	324	1691	14.47
14-4	131	141	1110	317	1699	17.55
15-6	94	84	997	259	1434	13.30

Table 32. Fluorescent seed count data for the *420* FTL interval in two rQTL5^{420} BC$_1$F$_3$ fine mapping populations. Populations were generated from self-fertilisation of the Cvi*420*F$_3$(F$_2$ 2-3)/Col 1-5 10-2 and 10-6 lines (BC$_1$F$_2$ lines, see Figures 33 and 34). Genetic distance of the *420* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015).

Parent	Individual	Green (G)	Red (R)	Double -colour (D)	Non-colour (N)	Total (T)	cM
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	2-1	115	123	1041	262	1541	16.87
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	3-3	117	121	1039	265	1542	16.85
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	4-5	126	132	1065	245	1568	18.09
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	5-2	111	129	1018	254	1512	17.38
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	5-5	145	147	1065	274	1631	19.88
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	6-5	158	154	1108	285	1705	20.37
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	6-6	146	123	1109	275	1653	17.87
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	7-5	115	141	1063	296	1615	17.36
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	8-1	165	167	1134	249	1715	21.72
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	8-4	154	130	1141	268	1693	18.48
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	8-5	137	152	1145	300	1734	18.35
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	9-1	135	135	1150	329	1749	16.86
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	10-4	126	103	943	238	1410	17.83
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	10-6	112	119	1105	286	1622	15.43
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	11-1	130	140	1067	276	1613	18.44
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	12-2	105	125	1097	291	1618	15.40
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	12-4	141	157	1204	276	1778	18.47
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	13-2	110	134	1079	257	1580	16.87
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	13-6	72	73	708	213	1066	14.68
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	14-1	123	118	1191	313	1745	14.92
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	14-6	124	129	1019	238	1510	18.46
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-2	15-4	135	145	1078	262	1620	19.11
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-6	3-1	117	123	1179	321	1740	14.90
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-6	5-3	104	112	1125	284	1625	14.32
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-6	6-2	122	111	1069	295	1597	15.85
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-6	6-4	126	126	1209	345	1806	15.09
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-6	11-3	117	127	1083	281	1608	16.54
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-6	13-1	115	112	1183	290	1700	14.39
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-6	13-5	96	133	1161	313	1703	14.50
Cvi420F$_3$(F$_2$ 2-3)/Col 1-5 10-6	15-3	87	104	1060	295	1546	13.23

Table 33. Logarithm of odds (LOD) scores for association between rQTL5^{420} peak genotyping markers on chromosome 5 and *420* recombination frequency in the Cvi-0/*420* rQTL5^{420} fine mapping population. Genotyping markers are indicated by a number representing their location on chromosome 5 (chromosome-kilobase) e.g. marker 5-22402 is found on chromosome 5 at position 22402070bp.

Marker	LOD
5-21349	2.71925
5-21562	2.92983
5-21906	3.55358
5-22313	4.54365
5-22402	5.04052
5-22654	5.33332
5-23875	6.63628
5-24192	6.44643
5-25212	6.39918
5-26907	6.25899

Table 34. Candidate loci identified for rQTL5[420]. Position of the locus on chromosome 5 is indicated in base pairs, and the predicted gene function based on locus information from TAIR (Berardini *et al*. 2015) and Araport11 (Krishnakumar *et al*. 2015) is given for each locus. Polymorphisms between Col-0 and Cvi-0 accessions that are not shared between Cvi-0 and Can-0 accessions (Salk Genome Express Browser, Alonso-Blanco *et al*. 2016) are given for each locus. T-DNA insertion lines used for locus analysis are listed with insertion position (Salk T-DNA Express tool (The Salk Institute Genomic Analysis Laboratory)).

Position (bp)	Locus	Gene	Function	Col/Cvi polymorphism	T-DNA line	T-DNA line insertion position (bp)
21395944-21399419	*AT5G52800*	DNA primase	DNA replication	Synonymous SNP in exon 8	Salk 090163	21396049-21396385; Exon 1
21410852-21412340	*AT5G52830*	WRKY transcription factor WRKY27	Transcription regulation	Non-synonymous SNP in exon 3	Salk 048952	21411522-21411664; Exon 2
21431836-21433712	*AT5G52880*	F-box family protein	F-box domain, cyclin-like	Synonymous SNP in intron 1; non-synonymous SNP in intron 3	SAIL 369C12	21433100-21433688; Exon 1
21548473-21552470	*AT5G53140*	Protein phosphatase 2C family protein	Protein dephosphorylation	Two non-synonymous SNPs in promoter	Salk 010368	21548264-21548506; Upstream

Table 35. Fluorescent seed count data for the *420* FTL interval in rQTL5^{420} candidate T-DNA/*420* F_1 lines. Genetic distance of the *420* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015). The aggregate number of recombinant (G+R) and non-recombinant (D+N) seeds for the replicates of each T-DNA/*420* F_1 line were compared to the Col-0/*420* F_1 replicates and used to construct 2x2 contingency tables and perform two-tailed χ^2 tests to test for significant differences. P-values for each T-DNA/*420* F_1 are given in the final column.

Line	Replicate	Green (G)	Red (R)	Double -colour (D)	Non-colour (N)	Total (T)	cM	p-value
Salk 010368/*420* F_1	1	94	113	700	172	1079	21.49	
Salk 010368/*420* F_1	2	84	86	746	168	1084	17.15	
Salk 010368/*420* F_1	3	96	97	873	235	1301	16.14	
Salk 010368/*420* F_1	4	93	97	716	170	1076	19.57	
Salk 010368/*420* F_1	5	91	96	834	199	1220	16.73	
Salk 010368/*420* F_1	6	87	86	734	182	1089	17.40	
Salk 010368/*420* F_1	7	100	93	724	180	1097	19.49	
Salk 010368/*420* F_1	8	80	87	788	192	1147	15.81	
Salk 010368/*420* F_1	9	109	104	905	206	1324	17.64	
Salk 010368/*420* F_1	10	69	89	760	180	1098	15.61	
Salk 010368/*420* F_1	11	61	68	585	143	857	16.40	
Salk 010368/*420* F_1	12	80	92	735	187	1094	17.20	
Salk 010368/*420* F_1	13	89	78	780	208	1155	15.69	
Salk 010368/*420* F_1	14	80	82	695	190	1047	16.90	
Salk 010368/*420* F_1	15	75	85	737	168	1065	16.36	0.0287 8
Salk 090163/*420* F_1	1	149	120	1064	283	1616	18.33	
Salk 090163/*420* F_1	2	117	125	1075	280	1597	16.52	
Salk 090163/*420* F_1	3	134	133	1042	284	1593	18.47	
Salk 090163/*420* F_1	4	115	126	947	232	1420	18.72	
Salk 090163/*420* F_1	5	117	141	1098	254	1610	17.57	
Salk 090163/*420* F_1	6	137	126	956	273	1492	19.54	
Salk 090163/*420* F_1	7	152	160	1069	283	1664	20.94	
Salk 090163/*420* F_1	8	133	160	1003	223	1519	21.63	
Salk 090163/*420* F_1	9	126	137	1169	285	1717	16.71	
Salk 090163/*420* F_1	10	132	124	1045	293	1594	17.61	
Salk 090163/*420* F_1	11	135	157	1132	272	1696	19.03	
Salk 090163/*420* F_1	12	142	156	1146	276	1720	19.16	
Salk 090163/*420* F_1	13	137	122	1043	254	1556	18.32	
Salk 090163/*420* F_1	14	119	118	1059	280	1576	16.38	
Salk 090163/*420* F_1	15	151	147	1076	255	1629	20.37	
Salk 090163/*420* F_1	16	148	141	1121	255	1665	19.20	

Salk 090163/*420* F_1	17	145	139	1095	264	1643	19.11	0.8679
Salk 048952/*420* F_1	1	137	127	1045	242	1551	18.79	
Salk 048952/*420* F_1	2	130	121	944	232	1427	19.49	
Salk 048952/*420* F_1	3	141	137	1089	279	1646	18.62	
Salk 048952/*420* F_1	4	114	129	1067	276	1586	16.72	
Salk 048952/*420* F_1	5	129	139	1047	285	1600	18.45	
Salk 048952/*420* F_1	6	123	159	1086	295	1663	18.71	
Salk 048952/*420* F_1	7	131	123	1048	255	1557	17.92	
Salk 048952/*420* F_1	8	143	119	1013	269	1544	18.72	
Salk 048952/*420* F_1	9	121	150	1076	284	1631	18.29	
Salk 048952/*420* F_1	10	112	127	980	233	1452	18.10	
Salk 048952/*420* F_1	11	142	155	1079	235	1611	20.55	
Salk 048952/*420* F_1	12	135	124	957	265	1481	19.36	
Salk 048952/*420* F_1	13	126	114	1000	239	1479	17.81	
Salk 048952/*420* F_1	14	120	114	1014	230	1478	17.33	
Salk 048952/*420* F_1	15	116	115	1000	245	1476	17.12	
Salk 048952/*420* F_1	16	155	127	1080	264	1626	19.18	0.7967
SAIL 369C12/*420* F_1	1	161	172	1210	318	1861	19.87	
SAIL 369C12/*420* F_1	2	142	129	1092	254	1617	18.46	
SAIL 369C12/*420* F_1	3	148	168	1100	292	1708	20.63	
SAIL 369C12/*420* F_1	4	134	122	1065	250	1571	17.90	
SAIL 369C12/*420* F_1	5	151	136	1092	239	1618	19.67	
SAIL 369C12/*420* F_1	6	130	151	1218	304	1803	17.04	
SAIL 369C12/*420* F_1	7	163	160	1222	306	1851	19.32	
SAIL 369C12/*420* F_1	8	145	163	1138	303	1749	19.51	
SAIL 369C12/*420* F_1	9	139	137	1137	266	1679	18.07	
SAIL 369C12/*420* F_1	10	148	146	1012	232	1538	21.41	
SAIL 369C12/*420* F_1	11	147	157	1130	302	1736	19.39	
SAIL 369C12/*420* F_1	12	144	140	1052	259	1595	19.76	
SAIL 369C12/*420* F_1	13	131	140	1074	272	1617	18.46	
SAIL 369C12/*420* F_1	14	119	101	949	238	1407	17.10	
SAIL 369C12/*420* F_1	15	124	130	1125	269	1648	16.83	
SAIL 369C12/*420* F_1	16	144	151	1086	276	1657	19.75	0.9238

Table 36. Fluorescent seed count data for the *420* FTL interval in the CanQTL4 BC_1F_1 fine mapping population. Population was generated from cross-fertilisation of the Can*420*F_3 (F_2 95-1) line with Col-0 (see Figures 40 and 41). Genetic distance of the *420* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015).

Individual	Green (G)	Red (R)	Double-colour (D)	Non-colour (N)	Total (T)	cM
4-1	107	79	1131	316	1633	12.13
5-1	127	133	863	219	1342	21.74
5-4	103	121	1094	290	1608	15.07
8-1	94	91	876	257	1318	15.19
8-3	93	90	902	222	1307	15.15
10-1	173	173	949	230	1525	26.09
12-2	173	174	973	213	1533	26.02
14-3	109	105	934	232	1380	16.94
15-1	133	122	1032	287	1574	17.78
15-6	149	161	927	198	1435	24.64
19-1	102	108	953	281	1444	15.79
19-3	133	127	1052	268	1580	18.09
19-5	113	145	905	230	1393	20.65
20-3	116	132	1114	296	1658	16.28
24-2	78	84	718	209	1089	16.19
24-5	128	109	1118	318	1673	15.34
25-3	123	99	805	195	1222	20.21

Table 37. Fluorescent seed count data for the *420* FTL interval in three CanQTL4 BC_1F_2 fine mapping populations. Populations were generated from self-fertilisation of the Can*420*F_3 (F_2 95-1)/Col 4-1, 12-2 and 19-5 lines (BC_1F_1 lines, see Figures 40 and 41). Genetic distance of the *420* interval is calculated as cM = 100 x $(1-(1-2(G+R)/T)^{1/2})$. G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015).

Parent	Individual	Green (G)	Red (R)	Double -colour (D)	Non-colour (N)	Total (T)	cM
Can420F_3(F_2 95-1)/Col 4-1	1-4	94	92	1080	304	1570	12.65
Can420F_3(F_2 95-1)/Col 4-1	1-5	80	77	1040	320	1517	10.95
Can420F_3(F_2 95-1)/Col 4-1	2-2	98	94	1124	312	1628	12.59
Can420F_3(F_2 95-1)/Col 4-1	2-4	63	93	1030	289	1475	11.20
Can420F_3(F_2 95-1)/Col 4-1	3-1	83	74	1087	290	1534	10.82
Can420F_3(F_2 95-1)/Col 4-1	3-4	102	102	1167	366	1737	12.53
Can420F_3(F_2 95-1)/Col 4-1	4-1	89	81	1012	284	1466	12.36
Can420F_3(F_2 95-1)/Col 4-1	4-2	136	114	1125	302	1677	16.22
Can420F_3(F_2 95-1)/Col 4-1	4-4	122	116	1131	309	1678	15.36
Can420F_3(F_2 95-1)/Col 4-1	4-6	107	130	1108	300	1645	15.63
Can420F_3(F_2 95-1)/Col 4-1	5-4	89	105	1144	291	1629	12.72
Can420F_3(F_2 95-1)/Col 4-1	5-5	97	76	1077	326	1576	11.66
Can420F_3(F_2 95-1)/Col 4-1	5-6	106	93	1172	347	1718	12.35
Can420F_3(F_2 95-1)/Col 4-1	7-1	126	115	1170	316	1727	15.09
Can420F_3(F_2 95-1)/Col 4-1	7-4	71	83	964	261	1379	11.87
Can420F_3(F_2 95-1)/Col 4-1	8-1	74	79	1227	357	1737	9.23
Can420F_3(F_2 95-1)/Col 4-1	8-6	75	67	953	290	1385	10.84
Can420F_3(F_2 95-1)/Col 4-1	9-4	91	88	1082	322	1583	12.03
Can420F_3(F_2 95-1)/Col 4-1	10-3	72	58	1061	331	1522	8.94
Can420F_3(F_2 95-1)/Col 4-1	11-2	82	71	872	263	1288	12.68
Can420F_3(F_2 95-1)/Col 4-1	11-3	103	126	1059	274	1562	15.93
Can420F_3(F_2 95-1)/Col 4-1	12-2	77	78	1144	299	1598	10.22
Can420F_3(F_2 95-1)/Col 4-1	12-6	61	72	985	267	1385	10.11
Can420F_3(F_2 95-1)/Col 4-1	14-4	73	68	899	275	1315	11.37
Can420F_3(F_2 95-1)/Col 4-1	14-5	80	92	1022	302	1496	12.25
Can420F_3(F_2 95-1)/Col 4-1	15-1	81	78	1151	309	1619	10.36
Can420F_3(F_2 95-1)/Col 12-2	1-1	158	141	991	239	1529	21.97
Can420F_3(F_2 95-1)/Col 12-2	2-1	145	174	820	177	1316	28.22
Can420F_3(F_2 95-1)/Col 12-2	2-2	161	161	916	208	1446	25.53
Can420F_3(F_2 95-1)/Col 12-2	3-1	186	183	992	247	1608	26.44
Can420F_3(F_2 95-1)/Col 12-2	3-4	145	150	959	188	1442	23.13
Can420F_3(F_2 95-1)/Col 12-2	4-1	152	165	926	214	1457	24.84
Can420F_3(F_2 95-1)/Col 12-2	4-2	167	156	940	203	1466	25.21
Can420F_3(F_2 95-1)/Col 12-2	5-2	171	172	1075	272	1690	22.92

Can420F$_3$(F$_2$ 95-1)/Col 12-2	5-4	161	160	943	213	1477	24.81
Can420F$_3$(F$_2$ 95-1)/Col 12-2	5-5	169	170	916	210	1465	26.71
Can420F$_3$(F$_2$ 95-1)/Col 12-2	6-1	193	190	927	197	1507	29.88
Can420F$_3$(F$_2$ 95-1)/Col 12-2	7-6	166	170	969	217	1522	25.27
Can420F$_3$(F$_2$ 95-1)/Col 12-2	8-2	167	160	893	197	1417	26.62
Can420F$_3$(F$_2$ 95-1)/Col 12-2	8-4	178	170	900	163	1411	28.81
Can420F$_3$(F$_2$ 95-1)/Col 12-2	9-4	129	136	846	174	1285	23.35
Can420F$_3$(F$_2$ 95-1)/Col 12-2	9-5	170	173	939	230	1512	26.09
Can420F$_3$(F$_2$ 95-1)/Col 12-2	9-6	193	174	953	200	1520	28.09
Can420F$_3$(F$_2$ 95-1)/Col 12-2	10-1	179	192	928	205	1504	28.82
Can420F$_3$(F$_2$ 95-1)/Col 12-2	11-1	157	172	961	195	1485	25.37
Can420F$_3$(F$_2$ 95-1)/Col 12-2	11-6	185	152	865	220	1422	27.47
Can420F$_3$(F$_2$ 95-1)/Col 12-2	12-2	180	188	934	190	1492	28.82
Can420F$_3$(F$_2$ 95-1)/Col 12-2	12-4	169	179	942	197	1487	27.07
Can420F$_3$(F$_2$ 95-1)/Col 12-2	12-6	181	154	1016	226	1577	24.16
Can420F$_3$(F$_2$ 95-1)/Col 19-5	1-1	154	167	1007	254	1582	22.92
Can420F$_3$(F$_2$ 95-1)/Col 19-5	3-1	182	195	891	189	1457	30.54
Can420F$_3$(F$_2$ 95-1)/Col 19-5	3-3	198	187	934	193	1512	29.95
Can420F$_3$(F$_2$ 95-1)/Col 19-5	4-2	152	174	945	202	1473	25.34
Can420F$_3$(F$_2$ 95-1)/Col 19-5	4-3	194	223	982	194	1593	30.97
Can420F$_3$(F$_2$ 95-1)/Col 19-5	7-1	167	166	961	234	1528	24.89
Can420F$_3$(F$_2$ 95-1)/Col 19-5	9-1	195	172	1101	270	1738	24.00
Can420F$_3$(F$_2$ 95-1)/Col 19-5	11-4	140	136	916	219	1411	21.98
Can420F$_3$(F$_2$ 95-1)/Col 19-5	11-5	190	185	888	181	1444	30.67
Can420F$_3$(F$_2$ 95-1)/Col 19-5	14-6	168	191	1006	234	1599	25.77
Can420F$_3$(F$_2$ 95-1)/Col 19-5	15-1	158	156	1049	269	1632	21.57
Can420F$_3$(F$_2$ 95-1)/Col 19-5	27-6	160	168	1072	279	1679	21.94
Can420F$_3$(F$_2$ 95-1)/Col 19-5	28-3	169	159	945	224	1497	25.05
Can420F$_3$(F$_2$ 95-1)/Col 19-5	31-5	136	152	908	214	1410	23.09
Can420F$_3$(F$_2$ 95-1)/Col 19-5	35-3	162	174	916	210	1462	26.49
Can420F$_3$(F$_2$ 95-1)/Col 19-5	37-2	181	185	1161	296	1823	22.64
Can420F$_3$(F$_2$ 95-1)/Col 19-5	37-5	179	185	995	246	1605	26.08
Can420F$_3$(F$_2$ 95-1)/Col 19-5	38-6	130	144	787	195	1256	24.92
Can420F$_3$(F$_2$ 95-1)/Col 19-5	43-6	184	198	846	171	1399	32.63
Can420F$_3$(F$_2$ 95-1)/Col 19-5	45-3	147	173	1116	273	1709	20.91
Can420F$_3$(F$_2$ 95-1)/Col 19-5	45-4	155	154	933	220	1462	24.02
Can420F$_3$(F$_2$ 95-1)/Col 19-5	47-1	181	182	1030	258	1651	25.15
Can420F$_3$(F$_2$ 95-1)/Col 19-5	47-4	180	166	986	241	1573	25.16
Can420F$_3$(F$_2$ 95-1)/Col 19-5	48-3	183	156	1057	263	1659	23.10
Can420F$_3$(F$_2$ 95-1)/Col 19-5	50-2	166	157	1044	236	1603	22.73
Can420F$_3$(F$_2$ 95-1)/Col 19-5	51-2	191	169	899	175	1434	29.44
Can420F$_3$(F$_2$ 95-1)/Col 19-5	54-4	197	226	960	194	1577	31.92
Can420F$_3$(F$_2$ 95-1)/Col 19-5	54-5	200	182	1088	243	1713	25.57

Can420F_3(F_2 95-1)/Col 19-5	57-4	172	176	1046	236	1630	24.30
Can420F_3(F_2 95-1)/Col 19-5	60-1	170	165	793	178	1306	30.22
Can420F_3(F_2 95-1)/Col 19-5	60-6	134	151	1074	269	1628	19.39
Can420F_3(F_2 95-1)/Col 19-5	61-5	174	176	877	199	1426	28.65

Table 38. Logarithm of odds (LOD) scores for association between CanQTL4 peak genotyping markers on chromosome 4 and *420* recombination frequency in the Can-0/*420* fine mapping population. Genotyping markers are indicated by a number representing their location on chromosome 4 (chromosome-kilobase) e.g. marker 4-230 is found on chromosome 4 at position 230388bp.

Marker	LOD
4-230	10.6309
4-1160	10.9249
4-2117	10.9924
4-2173	11.1312
4-2392	11.1312
4-2506	11.1312
4-2618	11.1312
4-2807	11.1
4-2852	11.0776
4-3363	11.0774
4-4636	11.0577
4-4844	10.6628
4-4871	10.6628

Table 39. Candidate loci identified for CanQTL4. Position of the locus on chromosome 4 is indicated in base pairs, and the predicted gene function based on locus information from TAIR (Berardini *et al*. 2015) and Araport11 (Krishnakumar *et al*. 2015) is given for each locus. Polymorphisms between Col-0 and Can-0 accessions that are not shared between Can-0 and Cvi-0 accessions (Salk Genome Express Browser, Alonso-Blanco *et al*. 2016) are given for each locus. T-DNA insertion lines used for locus analysis are listed with insertion position (Salk T-DNA Express tool (The Salk Institute Genomic Analysis Laboratory)).

Position (bp)	Locus	Gene	Function	Col/Can polymorphism	T-DNA line	T-DNA line insertion position (bp)
5196787-5198238	*AT4G08250*	GRAS family transcription factor	Transcription regulation	Three SNPs in promoter; one non-synonymous SNP in exon 1	N/A	N/A
5200847-5201865	*AT4G08260*	Protein phosphatase 2C family protein	Protein dephosphorylation	Deletion in Can-0	N/A	N/A
5248864-5252150	*AT4G08310*	DNA ligase	Ligase activity	One non-synonymous SNP in exon 1; five SNPs in intron 1; one non-synonymous SNP in exon 2; one synonymous SNP in exon 3; one SNP in intron 3; one synonymous SNP in exon 9	Salk 146231 Salk 047663	5251892-5252238; Exon 1 5252084-5252437; Exon 1
5252811-5255200	*AT4G08320*	TPR8: carboxylate clamp (CC)-tetratricopeptide repeat (TPR) protein	Potential to interact with Hsp90/Hsp70 as a co-chaperone	Multiple non-synonymous SNPs in all exons	GK-219G07 SAIL 731H04	5253395-5253590; Exon 3 5254495-5254533; Exon 8

Table 40. Fluorescent seed count data for the *420* FTL interval in CanQTL4 candidate T-DNA/*420* F_1 lines. Genetic distance of the *420* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015). The aggregate number of recombinant (G+R) and non-recombinant (D+N) seeds for the replicates of each T-DNA/*420* F_1 line were compared to the Col-0/*420* F_1 replicates and used to construct 2x2 contingency tables and perform two-tailed χ^2 tests to test for significant differences. P-values for each T-DNA/*420* F_1 are given in the final column.

Line	Replicate	Green (G)	Red (R)	Double-colour (D)	Non-colour (N)	Total (T)	cM	p-value
GK 219G07/*420* F_1	1	88	93	765	197	1143	17.34	
GK 219G07/*420* F_1	2	102	116	718	171	1107	22.14	
GK 219G07/*420* F_1	3	105	113	835	196	1249	19.32	
GK 219G07/*420* F_1	4	89	109	743	170	1111	19.78	
GK 219G07/*420* F_1	5	101	117	769	189	1176	20.67	
GK 219G07/*420* F_1	6	115	93	796	198	1202	19.14	
GK 219G07/*420* F_1	7	95	104	764	178	1141	19.30	
GK 219G07/*420* F_1	8	129	108	889	214	1340	19.61	
GK 219G07/*420* F_1	9	108	108	788	208	1212	19.78	
GK 219G07/*420* F_1	10	105	96	759	190	1150	19.35	
GK 219G07/*420* F_1	11	110	119	767	200	1196	21.45	
GK 219G07/*420* F_1	12	120	105	799	190	1214	20.67	
GK 219G07/*420* F_1	13	107	130	863	205	1305	20.20	
GK 219G07/*420* F_1	14	106	109	833	189	1237	19.23	
GK 219G07/*420* F_1	15	112	115	826	204	1257	20.07	
GK 219G07/*420* F_1	16	107	115	799	190	1211	20.42	0.031
SAIL731H04/*420* F_1	1	78	110	821	203	1212	16.95	
SAIL731H04/*420* F_1	2	92	74	708	175	1049	17.33	
SAIL731H04/*420* F_1	3	94	98	833	197	1222	17.19	
SAIL731H04/*420* F_1	4	101	92	790	188	1171	18.12	
SAIL731H04/*420* F_1	5	95	93	832	189	1209	16.99	
SAIL731H04/*420* F_1	6	88	79	770	187	1124	16.16	
SAIL731H04/*420* F_1	7	100	106	766	190	1162	19.66	0.1776
Salk 146231/*420* F_1	1	136	131	1109	251	1627	18.04	
Salk 146231/*420* F_1	2	154	153	1094	282	1683	20.30	
Salk 146231/*420* F_1	3	119	115	996	268	1498	17.08	
Salk 146231/*420* F_1	4	126	138	1083	247	1594	18.22	
Salk 146231/*420* F_1	5	132	135	969	253	1489	19.91	
Salk 146231/*420* F_1	6	126	134	1090	270	1620	17.60	
Salk 146231/*420* F_1	7	135	133	1055	282	1605	18.39	
Salk 146231/*420* F_1	8	142	158	1038	220	1558	21.59	
Salk 146231/*420* F_1	9	145	139	1090	269	1643	19.11	

Salk 146231/*420* F_1	10	141	141	1097	266	1645	18.94	
Salk 146231/*420* F_1	11	117	134	897	227	1375	20.32	
Salk 146231/*420* F_1	12	138	186	1060	234	1618	22.57	
Salk 146231/*420* F_1	13	135	148	1057	226	1566	20.09	
Salk 146231/*420* F_1	14	138	116	1043	244	1541	18.13	
Salk 146231/*420* F_1	15	129	162	1060	290	1641	19.67	
Salk 146231/*420* F_1	16	118	157	1077	248	1600	18.99	0.3276
Salk 047663/*420* F_1	1	107	121	1063	253	1544	16.06	
Salk 047663/*420* F_1	2	122	124	1061	280	1587	16.93	
Salk 047663/*420* F_1	3	123	123	1027	245	1518	17.79	
Salk 047663/*420* F_1	4	134	140	1025	256	1555	19.53	
Salk 047663/*420* F_1	5	113	101	1011	282	1507	15.38	
Salk 047663/*420* F_1	6	117	123	1066	269	1575	16.62	
Salk 047663/*420* F_1	7	125	119	978	257	1479	18.14	
Salk 047663/*420* F_1	8	128	109	1039	270	1546	16.73	
Salk 047663/*420* F_1	9	130	132	1037	270	1569	18.39	
Salk 047663/*420* F_1	10	119	128	902	240	1389	19.73	
Salk 047663/*420* F_1	11	159	134	1129	267	1689	19.19	
Salk 047663/*420* F_1	12	136	125	1015	268	1544	18.64	
Salk 047663/*420* F_1	13	145	130	1027	290	1592	19.10	
Salk 047663/*420* F_1	14	131	134	1080	254	1599	18.24	0.3422

Table 41. Fluorescent seed count data for the *420* FTL interval in rQTL1^{420} BC_1F_1 and BC_1F_2 fine mapping lines. Parent line denotes the line that was self-fertilised to generate the individual. Genetic distance of the *420* interval is calculated as cM = 100 x $(1-(1-2(G+R)/T)^{1/2})$. G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015).

Parent	Individual	Green (G)	Red (R)	Double-colour (D)	Non-colour (N)	Total (T)	cM
Cvi420F_3(F_2 27-3)/Col	1-1	78	115	958	288	1439	14.46
Cvi420F_3(F_2 27-3)/Col	1-2	76	72	729	203	1080	14.80
Cvi420F_3(F_2 27-3)/Col 1-1	1-4	119	111	1142	305	1677	14.81
Cvi420F_3(F_2 13-1)/Col	1-4	63	59	755	218	1095	11.84
Cvi420F_3(F_2 13-1)/Col 1-4	1-8	93	88	1019	286	1486	13.03

Table 42. Fluorescent seed count data for the *420* FTL interval in rQTL1^{420}NILs. Genetic distance of the 420 interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015). The aggregate number of recombinant (G+R) and non-recombinant (D+N) seeds for Col-0/*420* F_1 replicates (see Table 26), and rQTL1^{420}NIL replicates were used to construct a 2x2 contingency table and perform a two-tailed χ^2 test to test for significant differences.

Line	Replicate	Green (G)	Red (R)	Double-colour (D)	Non-colour (N)	Total (T)	cM
rQTL1(420)NIL	1	123	131	1107	264	1625	17.09
rQTL1(420)NIL	2	111	130	1066	284	1591	16.51
rQTL1(420)NIL	3	111	94	993	276	1474	15.04
rQTL1(420)NIL	4	122	117	1046	246	1531	17.07
rQTL1(420)NIL	5	121	115	1004	292	1532	16.82
rQTL1(420)NIL	6	94	119	945	228	1386	16.77
rQTL1(420)NIL	7	102	124	967	271	1464	16.86
rQTL1(420)NIL	8	117	126	1026	282	1551	17.14
rQTL1(420)NIL	9	123	113	1046	281	1563	16.45
rQTL1(420)NIL	10	102	106	1071	272	1551	14.46

$\chi^2(1) = 33.19261$, $p=1.67 \times 10^{-8}$

Table 43. Fluorescent seed count data for the *5.11* FTL interval in rQTL1^{420}NIL/*5.11* replicates. For comparable Col-0/*5.11* F_1 line data see Table 22. Genetic distance of the *5.11* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015). The aggregate number of recombinant (G+R) and non-recombinant (D+N) seeds for Col-0/*5.11* F_1 replicates (see Table 6), and rQTL1^{420}NIL/*5.11* replicates were used to construct a 2x2 contingency table and perform a two-tailed χ^2 test to test for significant differences.

Line	Replicate	Green (G)	Red (R)	Double-colour (D)	Non-colour (N)	Total (T)	cM
rQTL1(420)NIL/*5.11*	1	125	123	864	192	1304	21.28
rQTL1(420)NIL/*5.11*	2	102	127	838	200	1267	20.09
rQTL1(420)NIL/*5.11*	3	117	125	830	164	1236	22.00
rQTL1(420)NIL/*5.11*	4	110	96	713	191	1110	20.70
rQTL1(420)NIL/*5.11*	5	131	108	937	229	1405	18.77
rQTL1(420)NIL/*5.11*	6	98	102	681	176	1057	21.16
rQTL1(420)NIL/*5.11*	7	94	93	749	212	1148	17.89
rQTL1(420)NIL/*5.11*	8	144	126	880	217	1367	22.22

$\chi^2(1) = 1.950031$, p=0.3252

Table 44. Fluorescent seed count data for the *420* FTL interval in Col-0/*420* F_1 lines and *HEI10^{Col}420, HEI10^{Ler}420, HEI10^{Cvi}420* and *HEI10^{R264G}420* T_1 lines. Genetic distance of the *420* interval is calculated as cM = 100 x (1-(1-2(G+R)/T)$^{1/2}$). G is the number of seeds with only green fluorescence, R is the number of seeds with only red fluorescence and T is the total number of seeds analysed (Ziolkowski *et al*. 2015). The aggregate number of recombinant (G+R) and non-recombinant (D+N) seeds for *HEI10^{Col}420, HEI10^{Ler}420, HEI10^{Cvi}420* and *HEI10^{R264G}420* T_1 lines were each used to construct a 2x2 contingency table and perform a two-tailed χ^2 test to test for significant differences from Col-0/*420* F_1 lines.

Line	Individual	Green (G)	Red (R)	Double-colour (D)	Non-colour (N)	Total (T)	cM
HEI10^{Col}420	1	205	184	1072	235	1696	26.43
HEI10^{Col}420	2	244	250	1133	228	1855	31.63
HEI10^{Col}420	3	176	184	1082	268	1710	23.91
HEI10^{Col}420	4	147	166	841	152	1306	27.84
HEI10^{Col}420	5	204	183	1013	214	1614	27.86
HEI10^{Col}420	6	238	250	1157	239	1884	30.58
HEI10^{Col}420	7	155	158	1014	212	1539	22.98
HEI10^{Col}420	8	151	159	1102	269	1681	20.55
HEI10^{Col}420	9	173	165	1024	231	1593	24.13
HEI10^{Col}420	10	249	238	900	172	1559	38.74
HEI10^{Col}420	11	103	121	532	107	863	30.65
HEI10^{Col}420	12	147	169	1013	251	1580	22.54
HEI10^{Col}420	13	158	170	1085	234	1647	22.43
HEI10^{Col}420	14	265	253	898	142	1558	42.12
HEI10^{Col}420	15	157	177	994	211	1539	24.77
HEI10^{Col}420	16	216	263	886	153	1518	39.26
HEI10^{Col}420	17	264	249	899	120	1532	42.53
HEI10^{Col}420	18	134	155	822	161	1272	26.14
HEI10^{Col}420	19	132	138	862	183	1315	23.23
HEI10^{Col}420	20	131	134	838	179	1282	23.41
HEI10^{R264G}420	1	107	100	944	234	1385	16.27
HEI10^{R264G}420	2	163	132	1039	266	1600	20.55
HEI10^{R264G}420	3	144	175	1085	281	1685	21.17
HEI10^{R264G}420	4	184	172	1092	255	1703	23.72

HEI10^{R264G}*420*	5	147	157	1004	254	1562	21.85
HEI10^{R264G}*420*	6	201	176	960	177	1514	29.15
HEI10^{R264G}*420*	7	248	236	847	138	1469	41.60
HEI10^{R264G}*420*	8	110	111	483	101	805	32.85
HEI10^{R264G}*420*	9	166	160	892	187	1405	26.79
HEI10^{R264G}*420*	10	240	236	1107	206	1789	31.60
HEI10^{R264G}*420*	11	166	159	1147	288	1760	20.58
HEI10^{R264G}*420*	12	195	211	985	238	1629	29.18
HEI10^{R264G}*420*	13	167	177	1087	231	1662	23.45
HEI10^{R264G}*420*	14	220	221	1159	224	1824	28.14
HEI10^{R264G}*420*	15	173	169	1155	287	1784	21.48
HEI10^{R264G}*420*	16	154	140	1073	254	1621	20.17
HEI10^{R264G}*420*	17	175	191	783	138	1287	34.33
HEI10^{R264G}*420*	18	51	44	816	238	1149	8.64
HEI10^{R264G}*420*	19	88	75	1075	344	1582	10.90
HEI10^{R264G}*420*	20	167	153	908	205	1433	25.61
HEI10^{R264G}*420*	21	158	142	1109	272	1681	19.81
HEI10^{R264G}*420*	22	216	190	1110	220	1736	27.04
HEI10^{R264G}*420*	23	149	161	1194	310	1814	18.87
HEI10^{R264G}*420*	24	222	226	997	170	1615	33.28
HEI10^{R264G}*420*	25	209	250	911	159	1529	36.79
HEI10^{R264G}*420*	26	166	184	953	182	1485	27.29
HEI10^{Cvi}*420*	1	156	189	860	165	1370	29.55
HEI10^{Cvi}*420*	2	214	179	953	179	1525	30.39
HEI10^{Cvi}*420*	3	244	245	931	141	1561	38.89
HEI10^{Cvi}*420*	4	169	185	1010	260	1624	24.90
HEI10^{Cvi}*420*	5	132	130	1083	257	1602	17.97
HEI10^{Cvi}*420*	6	159	177	1051	212	1599	23.86
HEI10^{Cvi}*420*	7	102	94	616	160	972	22.75
HEI10^{Cvi}*420*	8	152	179	998	246	1575	23.86
HEI10^{Cvi}*420*	9	132	137	701	147	1117	28.00
HEI10^{Ler}*420*	1	140	164	1112	285	1701	19.84
HEI10^{Ler}*420*	2	206	195	995	241	1637	28.58

HEI10^{Ler}420	3	121	124	1145	298	1688	15.76
HEI10^{Ler}420	4	161	168	1081	294	1704	21.65
HEI10^{Ler}420	5	157	147	1144	284	1732	19.44
HEI10^{Ler}420	6	186	176	1116	295	1773	23.08
HEI10^{Ler}420	7	191	176	1249	282	1898	21.69
HEI10^{Ler}420	8	97	123	770	182	1172	20.97
HEI10^{Ler}420	9	175	155	1119	258	1707	21.68
HEI10^{Ler}420	10	142	131	1119	300	1692	17.70
HEI10^{Ler}420	11	156	147	1057	262	1622	20.86
HEI10^{Ler}420	12	172	163	940	211	1486	25.90
Col-0/*420*	1	119	139	1098	285	1641	17.20
Col-0/*420*	2	152	133	1021	254	1560	20.34
Col-0/*420*	3	146	134	1090	251	1621	19.10
Col-0/*420*	4	186	191	1261	319	1957	21.60
Col-0/*420*	5	150	148	1184	254	1736	18.96
Col-0/*420*	6	128	154	1143	266	1691	18.36
Col-0/*420*	7	107	100	946	244	1397	16.12
Col-0/*420*	8	121	101	855	204	1281	19.17
Col-0/*420*	9	121	128	1049	240	1538	17.77
Col-0/*420*	10	117	112	1012	241	1482	16.88
Col-0/*420*	11	145	148	1138	294	1725	18.74
Col-0/*420*	12	133	134	1164	286	1717	16.99

HEI10^{Col}*420* $\chi^2(1) = 395.7708$, $p=9.18 \times 10^{-88}$

HEI10^{Ler}*420* $\chi^2(1) = 33.98373$, $p=1.11 \times 10^{-8}$

HEI10^{Cvi}*420* $\chi^2(1) = 191.5251$, $p=2.95 \times 10^{-43}$

HEI10^{R264G}*420* $\chi^2(1) = 186.644$, $p=3.44 \times 10^{-42}$

Table 45. Fluorescent pollen flow cytometry count data for the three-colour *i3bc* FTL interval in Col-0/*i3bc* (Wild Type) and *HEI10*overexpressor/*i3bc* F_1 lines. Count data for each of the eight possible fluorescence classes is given for each individual. Fluorescence class indicated by letters denoting red (R), yellow (Y) and blue (B) fluorescence. For example, BYR denotes pollen showing red, yellow and blue fluorescence, where bYr denotes pollen showing only yellow fluorescence.

Genotype	**Replicate**	BYR	byr	bYr	ByR	BYr	byR	bYR	Byr	**Total**
Wild Type	1	22959	35024	111	41	1480	1448	4799	5637	71499
Wild Type	2	40147	56509	153	72	2443	2532	8540	9440	119836
Wild Type	3	24971	38383	121	59	1391	1573	5513	6984	78995
Wild Type	4	26872	36684	108	212	1586	1521	5713	6593	79289
Wild Type	5	36056	55040	175	151	2481	2609	8268	9422	114202
Wild Type	6	29694	39888	115	53	1825	1946	6729	7158	87408
Wild Type	7	27387	40104	107	49	1612	1771	5468	6293	82791
Wild Type	8	25144	39606	119	84	1547	1674	5739	5622	79535
Wild Type	9	20840	28144	76	46	1340	1381	4453	5350	61630
Wild Type	10	20088	30745	89	61	1111	1239	3484	4460	61277
Total		274158	400127	1174	828	16816	17694	58706	66959	836462
HEI10 overexpressor	1	6868	11307	254	237	1068	1081	3250	3926	27991
HEI10 overexpressor	2	4033	6668	156	141	634	697	2217	2539	17085
HEI10 overexpressor	3	5869	9663	232	133	812	919	2951	3048	23627
HEI10 overexpressor	4	6218	9922	251	190	946	1058	3250	3426	25261
HEI10 overexpressor	5	5864	9853	224	228	933	962	3010	3114	24188
HEI10 overexpressor	6	9593	16472	388	336	1491	1549	4807	5362	39998
HEI10 overexpressor	7	5156	7089	206	176	718	824	2324	2799	19292
HEI10 overexpressor	8	3436	5352	144	114	525	585	1655	2065	13876

HEI10 overexpressor	9	4041	7920	175	117	613	685	1941	2482	17974
HEI10 overexpressor	10	4497	7816	163	88	685	747	2178	2707	18881
HEI10 overexpressor	11	7413	12843	248	238	1061	1094	3501	3939	30337
HEI10 overexpressor	12	3647	5936	129	122	516	569	1729	2126	14774
HEI10 overexpressor	13	3876	5750	123	118	569	624	1787	2093	14940
HEI10 overexpressor	14	3568	4352	95	139	469	484	1752	1933	12792
HEI10 overexpressor	15	4243	5676	140	168	665	575	2230	2365	16062
Total		78322	126619	2928	2545	11705	12453	38582	43924	317078

Table 46. Genetic distance of the three-colour *i3bc* FTL interval in Col-0/*i3bc* and *HEI10* overexpressor/*i3bc* F_1 lines and measurement of crossover coincidence. Genetic distances of the *I3b* and *I3c* intervals were calculated as *i3b* cM=(NbYr + NByR + NbYR + NByr) / Ntotal, *i3c* cM=(NbYr + NByR + NBYr + NbyR) / Ntotal. These values correspond to pollen count data from Table 45, where for example, NbYr is the number of pollen expressing only yellow fluorescence Ntotal is the sum of pollen counts from all eight fluorescence classes. Crossover interference was calculated as: Observed DCOs = (NbYr + NByR). Expected DCOs = (*i3b* cM/100)*(*i3c* cM/100)*Ntotal. Coefficient of interference = observed DCOs/expected DCOs. Interference =1-CoC (Yelina *et al*. 2013; Ziolkowski *et al.* 2015). The aggregate number of recombinant (i3b = Byr+bYR+ByR+bYr, i3c=bYr+ByR+BYr+byR) and non-recombinant pollen (i3b=BYR+byr+BYr+byR, i3c=BYR+byr+bYR+Byr) for *HEI10* overexpressor/*i3bc* F_1 replicates and Col-0/*i3bc* F_1 replicates were used to construct 2x2 contingency tables and perform two-tailed χ^2 tests to test for significant differences in genetic distance. The *i3bc* interference estimates for *HEI10* overexpressor/*i3bc* F_1 lines were compared to Col-0/*i3bc* F_1 replicates using a two sample t-Test assuming unequal variances to test for significant differences ($p=2.66 \times 10^{-6}$).

Genotype	Replicate	*i3b* cM	*i3c* cM	Expected DCOs	Observed DCOs	1-CoC
Wild Type	1	14.81	4.31	456.10	152	0.67
Wild Type	2	15.19	4.34	789.96	225	0.72
Wild Type	3	16.05	3.98	504.54	180	0.64
Wild Type	4	15.92	4.32	545.72	320	0.41
Wild Type	5	15.78	4.74	854.40	326	0.62
Wild Type	6	16.08	4.51	633.38	168	0.73
Wild Type	7	14.39	4.27	509.41	156	0.69
Wild Type	8	14.54	4.31	497.83	203	0.59
Wild Type	9	16.10	4.61	457.84	122	0.73
Wild Type	10	13.21	4.08	330.22	150	0.55
Mean		15.21	4.35	557.94	200.20	0.64
HEI10 overexpressor	1	27.39	9.43	723.12	491	0.32
HEI10 overexpressor	2	29.58	9.53	481.49	297	0.38
HEI10 overexpressor	3	26.94	8.87	564.56	365	0.35
HEI10 overexpressor	4	28.17	9.68	688.85	441	0.36
HEI10 overexpressor	5	27.19	9.70	638.08	452	0.29
HEI10 overexpressor	6	27.23	9.41	1025.08	724	0.29
HEI10 overexpressor	7	28.54	9.97	549.02	382	0.30
HEI10 overexpressor	8	28.67	9.86	392.18	258	0.34
HEI10 overexpressor	9	26.23	8.85	417.09	292	0.30

HEI10 overexpressor	10	27.20	8.91	457.81	251	0.45
HEI10 overexpressor	11	26.13	8.71	690.00	486	0.30
HEI10 overexpressor	12	27.79	9.04	371.30	251	0.32
HEI10 overexpressor	13	27.58	9.60	395.55	241	0.39
HEI10 overexpressor	14	30.64	9.28	363.65	234	0.36
HEI10 overexpressor	15	30.53	9.64	472.53	308	0.35
Mean		27.99	9.37	548.69	364.87	0.34

i3b **$\chi^2(1) = 23575.54$, $p < 2.2 \times 10^{-16}$, *i3c* $\chi^2(1) = 10549.36$, $p < 2.2 \times 10^{-16}$**

Appendix

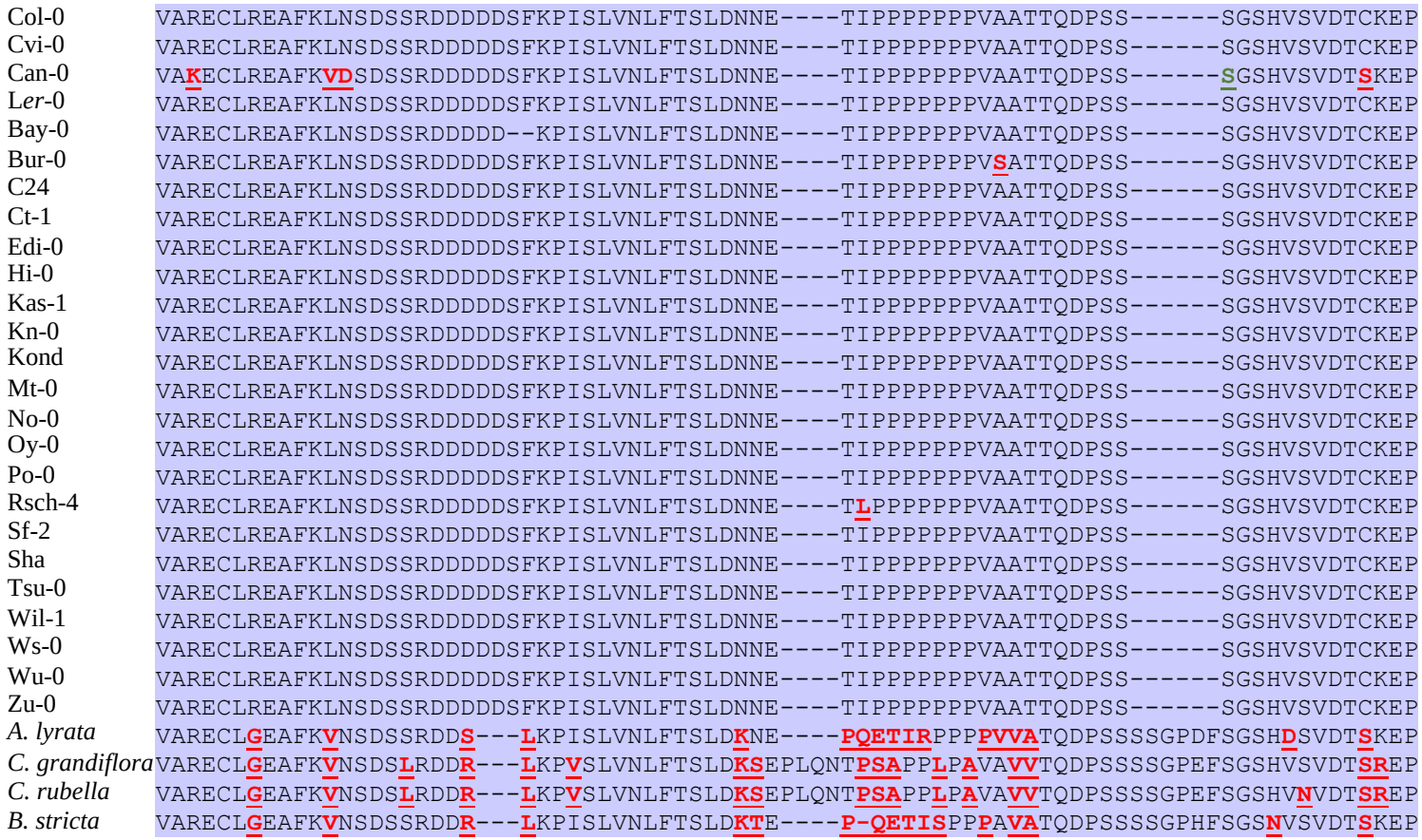
A

Col-0	taacattatgtatgccttg--tggtttaaatatag
Cvi-0	taacattatgtatgcgttt--tggtttaaatatag
Can-0	taacattatgtatgcgttg--tggtttaaatatag
L*er*-0	taacattatgtatgccttg--tggtttaaatatag
Bay-0	taacattatgtatgccttg--tggtttaaatatag
Bur-0	taacattatgtatgccttg--tggtttaaatatag
C24	taacattatgtatgccttg--tggtttaaatatag
Ct-1	taacattatgtatgccttg--tggtttaaatatag
Edi-0	taacattatgtatgccttg--tggtttaaatatag
Hi-0	taacattatgtatgccttg--tggtttaaatatag
Kas-1	taacattatgtatgccttg--tggtttaaatatag
Kn-0	taacattatgtatgccttg--tggtttaaatatag
Kond	taacattatgtatgccttg--tggtttaaatatag
Mt-0	taacattatgtatgccttg--tggtttaaatatag
No-0	taacattatgtatgccttg--tggtttaaatatag
Oy-0	taacattatgtatgccttg--tggtttaaatatag
Po-0	taacattatgtatgccttg--tggtttaaatatag
Rsch-4	taacattatgtatgccttg--tggtttaaatatag
Sf-2	taacattatgtatgccttg--tggtttaaatatag
Sha	taacattatgtatgccttg--tggtttaaatatag
Tsu-0	taacattatgtatgccttg--tggtttaaatatag
Wil-1	taacattatgtatgccttg--tggtttaaatatag
Ws-0	taacattatgtatgccttg--tggtttaaatatag
Wu-0	taacattatgtatgccttg--tggtttaaatatag
Zu-0	taacattatgtatgccttg--tggtttaaatatag
A. lyrata	taacattatgtatggcttgattggtttaaatatag
A. halleri	taacattatgtatggcttgattggtttaaatatag

B

Col-0	taatttaaggtaaggtttgactacttagATG
Cvi-0	taatttaaggcaaggtttggctacttagATG
Can-0	taatttaagttaagatttggctacttagATG
L*er*-0	taatttaaggtaagatttggctacttagATG
Bay-0	taatttaaggtaagatttggctacttagATG
Bur-0	taatttaaggtaaggtttggctacttagATG
C24	taatttaaggtaaggtttggctacttagATG
Ct-1	taatttaaggtaaggtttggctacttagATG
Edi-0	taatttaaggtaaggtttggctacttagATG
Hi-0	taatttaaggtaaggtttggctacttagATG
Kas-1	taatttaaggtaagatttggctacttagATG
Kn-0	taatttaaggtaagatttggctacttagATG
Kond	taatttaaggtaagatttggctacttagATG
Mt-0	taatttaaggtaaggtttggctacttagATG
No-0	taatttaaggtaaggtttggctacttagATG
Oy-0	taatttaaggtaagatttggctacttagATG
Po-0	taatttaaggtaaggtttgactacttagATG
Rsch-4	taatttaaggtaaggtttgactacttagATG
Sf-2	taatttaaggtaaggtttggctacttagATG
Sha	taatttaaggtaagatttgactacttagATG
Tsu-0	taatttaaggtaaggtttggctacttagATG
Wil-1	taatttaaggtaagatttggctacttagATG
Ws-0	taatttaaggtaaggtttggctacttagATG
Wu-0	taatttaaggtaaggtttgactacttagATG
Zu-0	taatttaaggtaaggtttggctacttagATG

Supplementary Figure 1. Multiple sequence alignments of the region around possible causative polymorphisms in *Arabidopsis thaliana* accession alleles and *Brassicacea* orthologs of *ARI7* and *PP2C*, the putative candidates for rQTL2^{420} and rQTL5^{420}. Sequences given for 25 representative *Arabidopsis thaliana* accessions (1001 Genomes Project, Alonso-Blanco *et al*. 2016), and any orthologs identified in *Brassicaceae* (JGI Phytozome BLAST, https://phytozome.jgi.doe.gov/pz/portal.html#!search?show=BLAST, on www.phytozome.jgi.doe.gov, 7th March 2018). **A.** 33bp of sequence from intron 14 of the *ARIADNE7* gene (At2g31510), believed to contain the polymorphism underlying rQTL2^{420}. **B.** 31bp of sequence from the promoter of the *PP2C* gene (At5g53140), believed to contain the polymorphism underlying rQTL5^{420}. The start codon of *PP2C* is denoted by the sequence given in capitalised letters. Polymorphisms differing from the Col-0 reference sequence are highlighted in red. Indel polymorphisms are indicated by dashed lines. *Brassicacea* species used: *Arabidopsis thaliana*, *Arabidopsis lyrata* and *Arabidopsis halleri.*



Supplementary Figure 2. Multiple protein sequence alignments of the region around possible causative polymorphisms in *Arabidopsis thaliana* accession alleles and *Brassicacea* orthologs of *TPR8*, the putative candidate for CanQTL4. Sequences given for 25 representative *Arabidopsis thaliana* accessions (1001 Genomes Project, Alonso-Blanco *et al*. 2016), and any orthologs identified in *Brassicaceae* (JGI Phytozome BLAST). 83 amino acid protein sequence from exon 2 of the *TPR8* gene (At4g08320), believed to contain the polymorphism underlying CanQTL4. Polymorphisms differing from the Col-0 reference sequence are highlighted in red if they denote a non-synonymous polymorphism resulting in an amino acid change, and in green if they denote a synonymous polymorphism that does not result in an amino acid change. Indel polymorphisms are indicated by dashed lines. *Brassicacea* species used: *Arabidopsis thaliana*, *Arabidopsis lyrata, Capsella grandiflora, Capsella rubella* and *Boechera stricta*.

Supplementary Table 1. Genotyping markers used for QTL mapping populations. Chromosome and chromosome position are indicated, as is the closest gene (TAIR10 annotation).

Chromosome	Marker name	Type	Enzyme	Product size in Col-0	Product size in Cvi-0 and Can-0	Forward primer	Reverse primer	Position (start bp) TAIR10	Closest gene
1	1242	SSLP		250	198	CCGGTTTTCGGATATGGTTT	TCGAGTCAATGGAAACTTCAAA	1242104	At1g04555
1	5292	CAPS	HindIII	95+95	190	TGTGCTTGGTTTTGGGTGTA	AGTTGCATTGCACAAAAGAGTT	5292435	At1g15380
1	6108	CAPS	HindIII	77+154	231	CAGTTCACGGGTCCAATACC	CCGGCATAAAACCAAAAAGA	6108789	At1g17750
1	7294	SSLP		199	162	TTCAAAACTGGAGCGTCGTC	GGCCCATCTTGTGTGTTTTG	7294957	At1g20930
1	10443	SSLP		245	141	TTCTCGATGGGATGATAGGTG	CCAATCAATGACCAACAAAAAG	10443987	At1g29830
1	11184	CAPS	HindIII	31+214	245	TTGAGGAATTGCTTCGATCC	GACATTGCTTCCACCAACCT	11184029	At1g31280
1	13278	CAPS	HindIII	23+151	174	TCGCATCTGAGAATTGCTTG	GATTGTTTCAGCAACCACCA	13278246	At1g35780
1	14122	SSLP		239	189	GCTAGCAGTCGAGTATTCTGTCGAG	CGTGTCCCACCATCATCAC	14122817	At1g37100
1	16908	SSLP		110	82	GCACAGAAAGACAAACCCAAAG	CGACCAGCAAGGTTGTTCTTAG	16907783	At1g44770
1	18587	CAPS	HindIII	82+83	165	TGATTGCTCGTAGCATGTGA	AATCTCAAAGACGACGCAAA	18587671	At1g50190
1	19331	CAPS	HindIII	93+72	165	TCGATGACGAATCATACATCAATA	AGGCAAGTCGAAGTTGATGC	19331674	At1g51190
1	19540	SSLP		221	286	GTTCCCCGATTCATGTGAGA	CAAAAAGGGAAAAGCCCACT	19539729	At1g52440
1	19611	CAPS	EcoRI	95+88	183	TGACTTTGCTTCACACCGAGT	CCAGCCAACCAATCAATACC	19611233	At1g52650
1	20074	CAPS	HindIII	65+85	150	CTGCCTACACCGTCATCAAA	TCCTTCTCGCCATCTCAGTT	20074913	At1g53780
1	20907	CAPS	HindIII	110+134	244	TGGCTGTGTTGTTTGGAGTT	AAACCGTGCTTGTCCAAATC	20907282	At1g55915
1	21975	CAPS	HindIII	119+39	158	TTGCTTTTGAATTTATGAGTGGAA	GAATATTTGCCAAGCCATCG	21975829	At1g59750
1	22709	CAPS	HindIII	169+47	216	ATCCCACGAATCGAAATCAG	GCGTTAAAGAGTTGGCATCA	22709104	At1g61560
1	24743	SSLP		234	160	GAGGCACCGAAAATGGATTA	CCAATCGGATTATAGTGTGAACTTT	24743355	At1g66345
1	25323	CAPS	HindIII	73+79	152	AATGCATCCGGTTTACAAGC	ACGCTGCAGAGCTAAGTTCC	25323140	At1g67560
1	26352	SSLP		231	131	CATAAGAGCCCCGATACTACTCA	CAAGGAGATGTTGGGCTTTG	26357768	At1g69980
1	30413	SSLP		135	104	CCAGCCACAGCTTCTTTCTGA	TTGATTGAATAATGGTTCTTGTGATGA	30412519	At1g80950
2	132	SSLP		229	162	TCCAATGGGCCACAAATTAAC	TTTGTGCTTTGATTACTGCAAGTG	132648	At2g01250
2	2346	SSLP		347	261	GGCAAATTTGGTTGGCTCTC	TGTTTTGTGCTATTTGTGTCAACC	2346993	At2g06020
2	4302	SSLP		283	163	CCAACGTCACCTCCTCCTTA	AACGGCTATGACTATCCAATTAAGA	4302041	At2g10921
2	6789	SSLP		112	82	GCGTTTTGTATCATCAAAGGTTCC	CGCAATTTCTCGAACTTCCTTT	6789815	At2g15560

2	7170	SSLP		347	245	TGCAACATCACGGTTTTCTC	GCAATTTTTGGCGTGTGTAG	7170083	At2g16540
2	7852	SSLP		220	170	ATATTTTCGCGGTGTGCAA	TTAGCTTATTTGTCAATCAAAGATCA	7852059	At2g18070
2	8078	SSLP		211	146	CGCTTATATCGTTCGCGTTT	AACACGGCGGTTTTATGAAC	8078853	At2g18620
2	8799	SSLP		155	116	CTCGAGTGATGATCCCGAAT	CTCCGTCGGAATTGACCTTA	8799647	At2g20400
2	9652	SSLP		249	166	GGACGTGGGATGTTTGATGT	CGCCTCTCTCATTCCCTCAT	9652244	At2g22710
2	10268	CAPS	BamHI	160+40	187	TCATTTGAAGCCAGTTGGAA	TTTCCATCTTGGTTGGGTTC	10268940	At2g24160
2	10540	CAPS	EcoRI	90+62	152	AGTGTGTCGCGAAGGTGAG	TGCAACGTCTATTCCCCATA	10540061	At2g24750
2	10694	SSLP		249	82	CGTACCTAGATCGTAGATCAAAAGTG	GCATTACCACTACGCTCACC	10694128	At2g25130
2	10702	CAPS	EcoRI	101+64	165	AAAATCATTCAGGCACCAACT	ATCCATCAATGGAAGGTGGA	10702633	At2g25150
2	10870	CAPS	BamHI	100+29	129	TTTGACCCTGTCCCTTCAAA	TGGGCTTGTGAATGTTGACT	10870606	At2g25540
2	10950	CAPS	EcoRI	122+54	176	AGCGAGAAGGATCAGTTGGA	GCGAAGTGAATGTCGAGGA	10950454	At2g25700
2	11321	CAPS	EcoRI	81+79	160	TGAAACTCCAAACGCATCAG	TTACCAGTGCAAATGCGAAA	11321269	At2g26610
2	11845	SSLP		216	157	TCAATCGATGAAGTTCCGTCT	TGGTTGAGTGCTGAGCAAAT	11845188	At2g27780
2	12121	SSLP		229	157	CAACGAAGAGAGGGAAACAAA	CGTTTCCAAATGACAAATGGT	12121783	At2g28355
2	12399	CAPS	HindIII	93+98	191	TGCATTAGCAGGAGCTCTAGTC	GAGGCCTTTACTCTGGCTCA	12399065	At2g28880
2	12618	CAPS	HindIII	45+177	222	GCAATCTCAATCCTCCGACT	CTTCCGATGTGGGACAAAGT	12618840	At2g29420
2	12768	CAPS	HindIII	181+41	222	AGCGCGATTAGGAGTGTCAT	TGACCAGAGCCATCTAAAAGC	12768352	At2g29960
2	13157	CAPS	NdeI	55+155	210	CTCTTTCACCTGCAAATCCA	CAACGTGCTTCAGTCGATGT	13157576	At2g30920
2	13188	CAPS	XhoI	115+111	226	CAGAGTTTTCTCTGCCACAGG	GCAAAATGAAGATAAGAGGAAGTTG	13188597	At2g30990
2	13195	CAPS	NcoI	156+31	187	GGGAGTTGATGTCCTCATGC	GCTTAATCTGGCCAAGCAAC	13195774	At2g31010
2	13235	CAPS	HindIII	77+166	243	GAGTAACCCAAAACCCACGA	TTGGTAACAAAACGAACAGCTT	13235879	At2g31081
2	13427	CAPS	PstI	140+65	205	GTGGAGATTATTCCGCCAGA	CAAGCAAAACAGCCGTTACA	13427779	At2g31530
2	13521	CAPS	HindIII	139+51	190	CCCAGCATTCCTCAATCAAT	GCCAAGTCCAAAAGTTGTCC	13521040	At2g31800
2	13690	CAPS	HindIII	46+181	227	GCAACTTGCAAGCAACCTTT	AAGGTGAAGAAATTTTTGAACTGTG	13690469	At2g32250
2	13864	CAPS	HindIII	150+50	200	AACCTCAACATTTCCGAACG	CGTACTCACGAGTGGACGAT	13864607	At2g32690
2	13959	CAPS	HindIII	91+114	205	CTGATGCCAAATGTCATTGAA	TTGTTCCCGGAAACTTCATC	13959658	At2g32905
2	14176	SSLP		205	146	TTGAGAGATTTTGCAATAGTAAGCA	CGAATCAATCTTATCAACTCTTCTTG	14176271	At2g33470
2	14232	CAPS	HindIII	74+156	230	GCTGGTTGTTGTGTTGGTTG	ACGCAAAAACCGGCTTAATA	14232466	At2g33620
2	14476	CAPS	BamHI	49+171	220	GCTCTCTTCCAGTTGCGTTT	TGCGACTGAGAACAACCAAG	14476387	At2g34300
2	15445	CAPS	HindIII	101+111	212	GTAACCGGTCACCAAAATGG	TGAAATGAACTCAGCCAACG	15445467	At2g36830

2	15642	CAPS	HindIII	21+147	168	TCGATTTGGTGGTGAATTTG	GATGTGGAGGGAGAAGTGGA	15642866	At2g37250
2	16011	CAPS	BamHI	100+60	161	ACTCTCGCCGACGAAGATAA	ACGGCTTCAACAACGTTACC	16011942	At2g38230
2	16339	SSLP		202	142	CACGAGCAATCCTTGTTTCA	GGGAAAAAGAAAGACCCACA	16339573	At2g39175
2	17165	CAPS	BamHI	57+130	187	TCATCAACGTTGCTCATAAACC	CGTCTTGACCGGTGAGTTCT	17165057	At2g41180
2	18797	SSLP		199	137	TGCTTTATTTGTTTTTGCCATTT	TGTAGGTGTAGATGGGCGTATT	18797975	At2g45630
2	19048	SSLP		306	230	GACCCATATCGTAGGCCACT	TTTTACCAGCCTCCATCGAC	19048891	At2g46410
3	1031	SSLP		419	345	ATGCCTTGGTTTCAATTTGG	TACCCGCTCCTTGACAGTTT	1031481	At3g03980
3	1746	CAPS	HindIII	23+146	169	TGAATGCACAGTCAGAGCTAAA	TTTCTAGAGATTACCTCCCTTTTTGA	1746235	At3g05850
3	2604	CAPS	HindIII	86+109	195	CAGCAGACAAATGGCAAAAT	AGAGTTGCACATCAGTTTGGAA	2604955	At3g08570
3	2718	SSLP		223	185	ACAACTGGGCGACTCACCTT	CGTAAACACAAACTGCGAGGT	2718687	At3g08940
3	3621	CAPS	HindIII	121+31	152	ACGTACCACCATCCCAATGT	TTTTTGTTGTTACCCCCAATG	3621888	At3g11500
3	4126	SSLP		310	169	GGAATAATGGATTCCTCTCTCG	TGTTTGAATGTTGACAATGAGC	4126508	At3g12930
3	4715	CAPS	HindIII	176+22	198	TGGCTTGTAGGTGTTGTTGAA	CAAAGCAGCCATTGATGATG	4715997	At3g14205
3	5419	SSLP		212	152	TTTGCTTTTTCTTTTCCTTTTTG	TTCGCAAAGCTAAGCAACCT	5419342	At3g15980
3	7638	CAPS	HindIII	105+81	186	ACGGCGAGCTAGAAACTGTC	GTTCGAGGTTACCGAGATGG	7638911	At3g21690
3	8140	CAPS	HindIII	123+48	171	TCCTGGATTGTTGCTTCCTC	TTGATCCCTCCCAAATTTACC	8140489	At3g22960
3	8935	CAPS	HindIII	114+36	150	CCAAAACCAACCACTGCTTT	TGAGAAGTCTGGTGAAAGTGGA	8935217	At3g24515
3	9194	SSLP		383	227	TTCCAAATTTTGACCGAGACT	TTCGCTCATTTGGACAGTTG	9194020	At3g25250
3	9404	SSLP		384	297	AACGGTCCAGGTTCCTCCTC	TTGGTTTTAAGGCTCTGGAATCA	9404279	At3g25760
3	10695	SSLP		161	122	GAGGGATGCAAGGAGGATCA	TTCATCACATCAACGCTCCAA	10695968	At3g28540
3	11649	SSLP		228	188	TTTAGCCAAACATGCCCAAAT	CCAAGCGCCAAAACTACCTC	11649496	At3g29770
3	12356	SSLP		455	315	CTACGCCCGGTGTATTTGGA	GCTTGTGAGGCTATGTGGCTTA	12356948	At3g30730
3	15949	SSLP		465	382	CCACCCTCCAGGGAAGAAGT	GGCAGCGACTGGCTTGTTTA	15949551	At3g44250
3	16008	SSLP		237	147	TCCGTTTTTCAGCAATTAGGA	GCCATCCCTACACACAAACC	16008197	At3g44330
3	16679	CAPS	HindIII	122+69	191	ATCCATTAGCAAGGCGATGT	CCGTGAGTTTGGGAATCAAT	16679705	At3g45470
3	17233	CAPS	HindIII	159+31	190	CACAGAACCCAAATCCGAGT	GATTCGCGTCTTGTCTCAGG	17233598	At3g46800
3	18459	CAPS	HindIII	114+42	156	TCAGAGACAGCAGAGGTGTGA	CGAAGATTTGCGAGAGAACA	18459071	At3g49765
3	19064	CAPS	HindIII	124+52	176	TTGTCGCTGAAGTTGGTTTG	AACCACGACGGTTGGATAAA	19604527	At3g52880
3	19165	SSLP		284	234	TACGTCGCCCTCGAAGAAAT	GCGCTACATACGCACCACAT	19165521	At3g51660
3	19604	CAPS	HindIII	124+52	176	TTGTCGCTGAAGTTGGTTTG	AACCACGACGGTTGGATAAA	19604527	At3g52880

3	21008	SSLP		211	172	CCGACGTTGTGTTTCTATTTCC	TGAGGGAACAAGGACCTAACCA	21008135	At3g56710
4	5	SSLP		270	191	ACGAAAGAATGGCTGGGTAA	CGGCTTGAATAGCCCAGATA	2824	At4g00005
4	230	SSLP		267	209	GCGTTCACCTTTAGCATTCCA	GCAGCTACACTCATGCCCTCT	230388	At4g00520
4	1160	SSLP		545	237	CATTACCATTGATGTTGAAGCAA	CACGACCGACCTAAAGCAAT	1159651	At4g02660
4	1642	CAPS	HindIII	112+87	199	TTATGAACCCGCGGTTAAAG	TGCAGTATCATCGACACACG	1642962	At4g03700
4	2173	CAPS	HindIII	240	87+153	GTGGATCCCCAAGATACGTC	CACCTCCAAGGCATCAGAAT	2173145	At4g04405
4	2392	CAPS	EcoRI	175+72	247	AACACCAGCACTCCAAATGTC	CATCAAAGAAGGTGAGCCAAA	2392247	At4g04710
4	2506	CAPS	EcoRI	250	54+196	GTTTTTGGATGGATGAACTATCAC	CATTTCCTCCACAATGATTGAA	2505775	At4g04925
4	2611	SSLP		243	201	CTGGAGGAAAGGTTGGTGAA	TGAGCCTCCTTTCTGATTGA	2611319	At4g05091
4	2618			155	109	CCCAAATTACAAGATTTGTACCG	TGGCTAAAACTCTAGTAACAAAGCTG	2618097	At4g05100
4	2807	SSLP		250	182	CAGCAACGTCTCATGGACAA	GAGGTATGCCTAATCTAAATTCTGGT	2807397	At4g05520
4	3363	CAPS	HindIII	204	158+46	AGGCAAACCAGAACCTCAAA	ATCCTCACGTGGATGAGTCC	3363999	At4g06537
4	4636	SSLP		220	175	ACTCCATTCCATCACAAACG	CCAAAGTCAAATGCAATAATGG	4636841	At4g07820
4	4844	CAPS	BamHI	278	228+50	TTTCTCTTCCACATCCTTTCC	TTAAATGGTGTTGAACGGATTC	4844517	At4g08022
4	4871	CAPS	NdeI	310	225+85	GCAGCCTTTAAGGAGCAGAA	AAACGCCCTCGACCTATTTT	4871784	At4g08033
4	5153	CAPS	NdeI	237+113	350	TGTAGGTTTTATGTTTCGGGTTT	TTTTGGTTCACCGCTAAAAA	5153552	At4g08160
4	5258	SSLP		231	128	AAGCAATGTTGCGTGAAATG	TCCCGAAGGTTGAATCACTT	5258454	At4g08333
4	5742	SSLP		349	198	TCCCCAGTCATCACGATACA	TCGGTCATAGGCCAAAGAAA	5742329	At4g08953
4	6781	CAPS	NdeI	76+76	152	GGATGGGAAGACCTTTGCTT	GTTCCATATCTGACGCAGCA	6781662	At4g11050
4	7157	SSLP		334	262	ACATTAGCGGAGGCCACTT	ATGGGCAAAAGCTTCCAGTA	7157357	At4g11911
4	7807	SSLP		258	201	GATCCCAAACCTTCAGCAAG	GTTTGTGGTCCTGGTGCTTT	7807196	At4g13430
4	7953			331	261	TCGGGAGACTGATGAGAGATG	AGGTAACAAACCCGGTCGAT	7953372	At4g13690
4	8198	CAPS	HindIII	81+163	244	ACAAAACCGAACCCCAAAAG	GCTCTGAACAATGCAGGATG	8198181	At4g14220
4	9210	CAPS	HindIII	86+134	220	ATGAGGAGACCGGGGTAAGT	AGTTGGGAACCCTGTTCCTT	9210030	At4g16280
4	9652	SSLP		234	172	GTTGCCCACTTGTGTGGTCT	TCTTGTTTGGATGTGAAATTGGA	9652287	At4g17200
4	10721	SSLP		323	188	ACAATTTTTAGTCTGTCTAGCGTGA	CGAAATGCAGTTCACATCGT	10721530	At4g19710
4	11124	CAPS	HindIII	116+41	157	TCGATAACCACTTAATTGTTGAGA	CCAAATGTCTCATCTCGTCGT	11124218	At4g20740
4	11352	SSLP		164	127	AGAGGCTTGAAGCAAAACCA	TTACGTGCGCATATCTGACC	11352024	At4g21340
4	12198	CAPS	SpeI	206	108+98	TGTGTCCTCTTACTCCAGCAAA	CCCGAGACCTTTTCTCACAG	12201331	At4g23350
4	14139	SSLP		240	174	GCCTAGCTGCATCGTACACC	TGCAGATGACATTTTACACTCCA	14139014	At4g28630

4	17158	SSLP		214	151	TCGATGCTTATGCTATTTGAGTAGA	TACCGGAAAAAGTGGCAGTC	17158265	At4g36260
4	18510	SSLP		215	161	TGACGGCAGATTCAGAGAGA	AGGGAGGACGAAGAATGAGG	18510489	At4g39900
5	1414	CAPS	HindIII	113+86	199	GAGATAGAGAGAGAAAAGGACGGTAA	AAGAAAAGAAGGTAATCACACAACG	1414820	At5g04860
5	2628	CAPS	HindIII	111+43	154	TGCCATAATGCAAGCAAAGT	ATGAATCCTGGCCGTTGATA	2628593	At5g08170
5	3750	SSLP		137	97	ATGGTGGACCTGGGGGTAAC	GCATGTAGGAAACACAAATCCTGA	3750331	At5g11660
5	7064	SSLP		267	220	ACTGGCCTCGCCTTTCACTA	AATCACAACTGTGCCCTCGTT	7064379	At5g20840
5	8569	CAPS	HindIII	61+156	217	ATTTGCCACAGATCCCAAAA	GTCGCTATTTCGTGGAAACG	8569949	At5g24910
5	9437	CAPS	HindIII	128+62	190	TCGTTTCACAACTTCTTCTTCG	CGGATCGTGTGGAAGAGACT	9437511	At5g26820
5	9566	SSLP		346	242	TCAGCTCGTCAAAACAAGTACAA	GCAAGGCTTATGGATGCACT	9566527	At5g27180
5	14866	CAPS	HindIII	27+181	208	CTGCTTTGAGTAGCCCGATA	GCGATTTCTTGGATTATGCTG	14866728	At5g37475
5	15128	SSLP		350	198	CGTGGATTCACCATGAAATG	TTCCGAAGAATTTGGAATGG	15128892	At5g37980
5	16542	CAPS	HindIII	165+21	186	GAGATGTGGGAAAGGCAAAC	CTCTCGGGGTTCCAACTACA	16542416	At5g41350
5	19994	SSLP		169	109	TCTAAACCGAACTAAACCGTGAA	CAAACCAAAACCTACTTTTTCCAA	19994907	At5g49320
5	20312	CAPS	HindIII	173+22	195	TCAAATGAATCAGGGGATTTG	CGATTCGACACTGAATCACG	20312608	At5g49930
5	20437	CAPS	NcoI	29+142	171	CTACGCTCCACCTAGCCAAG	GCACGAACGAAGTAGGTTCC	20437184	At5g50200
5	20685	CAPS	EcoRI	61+149	210	CCGAGTTAAGCCCAATTTGA	TCAATTCGAACCGAAAGCAT	20685339	At5g50830
5	20780	CAPS	HindIII	45+197	242	GTGCGCAGCTTCACATTAAC	CCACCTGAAGGGTCTTCATT	20780202	At5g51120
5	20899	CAPS	EcoRI	122+127	249	CCACGCCATAACTGAAACCT	CTAAACCGGCAAGCAAATTC	20899170	At5g51451
5	21349	CAPS	EcoRI	109+41	150	TGTATTTTGGATTTTGGTTCCAG	TCGTCCAGCCTTTTAGTTGC	21349815	At5g52630
5	21562	CAPS	NdeI	76+174	250	TCAAGGTCGAAGCGAAAACT	CAGGAAGGCGAGAAATTTGA	21562910	At5g53160
5	21906	CAPS	NdeI	45+135	180	TGCTTTCGAACACCAAAATG	TGGCATTTGTTCATTTGCTT	21906244	At5g53960
5	22313	CAPS	XhoI	50+200	250	ACGGATTCCACCGCTACTTA	GCAGCGGCTAATTTCTTCCT	22313360	At5g54970
5	22402	CAPS	HindIII	110+56	166	GGGTCCCACACTCACCTCT	GGGTTTAAAATGGGTTTCTCTTT	22402070	At5g55230
5	23058	SSLP		242	176	TGGAGTATAAGTTTTCAAGAACAGC	CCGAATGCCAAAGAAAAATC	23058119	At5g56980
5	23186	CAPS	EcoRI	101+74	175	TAACGAGCAGCCATGTTTGA	AGGCAAATGTACCCACCTTG	23186865	At5g57220
5	23875	CAPS	HindIII	78+112	190	CTGTCGATGACGAACTCGAA	TAATGGGCCTCGTTTGAAAT	23875653	At5g59140
5	24192	CAPS	HindIII	102+122	224	TCAAATGGGATCAAAAACAACA	GGATTCGAGTTCCACGAGAA	24192726	At5g60070
5	25212	SSLP		159	123	GCGGTGGCAGTAGGTTAAAA	TCTAATACCGGCAATAAAACTTGA	25212874	At5g62780
5	25340	SSLP		240	190	CATGCGTATATACATGTGATAATGTGA	TTTTTGATGCTGACTTTTCAAAC	25340517	At5g63180
5	26907	SSLP		270	200	TGTGGATCTTTATGACGTGTGC	ACCATCTACTTCCATTCAAATAACG	26907352	At5g67420

Supplementary Table 2. T-DNA insertion genotyping primers.

Oligo name	Oligo sequence (5' to 3')
SALK_040837 LP	TCCTTCAGTTGATGCATACCC
SALK_040837 RP	AGAAGCAGTGGCTTCTCATTG
SALK_140743 LP	TCATTCGAATTTCGGGTACAC
SALK_140743 RP	TCAATGCCCTAACAAAACCAG
SALK_126570 LP	GCTTGTTCACGAAGAAGTTGG
SALK_126570 RP	CGCTTCAAAAGCTGAATTGTC
GK_804H10 LP	CAGTTAAGTTATGCGCGAAGG
GK_804H10 RP	GCCTTTGGTAACAGAATGCTG
GK_219G07 LP	CACAGATTCGTTATGCGGATC
GK_219G07 RP	CAATGAAACCATACCACCACC
GK_129F04 LP	CCATGAAGCATTTTCCTAAGG
GK_129F04 RP	TTTGACCGTATTTACGCCAAG
GK_129F04 RP2	cgccaagcccacaaatcagaag
GK_413F08 LP	CGATTACGAGTTTCCTCGTTG
GK_413F08 RP	CTCCAGGCTCTTGATTGTTTG
SALK_090163 LP	TTTGATGTTCTTCGATCGACC
SALK_090163 RP	ATTCAACATTTTCTTGCTCCG
SALK_053625 LP	AGCAATCATGGTACCATCTGC
SALK_053625 RP	TCCGAAAGTGATTGATCAAGG
SALK_048952 LP	CAAGTGCTCTCTCTCCCACAC
SALK_048952 RP	ACGTGGATGAGTGTGTTCTCC
SALK_010368 LP	GAATGCCAAATCTTATGCGAG
SALK_010368 RP	GCTGAACGAATCAGCTTTTTG
SAIL_369C12 LP	GCAGAAGCTTCTGAGGAAAGAG
SAIL_369C12 RP	GCCATTCAAAGATCAATCCTC
SAIL_1170H08 LP	TTGCTTGGTTCCAAGAGCTAC
SAIL_1170H08 RP	TGTGTGATCTCCCCAGAAAAG
SALK_093558 LP	TCAATGCCCTAACAAAACCAG
SALK_093558 RP	AGCTTGCACAAATCAGTGAGG
SALK_076600 LP	CAATGCATTCAGCACATCATC
SALK_076600 RP	CGATTCTGAGACGATTTCGAG
SALK_146231 LP	AATCGAGAAACGTGATTGGTG
SALK_146231 RP	TTATCATTTCTTCGTTTCGCC
SALK_047663 LP	GGTGGTGGTATGGTTTCATTG
SALK_047663 RP	TCACCGTCCTTCAAATCAATC
SALK_120118 LP	GACACAATCGCAAACTGAAAAG
SALK_120118 RP	GAGATTGCACAAGAAGAACGG
SALK_014248 LP	CTCCACGCTTACACATTCCTC
SALK_014248 RP	AAAAATCATGGTTGCGATGAC
SAIL_731H04 LP	GTTGGTTTGGCTTTTCAGTTG
SAIL_731H04 RP	GGAAAATATGCCGAGGCTATC
SALK_089877 LP	GACTCAAAGCCATCTTCATCG
SALK_089877 RP	GCCATATGTGTGGTTCTCTGG
SALK_137974 LP	TGAGCTGTTTTCATCCTCTCG

SALK_137974 RP	ATACATTGGCACGACCTCTTG
SALK_109522 LP	TCGACATGCTTATCACGATTG
SALK_109522 RP	TCCGCACAAAAGGTACAATTC
SALK_014797 LP	TTCTCGTACCTGCACAAATCC
SALK_014797 RP	TGGAAACGACGAACCTTAATAAG
SALK_111410 LP	AGATGGACATTCGCGTAATTG
SALK_111410 RP	CACACACACAAATTTGCAAGG
SALK_145153 LP	GGCACAAAAGGAGGAAAGAAC
SALK_145153 RP	AGATGGACATTCGCGTAATTG
SALK_024459 LP	TTATGTGGGAACAGCTGAAGC
SALK_024459 RP	TTCCTCTCATTTCCTTCTCCC
GK-025G08 LP	AATGCCGATGCTGTTGAATAC
GK-025G08 RP	TCTTTTGTGATAACCCAACGC
GK-380E06 LP	ACTCCTCAAGGAGTGAAAGGC
GK-380E06 RP	ACTCCATCATTGCAGATCTGC
SALK_054181 LP	TTTCTGCTCTTTTGGCTTGAG
SALK_054181 RP	GCCATTGTTGCAAAGCTACTC
SALK_082541 LP	CAGATCAGCCATAGCCTTTTG
SALK_082541 RP	TTCGATACCTTTGCATTGAGC
SALK_027620 LP	ATCGGCATATGCATTTGAAAC
SALK_027620 RP	AATTCGGGTTTTGTACCTTGG
SALK_018273 LP	ATAGGCTTCTCTGATTTGGCC
SALK_018273 RP	GAACTCGGAAGACCTCAATCC
SALK_050436 LP	AAAAGGTTGAAAGCAAAAGGG
SALK_050436 RP	GATGTTTGTACCACCACCACC
SALK_027514 LP	TCGAGATCCTTTGGTATCACTG
SALK_027514 RP	ATTTCACCACGACTCATTTCG
GK-345C06 LP	TTGGGAAGAGAAGCCATGTAG
GK-345C06 RP	TGGGGTGTTTCTGAAGACAAC
GK-523B12 LP	AAGTTGTTGGTGGTCGTGAAG
GK-523B12 RP	TGCAAAAAGAGCATGTGTTTG
GK-699D09 LP	GGATGACGAGGTCTAAGCTCC
GK-699D09 RP	TAACATGGAACTGGCAGAACC
SALK_002393 LP	TGGGTTGGGAATGTTTATCAG
SALK_002393 RP	AATGCAAGTAGCAAACAACGC
SALK_064633 LP	GAGCAAGAGGTAACTGAGGGG
SALK_064633 RP	GGCGGTGTTACTTTACACGTG
SALK_059528 LP	AATCACCGCATTTCAAACAAG
SALK_059528 RP	CCCTAAATTTCTGCCAAAACC
SALK_038792 LP	ATGTTATCATGCACCGAGTCC
SALK_038792 RP	AATTTTACCAGATGGGCCATC
SALK_026225 LP	GCAATTGTGTAAGAAACGTCG
SALK_026225 RP	GTCATCACATCGTCGATGTTG

Supplementary Table 3. *HEI10* Cloning primers

Primer	Sequence
HEI10-XbaI	5' AATCTAGACTGGAATCAACAACGCAGTG 3'
HEI10-BamHI	5' TTGGATCCTAAGCCTTCAATGAACATCAC 3'

Supplementary Table 4. *HEI10* sequencing primers. Position on chromosome 1 given in basepairs (bp)

Primer	Sequence	Position (bp)
HEI10_243_F	taaaaagcttacagcatttgaatc	19961772
HEI10_770_F	tggtagatgaatgttccaatcaa	19962300
HEI10_1274_F	aaaaaccgagttacatacgttttcta	19962801
HEI10_1743_F	ttcacccaatgacacgaaaa	19963276
HEI10_2244_F	GTACCTCCGCGAAGAGACTG	19963777
HEI10_2741_F	tgctactgctgattatctctgga	19964271
HEI10_3254_F	gcccaatttctagggtttcc	19964787
HEI10_3772_F	AAGATGCCAGCAAGATTCTCA	19965304
HEI10_4256_F	aaggttcaaaaacctcctgttt	19965787
HEI10_4739_F	GACACCAAGAACCCGACTTT	19966272
HEI10_5256_F	tttgttgcttctattctcagGCTA	19966785

Glossary

Term	Definition
Accessions	*Genetically diverse inbred lines from natural Arabidopsis thaliana populations*
Allopolyploidy	*A polyploid formed from the union of two separate chromosome sets from different sources*
Anaphase I	*An intermediate stage of nuclear division in the first meiotic division, during which chromosomes are pulled to the poles of the cell*
Autopolyploidy	*A polyploid formed from the doubling of a single genome*
Backcross	*To cross a hybrid to one of its parents, or an organism with the same genetic characteristics as one of the parents*
Balancing selection	*Natural selection that results in a stable intermediate equilibrium of allele frequencies*
Bivalent	*Two homologous chromosomes paired at meiosis*
CentiMorgan	*The unit of genetic distance between two linked loci where one percent of the products of meiosis are recombinant*
Centromere	*A specialised region of repeat DNA on each eukaryotic chromosome that acts as a site for the binding of kinetochore proteins*
Chiasma (plural chiasmata)	*The site of crossing over; a cross shaped structure commonly observed between non-sister chromatids in meiosis*
Chromatin	*The substance of chromosomes, containing DNA and chromosomal proteins*
Chromosome axis	*Structural core of proteins at the base of the chromatin loops of each chromatid; forms lateral elements of the synaptonemal complex*
***Cis*-acting**	*Affects only nearby DNA on the same molecule*
CO interference	*A measure of the independence of crossovers from each other, calculated by subtracting the coefficient of coincidence from 1*
Co-chaperone	*A protein that assists in the correct folding of newly synthesised proteins*
Complementation	*The production of a wild type phenotype when two different mutations are combined in a diploid*
Diploid	*A cell having two chromosome sets; an organism having two chromosome sets in each of its cells*
Directional selection	*Selection that changes the frequency of an allele in a constant direction, either toward or away from fixation*
DNA heteroduplex	*A DNA double helix formed by annealing single strands from different sources, which may loop or buckle if the strands differ; produced in the vicinity of a chiasma from both parental DNAs*
DNA methylation	*Modification of the cytosine base of DNA by addition of a methyl group*

Double Holliday Junction	*The two single-stranded crossovers in close proximity produced during homologous recombination, according to the DSB model*
EMS	*Ethyl methanesulfonate, alkylating agent that induces chemical modification of nucleotides, resulting in mispairing and base changes*
Endogenous	*A DNA sequence that is usually present in an organism i.e. not a foreign sequence from another organism introduced by transgenics*
Epigenetic	*Heritable modifications that leave the DNA sequence unchanged e.g. DNA methylation or histone acetylation*
epiRIL	*epigenetic Recombinant Inbred Lines; lines derived from epigenetic mutants that contain different patterns of stable epigenetic modifications in the same genetic background*
Epistasis	*Situation where the differential phenotypic expression of a genotype at one locus depends on the genotype at another locus*
Euchromatin	*A less condensed chromosomal region*
Eukaryotes	*An organism whose cells contain a nucleus*
Exon	*Non-intron sections of the coding sequence of a gene, corresponding to the mRNA*
Flow cytometry	*Technique used to analyse the physical and chemical properties of particles in a fluid as it passes through a laser. Fluorescently labelled cell components are excited by the laser and emit light at varying wavelengths*
FTL	*Fluorescent Transgenic Lines, which carry two or more fluorescence transgenes defining intervals of the genome. Fluorescence segregation in the products of meiosis from a hemizygote can be used to measure the rate of meiotic recombination between the transgenes*
Gamete	*A specialised haploid cell that fuses with a gamete of the opposite sex or mating type to form a diploid zygote during fertilisation*
Gel electrophoresis	*A method of molecular separation in which DNA/RNA/proteins are separated in a gel matrix according to molecular size, with the use of an electrical field to draw the molecules through the gel in a predetermined direction*
Gene conversion tracts	*The section of DNA that undergoes conversion during meiosis when one allele directs the conversion of a partner allele into its own form*
Gene silencing	*A gene that is not expressed due to epigenetic regulation*
Gene synteny	*Where genes are arranged in a similar order or orientation in a block in different species*
Genetic heterogeneity	*The production of the same phenotype from different alleles of the same loci, or mutations at different loci*
Genetic linkage	*The association of genes on the same chromosome*
Genotype	*The specific allele composition of a cell, commonly referring to a specific locus*

Haploid	*A cell having one chromosome set, or an organism composed of such cells*
Haploinsufficiency	*Describes a gene that, in a diploid cell, cannot promote wild type function with only one copy (dose)*
Haplotype	*A genetic class described by a sequence of DNA or of genes that are together on the same physical chromosome*
Helitron	*DNA transposon, transposes via rolling-circle replication*
Hemizygous	*A gene present in only one copy in a diploid organism, e.g. an X-linked gene in a male mammal, or a transgenic insertion*
Heritability	*Broad heritability: the proportion of total phenotypic variance at the population level that is contributed by genetic variance. Narrow heritability: The proportion of phenotypic variance that can be attributed to additive genetic variance*
Heterochromatin	*Densely staining condensed chromosomal regions, characterised by histone modifications and DNA methylation*
Heterozygous	*A gene pair containing two different alleles in a single individual or cell*
Histone	*A type of basic protein that forms the unit around which DNA is coiled in the nucleosomes of eukaryotic chromosomes*
Histone modification	*Post-translational modification (addition of a chemical group, usually methyl, phosphoryl or acetyl) of the tail of a histone that protrudes from the core nucleosome; information required for correct chromatin assembly*
Homeologous chromosomes	*Partly homologous chromosomes, usually indicating some ancestral homology; found in polyploids and hybrids*
Homologous chromosomes	*Chromosomes that pair with each other at meiosis and contain the same genes*
Homologous recombination (crossover)	*The exchange of corresponding chromosome parts between homologs by breakage and reunion*
Homology	*State of having the same or similar sequence or structure*
Homozygous	*State of carrying an identical pair of alleles at one locus*
Hotspot	*A locus or region of the genome that displays a greater than average frequency of meiotic recombination*
Hypomorph	*A mutation that causes a partial loss of gene function*
Inter-homolog recombination	*Crossover between homologous chromosomes*
Inter-sister repair	*Repair of a double-strand break using the sister chromatid as a template; non-crossover*
Introgression	*The transfer of genetic information from one organism to another as a result of hybridisation between them and repeated backcrossing*
Intron	*A segment of largely unknown function within a gene that is initially transcribed, but the transcript is not found in the functional mRNA*
Inversion	*A chromosomal mutation consisting of the removal of a chromosome segment, its rotation through 180°, and its reinsertion in the same location*

Isolation By Distance	*The accumulation of local genetic variation under geographically limited dispersal; determines the distribution of gene frequencies over a geographic region*
Kinetochore	*A complex of proteins to which a nuclear spindle fiber attaches during nuclear division*
Linkage Disequilibrium	*Deviation in the frequency of haplotypes in a population from the frequency expected if the alleles at different loci are associated at random*
LOD	*Logarithm (base 10) of odds; statistical test used for linkage analysis, compares the likelihood of obtaining the test data if the two loci are linked to the likelihood of observing the same data by chance*
Meiosis	*Two successive nuclear divisions (with corresponding cell divisions) that produce gametes or spores that have one-half of the genetic material of the original cell*
Metaphase I	*An intermediate stage of nuclear division in the first meiotic division, during which chromosomes align along the equatorial plane of the cell*
Microtubules	*Part of the cytoskeleton, consisting of polymerised tubulin subunits forming a hollow tube*
Mitosis	*Somatic nuclear division (during cell division) that produces two daughter nuclei identical to the parent nucleus*
MMR system	*A system for repairing damage (mismatched nucleotides) to DNA that has already been replicated*
MNase	*Micrococcal nuclease; endo-exonuclease that preferentially digests DNA not protected by histone proteins*
Mullers ratchet	*Process by which the genomes of an asexual population accumulate deleterious mutations in an irreversible manner; also observed in regions of genomes which do not undergo recombination in sexual organisms*
Multivalent	*When more than two chromosomes are connected by synapsis or chiasmata during meiosis*
Mutagenesis	*Treatment of organisms with a mutagen to increase the mutation rate, and examination of progeny for specific mutant phenotypes*
NIL	*Near Isogenic Line; lines that are almost genetically identical, differing at only one locus or very few loci*
Non-crossover	*Resolution of a DSB intermediate without exchange of flanking segments of the genome*
Non-synonymous	*Mutational replacement of an amino acid with one having different chemical properties*
Nucleosome	*The basic unit of eukaryotic chromosome structure; a ball of eight histone molecules wrapped in two coils of DNA*
Phenotype	*The detectable outward manifestations of a specific genotype*
Phosphorylation	*The process of adding a phosphoryl group to a molecule*

Poisson distribution	*A mathematical expression giving the probability of observing various numbers of a particular event in a sample when the mean probability of an event on any one trial is very small*
Promoter	*A regulator region that is a short distance from the 5' end of a gene and acts as a binding site for RNA polymerase*
Prophase I	*The early stage of nuclear division during the first meiotic division, during which chromosomes condense and become visible*
QTL	*A gene affecting the phenotypic variation in a continuously varying trait*
Recessive	*An allele whose phenotypic effect is not expressed in a heterozygote; the phenotype of a homozygote for the recessive allele*
Recombinant	*Refers to an individual organism or cell having a genotype produced by recombination*
Relict	*Arabidopsis thaliana accessions with extreme pairwise sequence divergences from the European accessions, believed to reflect isolated populations that survived climate change in glacial refugia*
Retrotransposon	*Transposable elements that replicate via RNA intermediates*
RIL	*Recombinant Inbred Line; inbred lines generated by self-fertilisation from hybrids that have undergone recombination; organisms with chromosomes that incorporate an essentially permanent set of recombination events between chromosomes inherited from two or more inbred strains*
RT-PCR	*Reverse Transcription Polymerase Chain Reaction; technique used to detect RNA expression*
Semi-dominant	*Phenotype of the heterozygous genotype is distinct from, and often intermediate to, the phenotypes of the homoygous genotypes*
Sexual Dimorphism	*The difference in morphology or phenotype between males and females of the same species, caused by inheritance of a sexual pattern in the genetic material*
Sister chromatids	*Juxtaposed pair of chromatids arising from replication of a chromosome*
Site-directed mutagenesis	*Alteration of a specific part of a cloned DNA segment followed by reintroduction of the modified DNA back into an organism for assay of the mutant phenotype or protein*
Splicing	*A reaction that removes introns and joins together exons in RNA*
SUMOylation	*Post-translational modification of proteins by addition of a SUMO protein to an acceptor lysine, involved in many cellular processes inlcluding protein stability, response to stress and progression through the cell cycle*
Supergene	*Group of neighbouring genes on a chromosome which are inherited together due to close genetic linkage and are functionally related in an evolutionary sense*

Synapsis	*Close pairing of homologs at meiosis; linkage of chromosomes by the synaptonemal complex*
Synaptonemal Complex	*A structure that unites homologs during prophase of meiosis*
Synergism	*A feature of eukaryotic regulatory proteins, where transcriptional activation mediated by the interaction of several proteins is greater than the sum of the effects of the proteins taken individually*
Synonymous	*A mutation that changes one codon for an amino acid into another codon for the same amino acid*
T-DNA	*A part of the Ti plasmid that is inserted into the genome of the host plant cell; referring to an insertion of transgenic DNA in the genome*
Telomere	*The end of a chromosome, consisting of repeat sequences that protect the chromosome from deterioration*
Telophase I	*The late stage of nuclear division during the first meiotic division, when daughter nuclei reform*
Terminator	*A regulator region at the 3' end of a gene that acts as a site for the release of RNA polymerase from DNA, and termination of transcription*
***Trans*-acting**	*A diffusible regulatory molecule that binds a cis-acting element*
Transcription factor	*A protein that binds to a cis-acting regulatory element (e.g. an enhancer) and thereby, directly or indirectly, affects initiation of transcription*
Transformation	*The directed modification of a genome by the external application of DNA from a cell of a different genotype*
Translocation	*The relocation of a chromosomal segment to a different position in the genome*
Transposable element	*A general term for any genetic unit that can insert into a chromosome, excise, and reinsert elsewhere; includes insertion sequences and transposons*
Ubiquitination	*The process of adding ubiquitin chains to a protein targeted for degradation*
Univalent	*A single unpaired meiotic chromosome*
Zinc finger	*Protein structural motif, typically functions as an interaction module that binds DNA/RNA/proteins, with variations in structure providing binding specificity*
ZMM pathway	*A pathway for crossover formation that relies on a group of proteins collectively referred to as ZMMs*

Definitions adapted from those given in Introduction to Genetic Analysis (9th Edition), the Oxford English Dictionary, Kleckner (2006), Kim *et al*. (2006), Melamed-Bessudo *et al*. (2012), Shapiro (2003), Francis *et al.* (2007), Glazier (2002), Choi *et al*. (2017), Lichten *et al*. (1995), Sharbel *et al.* (2000), Broman and Sen (2009), Hadany and Comeron (2008), Lawrence *et al.* (2017), Alonso-Blanco and Koornneef (2000), Youds and Boulton (2011), Alonso-Blanco *et al*. (2016) and Mercier *et al*. (2015).

References

1. Abiola, O. et al. (2003) 'The nature and identification of quantitative trait loci: a community's view.', Nature reviews. Genetics, 4(11), pp. 911–6. doi: 10.1038/nrg1206.
2. Acquaviva, L. et al. (2013) 'The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination.', Science (New York, N.Y.), 339(6116), pp. 215–8. doi: 10.1126/science.1225739.
3. Agrawal, A. F. (2009) 'Spatial heterogeneity and the evolution of sex in diploids.', The American naturalist. The University of Chicago Press , 174 Suppl 1(S1), pp. S54-70. doi: 10.1086/599082.
4. Alani, E., Reenan, R. a G. and Kolodner, R. D. (1994) 'Interaction between mismatch repair and genetic recombination in Saccharomyces cerevisiae', Genetics, 137(1), pp. 19–39.
5. Allard, R. W. (1963) 'Evidence for Genetic Restriction of Recombination in the Lima Bean.', Genetics. Genetics Society of America, 48(10), pp. 1389–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17248187 (Accessed: 18 September 2016).
6. Allers, T. and Lichten, M. (2001) 'Differential Timing and Control of Noncrossover and Crossover Recombination during Meiosis', Cell, 106(1), pp. 47–57. doi: 10.1016/S0092-8674(01)00416-0.
7. Alonso-Blanco, C. et al. (2003) 'Analysis of Natural Allelic Variation at Seed Dormancy Loci of Arabidopsis thaliana', Genetics, 164(2). Available at: http://www.genetics.org/content/164/2/711 (Accessed: 6 September 2017).
8. Alonso-Blanco, C. et al. (2016) '1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana', Cell, 166(2), pp. 481–491. doi: 10.1016/j.cell.2016.05.063.
9. Alonso-Blanco, C. and Koornneef, M. (2000) 'Naturally occurring variation in Arabidopsis: an underexploited resource for plant genetics', Trends in Plant Science, 5(1), pp. 22–29. doi: 10.1016/S1360-1385(99)01510-1.
10. Alonso, J. M. et al. (2003) 'Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana', Science, 301(5633). Available at: http://science.sciencemag.org/content/301/5633/653 (Accessed: 13 August 2017).
11. Alves, I. et al. (2017) 'The impact of recombination on human mutation load and disease.', Philosophical transactions of the Royal Society of London. Series B, Biological sciences. The Royal Society, 372(1736), p. 20160465. doi: 10.1098/rstb.2016.0465.
12. Anderson, L. K. et al. (2014) 'Combined fluorescent and electron microscopic imaging unveils the specific properties of two classes of meiotic crossovers.', Proceedings of the National Academy of Sciences of the United States of America, 111(37), pp. 13415–20. doi: 10.1073/pnas.1406846111.
13. Andronic, L. (2012) 'Viruses as triggers of DNA rearrangements in host plants', Canadian Journal of Plant Science. Agricultural Institute of Canada, 92(6), pp. 1083–1091. doi: 10.4141/cjps2011-197.
14. Arends, D. et al. (2010) 'R/qtl: high-throughput multiple QTL mapping.', Bioinformatics (Oxford, England), 26(23), pp. 2990–2. doi: 10.1093/bioinformatics/btq565.
15. Arroyo, M. T. K. (1973) 'CHIASMA FREQUENCY EVIDENCE ON THE EVOLUTION OF AUTOGAMY IN LIMNANTHES FLOCCOSA (LIMNANTHACEAE)', Evolution. Wiley/Blackwell (10.1111), 27(4), pp. 679–688. doi: 10.1111/j.1558-5646.1973.tb00715.x.
16. Auton, A. et al. (2013) 'Genetic Recombination Is Targeted towards Gene Promoter Regions in Dogs', PLoS Genetics. Edited by I. Henderson. Public Library of Science, 9(12), p. e1003984. doi: 10.1371/journal.pgen.1003984.

17. Baduel, P. et al. (2016) 'HABITAT-ASSOCIATED LIFE HISTORY AND STRESS-TOLERANCE VARIATION IN ARABIDOPSIS ARENOSA.', Plant physiology, p. pp.15.01875-. doi: 10.1104/pp.15.01875.
18. Barth, S. et al. (2001) 'Influence of genetic background and heterozygosity on meiotic recombination in Arabidopsis thaliana', Genome. NRC Research Press Ottawa, Canada, 44(6), pp. 971–978. doi: 10.1139/g01-094.
19. Barton, N. H. and Charlesworth, B. (1998) 'Why sex and recombination?', Science (New York, N.Y.). American Association for the Advancement of Science, 281(5385), pp. 1986–90. doi: 10.1126/SCIENCE.281.5385.1986.
20. Barton, N. H. and Otto, S. P. (2005) 'Evolution of Recombination Due to Random Drift'. doi: 10.1534/genetics.104.032821.
21. Baudat, F. et al. (2000) 'Chromosome Synapsis Defects and Sexually Dimorphic Meiotic Progression in Mice Lacking Spo11', Molecular Cell, 6(5), pp. 989–998. doi: 10.1016/S1097-2765(00)00098-8.
22. Baudat, F. et al. (2010) 'PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice.', Science (New York, N.Y.), 327(5967), pp. 836–40. doi: 10.1126/science.1183439.
23. Baudat, F., Imai, Y. and de Massy, B. (2013) 'Meiotic recombination in mammals: localization and regulation', Nature Reviews Genetics. Nature Research, 14(11), pp. 794–806. doi: 10.1038/nrg3573.
24. Baudat, F. and de Massy, B. (2007) 'Cis- and Trans-Acting Elements Regulate the Mouse Psmb9 Meiotic Recombination Hotspot', PLoS Genetics. Public Library of Science, 3(6), p. e100. doi: 10.1371/journal.pgen.0030100.
25. Baudat, F. and Nicolas, A. (1997) 'Clustering of meiotic double-strand breaks on yeast chromosome III', Genetics, 94, pp. 5213–5218. Available at: http://www.pnas.org/content/94/10/5213.full.pdf (Accessed: 30 September 2017).
26. Bauer, E. et al. (2013) 'Intraspecific variation of recombination rate in maize', Genome Biology. BioMed Central, 14(9), p. R103. doi: 10.1186/gb-2013-14-9-r103.
27. Bell, G. The Masterpiece of Nature: The Evolution and Genetics of Sexuality 1st Edition. London: *Croom Helm Ltd*, 1982. Print
28. Bentsink, L. et al. (2000) 'Genetic Analysis of Seed-Soluble Oligosaccharides in Relation to Seed Storability of Arabidopsis', Plant Physiology, 124(4). Available at: http://www.plantphysiol.org/content/124/4/1595.short (Accessed: 6 September 2017).
29. Berardini, T. Z. et al. (2015) 'The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome', genesis, 53(8), pp. 474–485. doi: 10.1002/dvg.22877.
30. Berchowitz, L. E. et al. (2009) 'A positive but complex association between meiotic double-strand break hotspots and open chromatin in Saccharomyces cerevisiae', Genome Research. Cold Spring Harbor Laboratory Press, 19(12), pp. 2245–2257. doi: 10.1101/gr.096297.109.
31. Berchowitz, L. E. and Copenhaver, G. P. (2007) 'Fluorescent Arabidopsis tetrads: a visual assay for quickly developing large crossover and crossover interference data sets', www.natureprotocols.com, 3(1). doi: 10.1038/nprot.2007.491.
32. Berchowitz, L. E. and Copenhaver, G. P. (2010) 'Genetic interference: don't stand so close to me.', Current genomics, 11(2), pp. 91–102. doi: 10.2174/138920210790886835.
33. Berkemeier, J., & Linnert, G. (1987). Effects of high and low temperature on meiotic recombination in normal and oligochiasmatic lines of Vicia faba. *Biologisches Zentralblatt*, *106*(2), 219-230.

34. Bhalla, N. et al. (2008) ‘ZHP-3 Acts at Crossovers to Couple Meiotic Recombination with Synaptonemal Complex Disassembly and Bivalent Formation in C. elegans’, PLoS Genetics. Edited by R. S. Hawley. Public Library of Science, 4(10), p. e1000235. doi: 10.1371/journal.pgen.1000235.
35. Bishop, D. K. et al. (1992) ‘DMC1: a meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression.’, Cell. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 69(3), pp. 439–56. doi: 10.1016/0092-8674(92)90446-J.
36. Bishop, D. K. and Zickler, D. (2004) ‘Early decision: Meiotic crossover interference prior to stable strand exchange and synapsis’, Cell, 117(1), pp. 9–15. doi: 10.1016/S0092-8674(04)00297-1.
37. Bocquet, N. et al. (2014) ‘Structural and mechanistic insight into Holliday-junction dissolution by Topoisomerase IIIα and RMI1’, Nature Structural & Molecular Biology. Nature Publishing Group, 21(3), pp. 261–268. doi: 10.1038/nsmb.2775.
38. Bomblies, K. et al. (2010) ‘Local-Scale Patterns of Genetic Variability, Outcrossing, and Spatial Structure in Natural Stands of Arabidopsis thaliana’, PLoS Genetics. Edited by R. Mauricio. Public Library of Science, 6(3), p. e1000890. doi: 10.1371/journal.pgen.1000890.
39. Bomblies, K. et al. (2016) ‘The challenge of evolving stable polyploidy: could an increase in “crossover interference distance” play a central role?’, Chromosoma. doi: 10.1007/s00412-015-0571-4.
40. Bomblies, K., Higgins, J. D. and Yant, L. (2015) ‘Meiosis evolves: adaptation to external and internal environments.’, The New phytologist, 208(2), pp. 306–23. doi: 10.1111/nph.13499.
41. Borde, V. et al. (2009) ‘Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites’, The EMBO Journal. EMBO Press, 28(2), pp. 99–111. doi: 10.1038/emboj.2008.257.
42. Börner, G. V., Kleckner, N. and Hunter, N. (2004) ‘Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis’, Cell, 117, pp. 29–45. doi: 10.1016/S0092-8674(04)00292-2.
43. Borts, R. H. and Haber, J. E. (1987) ‘Meiotic recombination in yeast: alteration by multiple heterozygosities.’, Science (New York, N.Y.). American Association for the Advancement of Science, 237(4821), pp. 1459–65. doi: 10.1126/science.2820060.
44. Bouchabke, O. et al. (2008) ‘Natural Variation in Arabidopsis thaliana as a Tool for Highlighting Differential Drought Responses’, PLoS ONE. Edited by J. Kroymann. Public Library of Science, 3(2), p. e1705. doi: 10.1371/journal.pone.0001705.
45. Bouché, N. and Bouchez, D. (2001) ‘Arabidopsis gene knockout: phenotypes wanted’, Current Opinion in Plant Biology. Elsevier Current Trends, 4(2), pp. 111–117. doi: 10.1016/S1369-5266(00)00145-X.
46. Bourguet, D. et al. (2003) ‘Genetic recombination and adaptation to fluctuating environments: selection for geotaxis in Drosophila melanogaster’. doi: 10.1038/sj.hdy.6800283.
47. Brachi, B. et al. (2010) ‘Linkage and association mapping of Arabidopsis thaliana flowering time in nature.’, PLoS genetics. Public Library of Science, 6(5), p. e1000940. doi: 10.1371/journal.pgen.1000940.
48. Brennan, A. C. et al. (2014) ‘The genetic structure of Arabidopsis thaliana in the south-western Mediterranean range reveals a shared history between North Africa and southern Europe.’, BMC plant biology, 14, p. 17. doi: 10.1186/1471-2229-14-17.

49. Brick, K. et al. (2012) 'Genetic recombination is directed away from functional genomic elements in mice.', Nature. NIH Public Access, 485(7400), pp. 642–5. doi: 10.1038/nature11089.
50. Broman, K. W. et al. (1998) 'Comprehensive Human Genetic Maps: Individual and Sex-Specific Variation in Recombination', The American Journal of Human Genetics, 63(3), pp. 861–869. doi: 10.1086/302011.
51. Broman, K. W. et al. (2003) 'R/qtl: QTL mapping in experimental crosses', Bioinformatics, 19(7), pp. 889–890. doi: 10.1093/bioinformatics/btg112.
52. Broman, K. W. (2012) 'A shorter tour of R / qtl', Rqtl, pp. 1–8.
53. Broman, K. W. and Sen, S. (2009) 'A guide to QTL mapping with R/qtl'. Available at: http://www.rqtl.org/book/rqtlbook_ch04.pdf (Accessed: 30 September 2017).
54. Brown, G. M. et al. (1998) 'Genetic Analysis of Meiotic Recombination in Humans by Use of Sperm Typing: Reduced Recombination within a Heterozygous Paracentric Inversion of Chromosome 9q32-q34.3', The American Journal of Human Genetics, 62(6), pp. 1484–1492. doi: 10.1086/301863.
55. Brown, J. and Sundaresan, V. (1991) 'A recombination hotspot in the maize A1 intragenic region', Theoretical and Applied Genetics. Springer-Verlag, 81(2), pp. 185–188. doi: 10.1007/BF00215721.
56. Brown, M. S. and Bishop, D. K. (2014) 'DNA strand exchange and RecA homologs in meiosis.', Cold Spring Harbor perspectives in biology. Cold Spring Harbor Laboratory Press, 7(1), p. a016659. doi: 10.1101/cshperspect.a016659.
57. Buard, J. et al. (2009) 'Distinct histone modifications define initiation and repair of meiotic recombination in the mouse.', The EMBO journal. EMBO Press, 28(17), pp. 2616–24. doi: 10.1038/emboj.2009.207.
58. Busygina, V. et al. (2012) 'Novel attributes of Hed1 affect dynamics and activity of the Rad51 presynaptic filament during meiotic recombination.', The Journal of biological chemistry. American Society for Biochemistry and Molecular Biology, 287(2), pp. 1566–75. doi: 10.1074/jbc.M111.297309.
59. Caceres, M., Puig, M. and Ruiz, A. (2001) 'Molecular Characterization of Two Natural Hotspots in the Drosophila buzzatii Genome Induced by Transposon Insertions', Genome Research. Cold Spring Harbor Laboratory Press, 11(8), pp. 1353–1364. doi: 10.1101/gr.174001.
60. Cai, X. et al. (2003) 'The Arabidopsis SYN1 cohesin protein is required for sister chromatid arm cohesion and homologous chromosome pairing.', Journal of cell science. The Company of Biologists Ltd, 116(Pt 14), pp. 2999–3007. doi: 10.1242/jcs.00601.
61. Campbell, N. Reece, J. Urry, L. Cain, M. Wasserman, S. Minorsky, P. and Jackson, R. Biology 8th Edition. San Francisco: *Pearson Education, Inc.*, 2008. Print
62. Cao, J. et al. (2011) 'Whole-genome sequencing of multiple Arabidopsis thaliana populations.', Nature genetics. Nature Publishing Group, 43(10), pp. 956–63. doi: 10.1038/ng.911.
63. Carballo, J. A. et al. (2008) 'Phosphorylation of the Axial Element Protein Hop1 by Mec1/Tel1 Ensures Meiotic Interhomolog Recombination', Cell. Cell Press, 132(5), pp. 758–770. doi: 10.1016/J.CELL.2008.01.035.
64. Carpenter, A. E. et al. (2006) 'CellProfiler: image analysis software for identifying and quantifying cell phenotypes', Genome Biology, 7(10). doi: 10.1186/gb-2006-7-10-r100.
65. Catcheside, D. G. (1977) The genetics of recombination. Arnold. Available at: http://agris.fao.org/agris-search/search.do?recordID=US201300546522 (Accessed: 29 September 2017).

66. Chakraborty, U. and Alani, E. (2016) 'Understanding how mismatch repair proteins participate in the repair/anti-recombination decision', FEMS Yeast Research. Edited by U. Mortensen. Oxford University Press, 16(6), p. fow071. doi: 10.1093/femsyr/fow071.
67. Charlesworth, B. (1975) 'RECOMBINATION MODIFICATION IN A FLUCTUATING ENVIRONMENT'. Available at: http://www.genetics.org/content/genetics/83/1/181.full.pdf (Accessed: 19 April 2018).
68. Charlesworth, B. and Barton, N. H. (1996) 'Recombination load associated with selection for increased recombination', Genetical Research. Cambridge University Press, 67(01), p. 27. doi: 10.1017/S0016672300033450.
69. Charlesworth, D., Charlesworth, B. and Strobeck, D. C. (1976) 'EFFECTS OF SELFING ON SELECTION FOR RECOMBINATION'. Available at: http://www.genetics.org/content/genetics/86/1/213.full.pdf (Accessed: 16 March 2018).
70. Chelysheva, L. et al. (2005) 'AtREC8 and AtSCC3 are essential to the monopolar orientation of the kinetochores during meiosis', Journal of Cell Science, 118(20), pp. 4621–4632. doi: 10.1242/jcs.02583.
71. Chelysheva, L. et al. (2007) 'Zip4/Spo22 Is Required for Class I CO Formation but Not for Synapsis Completion in Arabidopsis thaliana', PLoS Genetics. Public Library of Science, 3(5), p. e83. doi: 10.1371/journal.pgen.0030083.
72. Chelysheva, L. *et al*. (2010) 'An Easy Protocol for Studying Chromatin and Recombination Protein Dynamics during *Arabidopsis thaliana* Meiosis: Immunodetection of Cohesins, Histones and MLH1', *Cytogenet Genome Res,* 129, pp. 143-153
73. Chelysheva, L. et al. (2012) 'The Arabidopsis HEI10 is a new ZMM protein related to Zip3.', PLoS genetics. Public Library of Science, 8(7), p. e1002799. doi: 10.1371/journal.pgen.1002799.
74. Chen, C. et al. (2005) 'The Arabidopsis ROCK-N-ROLLERS gene encodes a homolog of the yeast ATP-dependent DNA helicase MER3 and is required for normal meiotic crossover formation', Plant Journal, 43(3), pp. 321–334. doi: 10.1111/j.1365-313X.2005.02461.x.
75. Chen, W. et al. (1999) 'The role of the mismatch repair machinery in regulating mitotic and meiotic recombination between diverged sequences in yeast.', Genetics. Genetics, 151(4), pp. 1299–313. doi: 10.1016/0076-6879(87)54076-9.
76. Cheng, C. H. et al. (2006) 'SUMO modifications control assembly of synaptonemal complex and polycomplex in meiosis of Saccharomyces cerevisiae', Genes and Development, 20(15), pp. 2067–2081. doi: 10.1101/gad.1430406.
77. Cheng, Z. et al. (2001) 'High-Resolution Pachytene Chromosome Mapping of Bacterial Artificial Chromosomes Anchored by Genetic Markers Reveals the Centromere Location and the Distribution of Genetic Recombination Along Chromosome 10 of Rice', Genetics, 157(4). Available at: http://www.genetics.org/content/157/4/1749?ijkey=d45fba6564849579009ac5a63a0bb21adaea1b85&keytype2=tf_ipsecsha (Accessed: 5 September 2017).
78. Choi, K. et al. (2013) 'Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters.', Nature genetics. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 45(11), pp. 1327–36. doi: 10.1038/ng.2766.
79. Choi, K. et al. (2016) 'Recombination Rate Heterogeneity within Arabidopsis Disease Resistance Genes', PLOS Genetics. Edited by R. Michelmore. Public Library of Science, 12(7), p. e1006179. doi: 10.1371/journal.pgen.1006179.

80. Choi, K. et al. (2017) Nucleosomes and DNA methylation shape meiotic DSB frequency in Arabidopsis transposons and gene regulatory regions, bioRxiv. doi: https://doi.org/10.1101/160911.
81. Choulet, F. et al. (2014) 'Structural and functional partitioning of bread wheat chromosome 3B.', Science (New York, N.Y.). American Association for the Advancement of Science, 345(6194), p. 1249721. doi: 10.1126/science.1249721.
82. Chowdhury, D. et al. (2005) 'γ-H2AX Dephosphorylation by Protein Phosphatase 2A Facilitates DNA Double-Strand Break Repair', Molecular Cell. Cell Press, 20(5), pp. 801–809. doi: 10.1016/J.MOLCEL.2005.10.003.
83. Chowdhury, R. et al. (2009) 'Genetic Analysis of Variation in Human Meiotic Recombination', PLoS Genetics. Edited by G. P. Copenhaver. Public Library of Science, 5(9), p. e1000648. doi: 10.1371/journal.pgen.1000648.
84. Civardi, L. et al. (1994) 'The relationship between genetic and physical distances in the cloned al-h2 interval of the Zea mays L. genome', Genetics, 91, pp. 8268–8272. Available at: http://www.pnas.org/content/91/17/8268.full.pdf (Accessed: 5 September 2017).
85. Clarke, J. D. (2009) 'Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation.', Cold Spring Harbor protocols. Cold Spring Harbor Laboratory Press, 2009(3), p. pdb.prot5177. doi: 10.1101/pdb.prot5177.
86. Colegrave, N. (2002) 'Sex releases the speed limit on evolution', Nature. Nature Publishing Group, 420(6916), pp. 664–666. doi: 10.1038/nature01191.
87. Colomé-Tatché, M. et al. (2012) 'Features of the Arabidopsis recombination landscape resulting from the combined loss of sequence variation and DNA methylation.', Proceedings of the National Academy of Sciences of the United States of America, 109(40), pp. 16240–5. doi: 10.1073/pnas.1212955109.
88. Comeron, J. M. (2017) 'Background selection as null hypothesis in population genomics: insights and challenges from Drosophila studies.', Philosophical transactions of the Royal Society of London. Series B, Biological sciences. The Royal Society, 372(1736), p. 20160471. doi: 10.1098/rstb.2016.0471.
89. Consortium, G. et al. (2016) '1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana Resource 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana', CS78942 Cell, 166, pp. 481–491. doi: 10.1016/j.cell.2016.05.063.
90. Coop, G. et al. (2008) 'Patterns Among Humans Extensive Variation in Fine-Scale Recombination High-Resolution Mapping of Crossovers Reveals High-Resolution Mapping of Crossovers Reveals Extensive Variation in Fine-Scale Recombination Patterns Among Humans', 1395. doi: 10.1126/science.1151851.
91. Cooper, T. F. (2007) 'Recombination Speeds Adaptation by Reducing Competition between Beneficial Mutations in Populations of Escherichia coli', PLoS Biology. Edited by N. H. Barton. Public Library of Science, 5(9), p. e225. doi: 10.1371/journal.pbio.0050225.
92. Copenhaver, G. P. et al. (1999) 'Genetic definition and sequence analysis of Arabidopsis centromeres.', Science (New York, N.Y.). American Association for the Advancement of Science, 286(5449), pp. 2468–74. doi: 10.1126/SCIENCE.286.5449.2468.
93. Copenhaver, G. P., Housworth, E. A. and Stahl, F. W. (2002) 'Crossover Interference in Arabidopsis', Genetics, 160(4), pp. 1631–1639. Available at: http://www.genetics.org/content/160/4/1631.short (Accessed: 22 January 2015).
94. Copenhaver, G. P., Keith, K. C. and Preuss, D. (2000) 'Tetrad Analysis in Higher Plants. A Budding Technology'. Available at: http://www.plantphysiol.org/content/plantphysiol/124/1/7.full.pdf (Accessed: 22 April 2018).

95. Costa, Y. and Cooke, H. J. (2007) 'Dissecting the mammalian synaptonemal complex using targeted mutations'. doi: 10.1007/s10577-007-1142-1.
96. Couteau, F. et al. (1999) 'Random Chromosome Segregation without Meiotic Arrest in Both Male and Female Meiocytes of a dmc1 Mutant of Arabidopsis', The Plant Cell Online, 11(9). Available at: http://www.plantcell.org/content/11/9/1623 (Accessed: 18 August 2017).
97. Crismani, W. et al. (2012) 'FANCM Limits Meiotic Crossovers', Science, 336(6088). Available at: http://science.sciencemag.org/content/336/6088/1588.full (Accessed: 20 August 2017).
98. Crismani, W., Girard, C. and Mercier, R. (2013) 'Tinkering with meiosis.', Journal of experimental botany, 64(1), pp. 55–65. doi: 10.1093/jxb/ers314.
99. Daley, J. M. et al. (2014) 'Multifaceted role of the Topo IIIα-RMI1-RMI2 complex and DNA2 in the BLM-dependent pathway of DNA break end resection.', Nucleic acids research. Oxford University Press, 42(17), pp. 11083–91. doi: 10.1093/nar/gku803.
100. Dapper, A. L. and Payseur, B. A. (2017) 'Connecting theory and data to understand recombination rate evolution.', Philosophical transactions of the Royal Society of London. Series B, Biological sciences. The Royal Society, 372(1736), p. 20160469. doi: 10.1098/rstb.2016.0469.
101. Davies, B. et al. (2016) 'Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice.', Nature. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 530(7589), pp. 171–176. doi: 10.1038/nature16931.
102. Dawson, D. A. et al. (2007) 'Gene Order and Recombination Rate in Homologous Chromosome Regions of the Chicken and a Passerine Bird', Molecular Biology and Evolution. University of Nottingham, Nottingham (UK), 24(7), pp. 1537–1552. doi: 10.1093/molbev/msm071.
103. De Muyt, A. et al. (2012) 'BLM Helicase Ortholog Sgs1 Is a Central Regulator of Meiotic Recombination Intermediate Metabolism', Molecular Cell, 46(1), pp. 43–53. doi: 10.1016/j.molcel.2012.02.020.
104. Derzhavets, E. M., Korol, A. B., & Nevo, E. (1996). Increased male recombination rate in D. melanogaster correlated with population adaptation to stressful conditions. Dros. Inf. Serv, 77, 1-2.
105. Deshaies, R. J. and Joazeiro, C. A. P. (2009) 'RING Domain E3 Ubiquitin Ligases', Annual Review of Biochemistry. Annual Reviews , 78(1), pp. 399–434. doi: 10.1146/annurev.biochem.78.101807.093809.
106. DeVeaux, L. C. and Smith, G. R. (1994) 'Region-specific activators of meiotic recombination in Schizosaccharomyces pombe.', Genes & development. Cold Spring Harbor Laboratory Press, 8(2), pp. 203–10. doi: 10.1101/GAD.8.2.203.
107. Dion, É. et al. (2007) 'An Arabidopsis MLH1 mutant exhibits reproductive defects and reveals a dual role for this gene in mitotic recombination', The Plant Journal. Blackwell Publishing Ltd, 51(3), pp. 431–440. doi: 10.1111/j.1365-313X.2007.03145.x.
108. Dobzhansky, T. and Sturtevant, A. H. (1938) 'Inversions in the Chromosomes of Drosophila Pseudoobscura.', Genetics. Genetics Society of America, 23(1), pp. 28–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17246876 (Accessed: 30 September 2017).
109. Dooner, H. K. (1986) 'Genetic Fine Structure of the BRONZE Locus in Maize.', Genetics. Genetics, 113(4), pp. 1021–36. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17246338 (Accessed: 2 July 2016).
110. Dooner, H. K. (2002) 'Extensive Interallelic Polymorphisms Drive Meiotic Recombination into a Crossover Pathway', THE PLANT CELL ONLINE. American Society of Plant Biologists, 14(5), pp. 1173–1183. doi: 10.1105/tpc.001271.

111. Dooner, H. K. and He, L. (2008) ‘Maize Genome Structure Variation: Interplay between Retrotransposon Polymorphisms and Genic Recombination’, THE PLANT CELL ONLINE. American Society of Plant Biologists, 20(2), pp. 249–258. doi: 10.1105/tpc.107.057596.
112. Dowrick, G. (1956) ‘Thw Influence of Temperature on Meiosis’, Heredity, 11, pp. 37–49.
113. Drouaud, J. et al. (2006) ‘Variation in crossing-over rates across chromosome 4 of Arabidopsis thaliana reveals the presence of meiotic recombination “hot spots”.’, Genome research, 16(1), pp. 106–14. doi: 10.1101/gr.4319006.
114. Drouaud, J. et al. (2007) ‘Sex-specific crossover distributions and variations in interference level along Arabidopsis thaliana chromosome 4.’, PLoS genetics, 3(6), p. e106. doi: 10.1371/journal.pgen.0030106.
115. Drouaud, J. et al. (2013) ‘Contrasted patterns of crossover and non-crossover at Arabidopsis thaliana meiotic recombination hotspots.’, PLoS genetics, 9(11), p. e1003922. doi: 10.1371/journal.pgen.1003922.
116. Dumont, B. L. and Payseur, B. A. (2011) ‘Evolution of the Genomic Recombination Rate in Murid Rodents’, Genetics, 187(3). Available at: http://www.genetics.org/content/187/3/643.short (Accessed: 4 September 2017).
117. Duroc, Y. et al. (2017) ‘Concerted action of the MutLβ heterodimer and Mer3 helicase regulates the global extent of meiotic gene conversion.’, eLife. eLife Sciences Publications, Ltd, 6. doi: 10.7554/eLife.21900.
118. Durvasula, A. et al. (2017) ‘African genomes illuminate the early history and transition to selfing in Arabidopsis thaliana’. doi: 10.1073/pnas.1616736114.
119. Dvorák, J., Luo, M. C. and Yang, Z. L. (1998) ‘Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing aegilops species.’, Genetics. Genetics Society of America, 148(1), pp. 423–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9475752 (Accessed: 19 September 2016).
120. Ederveen, A. et al. (2015) ‘Modulating crossover positioning by introducing large structural changes in chromosomes’, BMC Genomics. BioMed Central, 16(1), p. 89. doi: 10.1186/s12864-015-1276-z.
121. Edwards, K., Johnstone, C. and Thompson, C. (1991) ‘A simple and rapid method for the preparation of plant genomic DNA for PCR analysis.’, Nucleic acids research, 19(6), p. 1349. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=333874&tool=pmcentrez&rendertype=abstract (Accessed: 22 January 2015).
122. Eichinger, C. S. and Jentsch, S. (2010) ‘Synaptonemal complex formation and meiotic checkpoint signaling are linked to the lateral element protein Red1’, Proceedings of the National Academy of Sciences, 107(25), pp. 11370–11375. doi: 10.1073/pnas.1004248107.
123. El-Soda, M. et al. (2014) ‘Genotype \times environment interaction {QTL} mapping in plants: lessons from \emph{Arabidopsis}’, Trends in Plant Science, 19(6), pp. 390–398.
124. Ellermeier, C. et al. (2010) ‘RNAi and heterochromatin repress centromeric meiotic recombination.’, Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 107(19), pp. 8701–5. doi: 10.1073/pnas.0914160107.
125. Emmanuel, E. et al. (2006) ‘The role of AtMSH2 in homologous recombination in Arabidopsis thaliana.’, EMBO reports. EMBO Press, 7(1), pp. 100–5. doi: 10.1038/sj.embor.7400577.

126. Esch, E. et al. (2007) ‘Using crossover breakpoints in recombinant inbred lines to identify quantitative trait loci controlling the global recombination frequency.’, Genetics, 177(3), pp. 1851–8. doi: 10.1534/genetics.107.080622.
127. Fasching, C. L. et al. (2015) ‘Top3-Rmi1 dissolve Rad51-mediated D loops by a topoisomerase-based mechanism’, Molecular Cell. Elsevier Inc., 57(4), pp. 595–606. doi: 10.1016/j.molcel.2015.01.022.
128. Feldmant, M. W. and Liberman, U. (1986) ‘An evolutionary reduction principle for genetic modifiers (recombinadon/mutation/migration/evolutlonarily stable strategy/evolutionary genetic stability)’, Evolution, 83, pp. 4824–4827. Available at: http://www.pnas.org/content/pnas/83/13/4824.full.pdf (Accessed: 19 April 2018).
129. Ferdous, M. et al. (2012) ‘Inter-homolog crossing-over and synapsis in Arabidopsis meiosis are dependent on the chromosome axis protein AtASY3.’, PLoS genetics. Public Library of Science, 8(2), p. e1002507. doi: 10.1371/journal.pgen.1002507.
130. Fernandes, J. et al. (2017) ‘FIGL1 and its novel partner FLIP form a conserved complex that regulates homologous recombination.’, bioRxiv. Available at: http://www.biorxiv.org/content/early/2017/07/06/159657 (Accessed: 21 August 2017).
131. Filiault, D. L. et al. (2008) ‘Amino acid polymorphisms in Arabidopsis phytochrome B cause differential responses to light’, Proceedings of the National Academy of Sciences, 105(8), pp. 3157–3162. doi: 10.1073/pnas.0712174105.
132. Fledel-Alon, A. et al. (2011) ‘Variation in Human Recombination Rates and Its Genetic Determinants’, PLoS ONE. Edited by L. Quintana-Murci. Public Library of Science, 6(6), p. e20321. doi: 10.1371/journal.pone.0020321.
133. Francis, K. E. et al. (2007) ‘Pollen tetrad-based visual assay for meiotic recombination in Arabidopsis.’, Proceedings of the National Academy of Sciences of the United States of America, 104(10), pp. 3913–8. doi: 10.1073/pnas.0608936104.
134. Fransz, P. et al. (2016) ‘Molecular, genetic and evolutionary analysis of a paracentric inversion in Arabidopsis thaliana’, The Plant Journal, 88(2), pp. 159–178. doi: 10.1111/tpj.13262.
135. Fu, H. et al. (2001) ‘The highly recombinogenic bz locus lies in an unusually gene-rich region of the maize genome’, Proceedings of the National Academy of Sciences. National Academy of Sciences, 98(15), pp. 8903–8908. doi: 10.1073/pnas.141221898.
136. Fu, H., Zheng, Z. and Dooner, H. K. (2002) ‘Recombination rates between adjacent genic and retrotransposon regions in maize vary by 2 orders of magnitude.’, Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 99(2), pp. 1082–7. doi: 10.1073/pnas.022635499.
137. Gerton, J. and Hawley, R. (2005) ‘HOMOLOGOUS CHROMOSOME INTERACTIONS IN MEIOSIS: DIVERSITY AMIDST CONSERVATION’, Nature Reviews Genetics, 6, pp. 477–487. doi: doi:10.1038/nrg1614.
138. GIBBS, P. E., MILNE, C. and CARRILLO, M. V. (1975) ‘Correlation Between the Breeding System and Recombination Index in Five Species of Senecio’, New Phytologist, pp. 619–626. doi: 10.1111/j.1469-8137.1975.tb01428.x.
139. Gill, K. S. et al. (1996) ‘Identification and High-Density Mapping of Gene-Rich Regions in Chromosome Group I of Wheat’, Genetics, 144(1883–1891). Available at: http://www.genetics.org/content/genetics/144/4/1883.full.pdf (Accessed: 5 September 2017).
140. Girard, C. et al. (2015) ‘AAA-ATPase FIDGETIN-LIKE 1 and Helicase FANCM Antagonize Meiotic Crossovers by Distinct Mechanisms’, PLOS Genetics. Edited by M. Lichten. Public Library of Science, 11(7), p. e1005369. doi: 10.1371/journal.pgen.1005369.

141. Giraut, L. et al. (2011) ‘Genome-wide crossover distribution in Arabidopsis thaliana meiosis reveals sex-specific patterns along chromosomes.’, PLoS genetics, 7(11), p. e1002354. doi: 10.1371/journal.pgen.1002354.
142. Glazier, A. M. (2002) ‘Finding Genes That Underlie Complex Traits’, Science, 298(5602), pp. 2345–2349. doi: 10.1126/science.1076641.
143. Goddard, M. R., Godfray, H. C. J. and Burt, A. (2005) ‘Sex increases the efficacy of natural selection in experimental yeast populations’, Nature. Nature Publishing Group, 434(7033), pp. 636–640. doi: 10.1038/nature03405.
144. Goh, C.-S. and Cohen, F. E. (2002) ‘Co-evolutionary Analysis Reveals Insights into Protein–Protein Interactions’, Journal of Molecular Biology. Academic Press, 324(1), pp. 177–192. doi: 10.1016/S0022-2836(02)01038-0.
145. Goldman, A. S. H. and Hulten, M. A. (1993) ‘Meiotic analysis by FISH of a human male 46,XY,t(15;20)(q11.2;q11.2) translocation heterozygote: quadrivalent configuration, orientation and first meiotic segregation’, Chromosoma. Springer-Verlag, 102(2), pp. 102–111. doi: 10.1007/BF00356027.
146. Gorlov, I. P., Ladygina, T. Y. and Borodin, P. M. (1993) ‘Chiasma distribution in the first bivalent of mice carrying a double insertion of homogeneously - staining regions in homo - and heterozygous states’, Heredity, 70, pp. 642–647. doi: 10.1038/hdy.1993.92.
147. Goymer, P. (2007) ‘Genetic variation: Synonymous mutations break their silence’, Nature Reviews Genetics, 8(2), p. 92. doi: 10.1038/nrg2056.
148. Grant, V. (1953) ‘Cytogenetics of the hybrid Gilia millefoliata achilleaefolia’, Chromosoma. Springer-Verlag, 5(1), pp. 372–390. doi: 10.1007/BF01271494.
149. Griffiths, A. Wessler, S. Lewontin, R. and Carroll, S. Introduction to Genetic Analysis 9th Edition. New York: *W. H. Freeman and Company*, 2008. Print
150. Haag, C. R. et al. (2017) ‘Low recombination rates in sexual species and sex-asex transitions.’, Philosophical transactions of the Royal Society of London. Series B, Biological sciences. The Royal Society, 372(1736), p. 20160461. doi: 10.1098/rstb.2016.0461.
151. Hadany, L. and Comeron, J. M. (2008) ‘Why are sex and recombination so common?’, Annals of the New York Academy of Sciences, 1133, pp. 26–43. doi: 10.1196/annals.1438.011.
152. Haley, C. S. and Knott, S. A. (1992) ‘A simple regression method for mapping quantitative trait loci in line crosses using flanking markers’, Heredity, 69(4), pp. 315–324. doi: 10.1038/hdy.1992.131.
153. Hammarlund, M. et al. (2005) ‘Heterozygous Insertions Alter Crossover Distribution but Allow Crossover Interference in Caenorhabditis elegans’, Genetics, 171(3).
154. HANSSON, B. et al. (2006) ‘Comparative gene mapping in Arabidopsis lyrata chromosomes 1 and 2 and the corresponding A. thaliana chromosome 1: recombination rates, rearrangements and centromere location’, Genetical Research. Cambridge University Press, 87(02), p. 75. doi: 10.1017/S0016672306008020.
155. Harfe, B. D. and Jinks-Robertson, S. (2000) ‘DNA Mismatch Repair and Genetic Instability’, Annual Review of Genetics, 34(1), pp. 359–399. doi: 10.1146/annurev.genet.34.1.359.
156. Haudry, A. et al. (2008) ‘Mating system and recombination affect molecular evolution in four Triticeae species’, Genetics Research, 90(1), pp. 97–109. doi: 10.1017/S0016672307009032.
157. He, L. and Dooner, H. K. (2009) ‘Haplotype structure strongly affects recombination in a maize genetic interval polymorphic for Helitron and retrotransposon insertions.’,

Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 106(21), pp. 8410–6. doi: 10.1073/pnas.0902972106.
158. Hellens, R. P. et al. (2000) 'pGreen: A versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation', Plant Molecular Biology, 42(6), pp. 819–832. doi: 10.1023/A:1006496308160.
159. Henderson, I. R. (2012) 'Control of meiotic recombination frequency in plant genomes.', Current opinion in plant biology, 15(5), pp. 556–61. doi: 10.1016/j.pbi.2012.09.002.
160. HENDERSON, S. A. and EDWARDS, R. G. (1968) 'Chiasma Frequency and Maternal Age in Mammals', Nature. Nature Publishing Group, 218(5136), pp. 22–28. doi: 10.1038/218022a0.
161. Henikoff, S. et al. (2000) 'Targeted screening for induced mutations', Nature Biotechnology. Nature Publishing Group, 18(4), pp. 455–457. doi: 10.1038/74542.
162. Hewitt, G. M. (1964) 'Population cytology of British grasshoppers', Chromosoma. Springer-Verlag, 15(2), pp. 212–230. doi: 10.1007/BF00285730.
163. Higgins, J. D. et al. (2004) 'The Arabidopsis MutS homolog AtMSH4 functions at an early step in recombination: evidence for two classes of recombination in Arabidopsis', Genes & Development, 18(20), pp. 2557–2570. doi: 10.1101/gad.317504.
164. Higgins, J. D. et al. (2008) 'AtMSH5 partners AtMSH4 in the class I meiotic crossover pathway in Arabidopsis thaliana, but is not required for synapsis', Plant Journal, 55, pp. 28–39. doi: 10.1111/j.1365-313X.2008.03470.x.
165. Hill, W. G. and Robertson, A. (1966) 'The effect of linkage on limits to artificial selection', Genetical Research. Cambridge University Press, 8(03), p. 269. doi: 10.1017/S0016672300010156.
166. Hoffmann, A. A., Parsons, P. A. (1991) Evolutionary genetics and environmental stress. Oxford University Press. Available at: http://agris.fao.org/agris-search/search.do?recordID=US201300695279 (Accessed: 29 March 2018).
167. Hollingsworth, N. M. and Brill, S. J. (2004) 'The Mus81 solution to resolution: generating meiotic crossovers without Holliday junctions.', Genes & development. Cold Spring Harbor Laboratory Press, 18(2), pp. 117–25. doi: 10.1101/gad.1165904.
168. Hollingsworth, N. M. and Byers, B. (1989) 'HOP1: a yeast meiotic pairing gene.', Genetics, 121(3).
169. Hollingsworth, N. M. and Ponte, L. (1997) 'Genetic Interactions Between HOP1, RED1 and MEK1 Suggest That MEK1 Regulates Assembly of Axial Element Components During Meiosis in the Yeast Saccharomyces cerevisiae', Genetics, 147(1).
170. Hollister, J. D. et al. (2012) 'Genetic Adaptation Associated with Genome-Doubling in Autotetraploid Arabidopsis arenosa', PLoS Genetics. Edited by R. Mauricio. Public Library of Science, 8(12), p. e1003093. doi: 10.1371/journal.pgen.1003093.
171. Hörandl, E. (2009) 'A combinational theory for maintenance of sex.', Heredity. Europe PMC Funders, 103(6), pp. 445–57. doi: 10.1038/hdy.2009.85.
172. Horton, M. W. et al. (2012) 'Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel.', Nature genetics. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 44(2), pp. 212–6. doi: 10.1038/ng.1042.
173. Huang, S. et al. (2016) 'A proposed regulatory framework for genome-edited crops', Nature Genetics. Nature Publishing Group, 48(2), pp. 109–111. doi: 10.1038/ng.3484.

174. Huang, Z., Footitt, S. and Finch-Savage, W. E. (2014) 'The effect of temperature on reproduction in the summer and winter annual Arabidopsis thaliana ecotypes Bur and Cvi', Annals of Botany. Oxford University Press, 113(6), pp. 921–929. doi: 10.1093/aob/mcu014.
175. Hunt, P. A. and Hassold, T. J. (2002) 'Sex matters in meiosis.', Science (New York, N.Y.). American Association for the Advancement of Science, 296(5576), pp. 2181–3. doi: 10.1126/science.1071907.
176. Hunter, N. et al. (1996) 'The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid.', The EMBO journal. European Molecular Biology Organization, 15(7), pp. 1726–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8612597 (Accessed: 30 September 2017).
177. Da Ines, O. et al. (2013) 'Effects of XRCC2 and RAD51B mutations on somatic and meiotic recombination in Arabidopsis thaliana.', The Plant journal : for cell and molecular biology, 74(6), pp. 959–70. doi: 10.1111/tpj.12182.
178. Ito, H. et al. (2006) 'Ecotype-specific and chromosome-specific expansion of variant centromeric satellites in Arabidopsis thaliana', Molecular Genetics and Genomics. Springer-Verlag, 277(1), pp. 23–30. doi: 10.1007/s00438-006-0172-2.
179. Jackson, N. et al. (2006) 'Reduced meiotic crossovers and delayed prophase I progression in AtMLH3-deficient Arabidopsis.', The EMBO journal. EMBO Press, 25(6), pp. 1315–23. doi: 10.1038/sj.emboj.7600992.
180. Jones, H. D. (2015) 'GM foods: is there a way forward?', The Proceedings of the Nutrition Society. Cambridge University Press, 74(3), pp. 198–201. doi: 10.1017/S0029665115002098.
181. JONES, N. and REES, H. (1966) 'Chiasma Frequencies and the Potential Genetic Variability of Lolium Populations', Nature. Nature Publishing Group, 211(5047), pp. 432–433. doi: 10.1038/211432a0.
182. Joyce, E. F. and McKim, K. S. (2010) 'Chromosome Axis Defects Induce a Checkpoint-Mediated Delay and Interchromosomal Effect on Crossing Over during Drosophila Meiosis', PLoS Genetics. Edited by G. P. Copenhaver. Public Library of Science, 6(8), p. e1001059. doi: 10.1371/journal.pgen.1001059.
183. Joyce, E. F. and McKim, K. S. (2011) 'Meiotic checkpoints and the interchromosomal effect on crossing over in Drosophila females', Fly. Taylor & Francis, 5(2), pp. 134–140. doi: 10.4161/fly.5.2.14767.
184. Juenger, T., Purugganan, M. and Mackay, T. F. C. (2000) 'Quantitative Trait Loci for Floral Morphology in Arabidopsis thaliana', Genetics, 156(3). Available at: http://www.genetics.org/content/156/3/1379.short (Accessed: 6 September 2017).
185. Kagawa, N., Nagaki, K. and Tsujimoto, H. (2002) 'Tetrad-FISH analysis reveals recombination suppression by interstitial heterochromatin sequences in rye (Secale cereale)', Molecular Genetics and Genomics. Springer-Verlag, 267(1), pp. 10–15. doi: 10.1007/s00438-001-0634-5.
186. Kauppi, L., Jeffreys, A. J. and Keeney, S. (2004) 'Where the crossovers are: recombination distributions in mammals.', Nature reviews. Genetics, 5(6), pp. 413–24. doi: 10.1038/nrg1346.
187. Kaur, H., De Muyt, A. and Lichten, M. (2015) 'Top3-Rmi1 DNA single-strand decatenase is integral to the formation and resolution of meiotic recombination intermediates.', Molecular cell, 57(4), pp. 583–94. doi: 10.1016/j.molcel.2015.01.020.
188. Kawakatsu, T. et al. (2016) 'Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions', Cell. Elsevier, 166(2), pp. 492–505. doi: 10.1016/j.cell.2016.06.044.

189. Kearsey, M. J. and Farquhar, a G. L. (1998) 'Short Review QTL analysis in plants ; where are we now?', Heredity (Edinb), 80(October 1997), pp. 137–142. doi: 10.1046/j.1365-2540.1998.00500.x.
190. Keeney, S., Giroux, C. N. and Kleckner, N. (1997) 'Meiosis-Specific DNA Double-Strand Breaks Are Catalyzed by Spo11, a Member of a Widely Conserved Protein Family', Cell, 88(3), pp. 375–384. doi: 10.1016/S0092-8674(00)81876-0.
191. Keeney, S. and Neale, M. J. (2006) 'Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation', Biochemical Society Transactions, 34(4). Available at: http://www.biochemsoctrans.org/content/34/4/523 (Accessed: 22 August 2017).
192. Kerk, D. et al. (2002) 'The complement of protein phosphatase catalytic subunits encoded in the genome of Arabidopsis.', Plant physiology. American Society of Plant Biologists, 129(2), pp. 908–25. doi: 10.1104/pp.004002.
193. Kerstes, N. A. et al. (2012) 'Antagonistic experimental coevolution with a parasite increases host recombination frequency', BMC Evolutionary Biology. BioMed Central, 12(1), p. 18. doi: 10.1186/1471-2148-12-18.
194. Kim, Y., Schumaker, K. S. and Zhu, J. (2006) 'EMS Mutagenesis of Arabidopsis', Methods in Molecular Biology, 323(6), pp. 101–103.
195. Kim, Y. and Stephan, W. (2002) 'Detecting a Local Signature of Genetic Hitchhiking Along a Recombining Chromosome'. Available at: http://www.genetics.org/content/genetics/160/2/765.full.pdf (Accessed: 19 April 2018).
196. Kleckner, N. et al. (2004) 'A mechanical basis for chromosome function.', Proceedings of the National Academy of Sciences of the United States of America, 101, pp. 12592–12597. doi: 10.1073/pnas.0402724101.
197. Kleckner, N. (2006) 'Chiasma formation: chromatin/axis interplay and the role(s) of the synaptonemal complex', Chromosoma. Springer-Verlag, 115(3), pp. 175–194. doi: 10.1007/s00412-006-0055-7.
198. Kleinboelting, N. et al. (2012) 'GABI-Kat SimpleSearch: new features of the Arabidopsis thaliana T-DNA mutant database', Nucleic Acids Research, 40(D1), pp. D1211–D1215. doi: 10.1093/nar/gkr1047.
199. Koehler, K. E. et al. (2002) 'Genetic Control of Mammalian Meiotic Recombination. I. Variation in Exchange Frequencies Among Males From Inbred Mouse Strains', Genetics, 162(1). Available at: http://www.genetics.org/content/162/1/297.short (Accessed: 4 August 2017).
200. Kong, A. et al. (2004) 'Recombination rate and reproductive success in humans', Nature Genetics. Nature Publishing Group, 36(11), pp. 1203–1206. doi: 10.1038/ng1445.
201. Kong, A. et al. (2008) 'Sequence Variants in the RNF212 Gene Associate with Genome-Wide Recombination Rate', Science, 319(5868), pp. 1398–1401. doi: 10.1126/science.1152422.
202. Kong, A. et al. (2013) 'Common and low-frequency variants associated with genome-wide recombination rate', Nature Genetics. Nature Publishing Group, 46(1), pp. 11–16. doi: 10.1038/ng.2833.
203. Koornneef, M., Alonso-Blanco, C. and Vreugdenhil, D. (2004) 'NATURALLY OCCURRING GENETIC VARIATION IN ARABIDOPSIS THALIANA', Annual Review of Plant Biology. Annual Reviews, 55(1), pp. 141–172. doi: 10.1146/annurev.arplant.55.031903.141605.
204. Kovalchuk, I. et al. (2003) 'Pathogen- induced systemic plant signal triggers DNA rearrangements', Nature, 423, pp. 760–762.

205. Kover, P. X. et al. (2009) ‘A Multiparent Advanced Generation Inter-Cross to Fine-Map Quantitative Traits in Arabidopsis thaliana’, PLoS Genetics. Edited by R. Mauricio. Public Library of Science, 5(7), p. e1000551. doi: 10.1371/journal.pgen.1000551.
206. Kraft, T. et al. (1998) ‘Positive Correlation Between Recombination Rates and Levels of Genetic Variation in Natural Populations of Sea Beet (Beta vulgaris subsp. maritima)’, Genetics, 150(3).
207. Krishnakumar, V. et al. (2015) ‘Araport: the Arabidopsis Information Portal’, Nucleic Acids Research. University of California, Irvine, 43(D1), pp. D1003–D1009. doi: 10.1093/nar/gku1200.
208. Kumar, R., Bourbon, H.-M. and de Massy, B. (2010) ‘Functional conservation of Mei4 for meiotic DNA double-strand break formation from yeasts to mice.’, Genes & development. Cold Spring Harbor Laboratory Press, 24(12), pp. 1266–80. doi: 10.1101/gad.571710.
209. Künzel, G., Korzun, L. and Meister, A. (2000) ‘Cytologically Integrated Physical Restriction Fragment Length Polymorphism Maps for the Barley Genome Based on Translocation Breakpoints’, Genetics, 154(1). Available at: http://www.genetics.org/content/154/1/397?ijkey=5a49bedcdeebeacc5463ad489e986b681c0ac83d&keytype2=tf_ipsecsha (Accessed: 5 September 2017).
210. Kurzbauer, M.-T. et al. (2012) ‘The Recombinases DMC1 and RAD51 Are Functionally and Spatially Separated during Meiosis in Arabidopsis’, The Plant Cell Online, 24(5). Available at: http://www.plantcell.org/content/24/5/2058 (Accessed: 24 August 2017).
211. Lam, S. Y. et al. (2005) ‘Crossover interference on nucleolus organizing region-bearing chromosomes in Arabidopsis.’, Genetics, 170(2), pp. 807–12. doi: 10.1534/genetics.104.040055.
212. Lambing, C. et al. (2015) ‘Arabidopsis PCH2 Mediates Meiotic Chromosome Remodeling and Maturation of Crossovers.’, PLoS genetics. Public Library of Science, 11(7), p. e1005372. doi: 10.1371/journal.pgen.1005372.
213. Lambing, C., Franklin, F. C. H. and Wang, C.-J. R. (2017) ‘Understanding and Manipulating Meiotic Recombination in Plants.’, Plant physiology. American Society of Plant Biologists, 173(3), pp. 1530–1542. doi: 10.1104/pp.16.01530.
214. Laurie, D. A., Palmer, R. W. and Hulten, M. A. (1984) ‘Studies on chiasma frequency and distribution in two fertile men carrying reciprocal translocations; one with a t(9;10) karyotype and one with a t(Y;10) karyotype’, Human Genetics. Springer-Verlag, 68(3), pp. 235–247. doi: 10.1007/BF00418394.
215. Lawrence, E. J., Griffin, C. H. and Henderson, I. R. (2017) ‘Modification of meiotic recombination by natural variation in plants’, Journal of Experimental Botany. Oxford University Press, 68(20), pp. 5471–5483. doi: 10.1093/jxb/erx306.
216. Leflon, M. et al. (2010) ‘Crossovers Get a Boost in Brassica Allotriploid and Allotetraploid Hybrids’, THE PLANT CELL ONLINE. American Society of Plant Biologists, 22(7), pp. 2253–2264. doi: 10.1105/tpc.110.075986.
217. Lenormand, T. et al. (2016) ‘Evolutionary mysteries in meiosis’. Philosophical Transactions B Royal Society, 371. doi: 10.1098/rstb.2016.0001.
218. Lenormand, T. and Otto, S. P. (2000) ‘The Evolution of Recombination in a Heterogeneous Environment’, Genetics, 156(1).
219. Li, C. et al. (2016) ‘Analysis of recombination QTLs, segregation distortion, and epistasis for fitness in maize multiple populations using ultra-high-density markers’, Theoretical and Applied Genetics. Springer Berlin Heidelberg, 129(9), pp. 1775–1784. doi: 10.1007/s00122-016-2739-6.

220. Li, W. et al. (2004) ‘The Arabidopsis AtRAD51 gene is dispensable for vegetative development but required for meiosis.’, Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 101(29), pp. 10596–601. doi: 10.1073/pnas.0404110101.
221. Lichten, M. and de Massy, B. (2011) ‘The Impressionistic Landscape of Meiotic Recombination’, Cell, 147(2), pp. 267–270. doi: 10.1016/j.cell.2011.09.038.
222. Lichten, M. and Goldman, A. S. H. (1995) ‘Meiotic recombination hotspots’, Allll u. Rev. Genetics, 29, pp. 423–444.
223. Liu, S. et al. (2009) ‘Mu Transposon Insertion Sites and Meiotic Recombination Events Co-Localize with Epigenetic Marks for Open Chromatin across the Maize Genome’, PLoS Genetics. Edited by H. S. Malik. Public Library of Science, 5(11), p. e1000733. doi: 10.1371/journal.pgen.1000733.
224. Lloyd, A. et al. (2018) ‘Plasticity of meiotic recombination rates in response to temperature in Arabidopsis’.
225. Loidl, J. (1989) ‘Effects of elevated temperature on meiotic chromosome synapsis in Allium ursinum’, Chromosoma, 97(6), pp. 449–458. doi: 10.1007/BF00295029.
226. López, E. et al. (2012) ‘Looking for natural variation in chiasma frequency in Arabidopsis thaliana’, Journal of Experimental Botany. Academic Press, Inc., Orlando, FL, 63(2), pp. 887–894. doi: 10.1093/jxb/err319.
227. Macaisne, N. et al. (2008) ‘SHOC1, an XPF Endonuclease-Related Protein, Is Essential for the Formation of Class I Meiotic Crossovers’, Current Biology, 18(18), pp. 1432–1437. doi: 10.1016/j.cub.2008.08.041.
228. Macaisne, N., Vignard, J. and Mercier, R. (2011) ‘SHOC1 and PTD form an XPF–ERCC1-like complex that is required for formation of class I crossovers’, Journal of Cell Science, 124(16). Available at: http://jcs.biologists.org/content/124/16/2687.short (Accessed: 19 August 2017).
229. Maguire, M. P. and Riess, R. W. (1994) ‘The relationship of homologous synapsis and crossing over in a maize inversion.’, Genetics, 137(1).
230. Maloof, J. N. (2003) ‘Genomic approaches to analyzing natural variation in Arabidopsis thaliana’, Current Opinion in Genetics and Development, 13(6), pp. 576–582. doi: 10.1016/j.gde.2003.10.003.
231. Manzano-Piedras, E. et al. (2014) ‘Deciphering the adjustment between environment and life history in annuals: lessons from a geographically-explicit approach in Arabidopsis thaliana.’, PloS one. Public Library of Science, 9(2), p. e87836. doi: 10.1371/journal.pone.0087836.
232. Marais, G. and Charlesworth, B. (2003) ‘Genome Evolution: Recombination Speeds Up Adaptive Evolution’, Current Biology, 13(2), pp. R68–R70. doi: 10.1016/S0960-9822(02)01432-X.
233. Marais, G., Charlesworth, B. and Wright, S. I. (2004) ‘Recombination and base composition: the case of the highly self-fertilizing plant Arabidopsis thaliana.’, Genome biology, 5(7), p. R45. doi: 10.1186/gb-2004-5-7-r45.
234. Marais, G. and Galtier, N. (2003) ‘Sex chromosomes: how X-Y recombination stops’, Current Biology, 13(16), pp. R641–R643. doi: 10.1016/S0960-9822(03)00570-0.
235. Martín, A. C. et al. (2014) ‘Licensing MLH1 sites for crossover during meiosis’, Nature Communications, 5, pp. 1–5. doi: 10.1038/ncomms5580.
236. Martin, H. C. et al. (2015) ‘Multicohort analysis of the maternal age effect on recombination’, Nature Communications. Nature Publishing Group, 6(1), p. 7846. doi: 10.1038/ncomms8846.

237. Martini, E. et al. (2011) 'Genome-Wide Analysis of Heteroduplex DNA in Mismatch Repair–Deficient Yeast Cells Reveals Novel Properties of Meiotic Recombination Pathways', PLoS Genetics. Edited by M. Lichten. Public Library of Science, 7(9), p. e1002305. doi: 10.1371/journal.pgen.1002305.
238. de Massy, B. (2013) 'Initiation of Meiotic Recombination: How and Where? Conservation and Specificities Among Eukaryotes', Annual Review of Genetics. Annual Reviews , 47(1), pp. 563–599. doi: 10.1146/annurev-genet-110711-155423.
239. de Massy, B. and Nicolas, A. (1993) 'The control in cis of the position and the amount of the ARG4 meiotic double-strand break of Saccharomyces cerevisiae.', The EMBO journal. European Molecular Biology Organization, 12(4), pp. 1459–66. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8467798 (Accessed: 30 September 2017).
240. Mccallum, C. M. et al. (2000) 'Targeted screening for induced mutations', NATURE BIOTECHNOLOGY, 18. Available at: http://biotech.nature.com (Accessed: 11 April 2018).
241. McDowell, J. M. et al. (1998) 'Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of Arabidopsis.', The Plant cell. American Society of Plant Biologists, 10(11), pp. 1861–74. doi: 10.1105/TPC.10.11.1861.
242. McVean, G. A. T. et al. (2004) 'The Fine-Scale Structure of Recombination Rate Variation in the Human Genome', Science, 304(5670). Available at: http://science.sciencemag.org/content/304/5670/581 (Accessed: 3 September 2017).
243. Meinke, D. W. et al. (1998) 'Arabidopsis thaliana: a model plant for genome analysis', Science, 282(5389), p. 662,679-682. doi: 10.1126/science.282.5389.662.
244. Melamed-Bessudo, C. et al. (2005) 'A new seed-based assay for meiotic recombination in Arabidopsis thaliana.', The Plant journal: for cell and molecular biology, 43(3), pp. 458–66. doi: 10.1111/j.1365-313X.2005.02466.x.
245. Melamed-Bessudo, C. and Levy, A. A. (2012) 'Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in Arabidopsis.', Proceedings of the National Academy of Sciences of the United States of America, 109(16), pp. E981-8. doi: 10.1073/pnas.1120742109.
246. Mercier, R. et al. (2005) 'Two meiotic crossover classes cohabit in Arabidopsis: One is dependent on MER3, whereas the other one is not', Current Biology, 15, pp. 692–701. doi: 10.1016/j.cub.2005.02.056.
247. Mercier, R. et al. (2015) 'The molecular biology of meiosis in plants.', Annual review of plant biology. Annual Reviews, 66, pp. 297–327. doi: 10.1146/annurev-arplant-050213-035923.
248. Mercier, R. and Grelon, M. (2008) 'Meiosis in plants: ten years of gene discovery.', Cytogenetic and genome research. Karger Publishers, 120(3–4), pp. 281–90. doi: 10.1159/000121077.
249. Merker, J. D. et al. (2008) 'The histone methylase Set2p and the histone deacetylase Rpd3p repress meiotic recombination at the HIS4 meiotic recombination hotspot in Saccharomyces cerevisiae', DNA Repair, 7(8), pp. 1298–1308. doi: 10.1016/j.dnarep.2008.04.009.
250. Miao, C. et al. (2013) 'Central region component1, a novel synaptonemal complex component, is essential for meiotic recombination initiation in rice.', The Plant cell. American Society of Plant Biologists, 25(8), pp. 2998–3009. doi: 10.1105/tpc.113.113175.
251. Michelmore, R. W. and Meyers, B. C. (1998) 'Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process.', Genome research. Cold Spring Harbor Laboratory Press, 8(11), pp. 1113–30. doi: 10.1101/GR.8.11.1113.

252. Mihola, O. (2009) ‘Histone H3 Methyltransferase A Mouse Speciation Gene Encodes a Meiotic’, Science J. Phys. Chem. C Annu. Rev. Phys. Chem. Synth. Met. J. Phys. Soc. Jpn. Phys. Rev. Lett. Nature Science J. Chem. Phys. A. V. Pisliakov, J. Chem. Phys, 373(124), pp. 1327–1100. doi: 10.1126/science.1163601.
253. Mikhailova, E. I. et al. (1998) ‘The effect of the wheat Ph1 locus on chromatin organisation and meiotic chromosome pairing analysed by genome painting’, Chromosoma. Springer-Verlag, 107(5), pp. 339–350. doi: 10.1007/s004120050316.
254. Mirouze, M. et al. (2012) ‘Loss of DNA methylation affects the recombination landscape in Arabidopsis.’, Proceedings of the National Academy of Sciences of the United States of America, 109(15), pp. 5880–5. doi: 10.1073/pnas.1120841109.
255. Mitchell-Olds, T. (2001) ‘Arabidopsis thaliana and its wild relatives: a model system for ecology and evolution’, Trends in Ecology & Evolution, 16(12), pp. 693–700. Available at: https://pdfs.semanticscholar.org/4800/a3f4e40d9e08669e5bc63bdc352673bb03dc.pdf (Accessed: 7 April 2018).
256. Mitchell-Olds, T. and Pedersen, D. (1998) ‘The Molecular Basis of Quantitative Genetic Variation in Central and Secondary Metabolism in Arabidopsis’, Genetics, 149(2). Available at: http://www.genetics.org/content/149/2/739.full (Accessed: 6 September 2017).
257. Mitchell-Olds, T. and Schmitt, J. (2006) ‘Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis’, Nature Reviews Genetics, 441. doi: 10.1038/nature04878.
258. Mladek, C., Guger, K. and Hauser, M.-T. (2003) ‘Identification and characterization of the ARIADNE gene family in Arabidopsis. A group of putative E3 ligases.’, Plant physiology. American Society of Plant Biologists, 131(1), pp. 27–40. doi: 10.1104/pp.012781.
259. Modrich, P. and Lahue, R. (2003) ‘Mismatch Repair in Replication Fidelity, Genetic Recombination, and Cancer Biology’, http://dx.doi.org/10.1146/annurev.bi.65.070196.000533. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA.
260. Monda, K. et al. (2011) ‘Environmental regulation of stomatal response in the Arabidopsis Cvi-0 ecotype’, Planta, 234(3), pp. 555–563. doi: 10.1007/s00425-011-1424-x.
261. Morgan, C. H. et al. (2017) ‘Are the effects of elevated temperature on meiotic recombination and thermotolerance linked via the axis and synaptonemal complex ?’ doi: 10.1098/rstb.2016.0470.
262. Morrell, P. L. et al. (2004) ‘Low levels of linkage disequilibrium in wild barley (Hordeum vulgare ssp. spontaneum) despite high rates of self-fertilization’. Available at: http://www.pnas.org/content/pnas/102/7/2442.full.pdf (Accessed: 16 March 2018).
263. Moses, M. J. (1968) ‘Synaptinemal Complex’, Annual Review of Genetics, 2(1), pp. 363–412. doi: 10.1146/annurev.ge.02.120168.002051.
264. Moynihan, T. P. et al. (1999) ‘The ubiquitin-conjugating enzymes UbcH7 and UbcH8 interact with RING finger/IBR motif-containing domains of HHARI and H7-AP1.’, The Journal of biological chemistry. American Society for Biochemistry and Molecular Biology, 274(43), pp. 30963–8. doi: 10.1074/JBC.274.43.30963.
265. Muller, H. J. (1964) ‘The relation of recombination to mutational advance’, Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. Elsevier, 1(1), pp. 2–9. doi: 10.1016/0027-5107(64)90047-8.
266. De Muyt, A. et al. (2007) ‘AtPRD1 is required for meiotic double strand break formation in Arabidopsis thaliana.’, The EMBO journal. European Molecular Biology Organization, 26(18), pp. 4126–37. doi: 10.1038/sj.emboj.7601815.

267. De Muyt, A. et al. (2009) ‘A High Throughput Genetic Screen Identifies New Early Meiotic Recombination Functions in Arabidopsis thaliana’, PLoS Genetics. Edited by G. P. Copenhaver. Recombination and Meiosis, 5(9), p. e1000654. doi: 10.1371/journal.pgen.1000654.
268. De Muyt, A. et al. (2014) ‘E3 ligase Hei10: a multifaceted structure-based signaling molecule with roles within and beyond meiosis’, Genes & Development. Cold Spring Harbor Laboratory Press, 28(10), pp. 1111–1123. doi: 10.1101/gad.240408.114.
269. Myers, S. et al. (2010) ‘Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination.’, Science (New York, N.Y.). American Association for the Advancement of Science, 327(5967), pp. 876–9. doi: 10.1126/science.1182363.
270. Nachman, M. (2002) ‘Variation in recombination rate across the genome: evidence and implications’, Current Opinion in Genetics & Development, 12(6), pp. 657–663. doi: 10.1016/S0959-437X(02)00358-1.
271. Nachman, M. W. and Payseur, B. A. (2012) ‘Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice.’, Philosophical transactions of the Royal Society of London. Series B, Biological sciences. The Royal Society, 367(1587), pp. 409–21. doi: 10.1098/rstb.2011.0249.
272. Neale, M. J., Pan, J. and Keeney, S. (2005) ‘Endonucleolytic processing of covalent protein-linked DNA double-strand breaks.’, Nature. NIH Public Access, 436(7053), pp. 1053–7. doi: 10.1038/nature03872.
273. NILSSON, N.-O. and PELGER, S. (1991) ‘The relationship between natural variation in chiasma frequencies and recombination frequencies in barley’, Hereditas. Blackwell Publishing Ltd, 115(2), pp. 121–126. doi: 10.1111/j.1601-5223.1991.tb03545.x.
274. Niu, H. et al. (2005) ‘Partner choice during meiosis is regulated by Hop1-promoted dimerization of Mek1.’, Molecular biology of the cell. American Society for Cell Biology, 16(12), pp. 5804–18. doi: 10.1091/mbc.E05-05-0465.
275. Niu, H. et al. (2009) ‘Regulation of Meiotic Recombination via Mek1-Mediated Rad54 Phosphorylation’, Molecular Cell. Cell Press, 36(3), pp. 393–404. doi: 10.1016/J.MOLCEL.2009.09.029.
276. Nonomura, K.-I. et al. (2004) ‘The Novel Gene HOMOLOGOUS PAIRING ABERRATION IN RICE MEIOSIS1 of Rice Encodes a Putative Coiled-Coil Protein Required for Homologous Chromosome Pairing in Meiosis’, The Plant Cell Online, 16(4). Available at: http://www.plantcell.org/content/16/4/1008/tab-figures-data (Accessed: 18 August 2017).
277. Nordborg, M. et al. (2005) ‘The Pattern of Polymorphism in Arabidopsis thaliana’, PLoS Biology. Edited by T. Mitchell-Olds. Public Library of Science, 3(7), p. e196. doi: 10.1371/journal.pbio.0030196.
278. Nordborg, M., Charlesworth, B. and Charlesworth, D. (1996) ‘The effect of recombination on background selection’, Genetical Research. Cambridge University Press, 67(02), p. 159. doi: 10.1017/S0016672300033619.
279. Ohta, K., Shibata, T. and Nicolas, A. (1994) ‘Changes in chromatin structure at recombination initiation sites during yeast meiosis.’, The EMBO journal. European Molecular Biology Organization, 13(23), pp. 5754–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7988571 (Accessed: 11 September 2016).
280. Oliver, S. G. et al. (1992) ‘The complete DNA sequence of yeast chromosome III’, Nature. Nature Publishing Group, 357(6373), pp. 38–46. doi: 10.1038/357038a0.
281. Otto (2007) ‘The Evolutionary Consequences of Polyploidy’, Cell. Cell Press, 131(3), pp. 452–462. doi: 10.1016/J.CELL.2007.10.022.

282. Otto, S. (2009) 'The Evolutionary Enigma of Sex', *The American Naturalist*, 174 (S1), pp. 1-14
283. Otto, S. P. and Barton, N. H. (2001) 'Selection for recombination in small populations.', Evolution; international journal of organic evolution, 55(10), pp. 1921–1931. doi: 10.2307/2680441.
284. Otto, S. P. and Lenormand, T. (2002) 'Resolving the paradox of sex and recombination', Nature Reviews Genetics, 3(4), pp. 252–261. doi: 10.1038/nrg761.
285. Page, D. and Grossniklaus, U. (2002) 'The art and design of genetic screens: Arabidopsis thaliana', Nature Reviews Genetics, 3, pp. 124–136. doi: 10.1038/nrg730.
286. Page, S. L. and Hawley, R. S. (2004) 'THE GENETICS AND MOLECULAR BIOLOGY OF THE SYNAPTONEMAL COMPLEX', Annual Review of Cell and Developmental Biology. Annual Reviews, 20(1), pp. 525–558. doi: 10.1146/annurev.cellbio.19.111301.155141.
287. Paigen, K. et al. (2008) 'The Recombinational Anatomy of a Mouse Chromosome', PLoS Genetics. Edited by M. Lichten. Public Library of Science, 4(7), p. e1000119. doi: 10.1371/journal.pgen.1000119.
288. Pan, J. et al. (2011) 'A Hierarchical Combination of Factors Shapes the Genome-wide Topography of Yeast Meiotic Recombination Initiation', Cell, 144(5), pp. 719–731. doi: 10.1016/j.cell.2011.02.009.
289. Pan, Q. et al. (2017) 'Complexity of genetic mechanisms conferring nonuniformity of recombination in maize', Scientific Reports. Nature Publishing Group, 7(1), p. 1205. doi: 10.1038/s41598-017-01240-2.
290. Panizza, S. et al. (2011) 'Spo11-accessory proteins link double-strand break sites to the chromosome axis in early meiotic recombination.', Cell, 146(3), pp. 372–83. doi: 10.1016/j.cell.2011.07.003.
291. Pâques, F. and Haber, J. E. (1999) 'Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae.', Microbiology and molecular biology reviews: MMBR, 63(2), pp. 349–404. doi: .
292. Paran, I. and Zamir, D. (2003) 'Quantitative traits in plants: Beyond the QTL', Trends in Genetics, 19(6), pp. 303–306. doi: 10.1016/S0168-9525(03)00117-3.
293. Parket, A., Inbar, O. and Kupiec, M. (1995) 'Recombination of Ty elements in yeast can be induced by a double-strand break.', Genetics, 140(1).
294. Parniske (1997) 'Novel Disease Resistance Specificities Result from Sequence Exchange between Tandemly Repeated Genes at the Cf-4/9 Locus of Tomato', Cell. Cell Press, 91(6), pp. 821–832. doi: 10.1016/S0092-8674(00)80470-5.
295. Parvanov, E. D., Petkov, P. M. and Paigen, K. (2010) 'Prdm9 Controls Activation of Mammalian Recombination Hotspots', Science, 327(5967).
296. Pasternak, T. et al. (2005) 'Morphogenic effects of abiotic stress: reorientation of growth in Arabidopsis thaliana seedlings', Environmental and Experimental Botany. Elsevier, 53(3), pp. 299–314. doi: 10.1016/J.ENVEXPBOT.2004.04.009.
297. Pecinka, A. et al. (2011) 'Polyploidization increases meiotic recombination frequency in Arabidopsis', BMC Biology. BioMed Central, 9(1), p. 24. doi: 10.1186/1741-7007-9-24.
298. Perrella, G. et al. (2010) 'Histone hyperacetylation affects meiotic recombination and chromosome segregation in Arabidopsis.', The Plant journal : for cell and molecular biology, 62(5), pp. 796–806. doi: 10.1111/j.1365-313X.2010.04191.x.
299. Petes, T. D. (2001) 'Meiotic recombination hot spots and cold spots', Nature Reviews Genetics. Nature Publishing Group, 2(5), pp. 360–369. doi: 10.1038/35072078.

300. Pignatta, D. et al. (2014) ‘Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting.’, eLife. eLife Sciences Publications, Ltd, 3, p. e03198. doi: 10.7554/eLife.03198.
301. Prakash, R. et al. (2009) ‘Yeast Mphl helicase dissociates Rad51-made D-loops: Implications for crossover control in mitotic recombination’, Genes and Development, 23(1), pp. 67–79. doi: 10.1101/gad.1737809.
302. Prakash, S. and Chopra, V. L. (1988) ‘Introgression of Resistance to Shattering in Brassica napus from Brassica juncea through Non-Homologous Recombination’, Plant Breeding, 101(2), pp. 167–168. doi: 10.1111/j.1439-0523.1988.tb00283.x.
303. Prasad, B. D., Goel, S. and Krishna, P. (2010) ‘In Silico Identification of Carboxylate Clamp Type Tetratricopeptide Repeat Proteins in Arabidopsis and Rice As Putative Co-Chaperones of Hsp90/Hsp70’, PLoS ONE. Edited by K. Schumacher. Public Library of Science, 5(9), p. e12761. doi: 10.1371/journal.pone.0012761.
304. Ptak, S. E. et al. (2005) ‘Fine-scale recombination patterns differ between chimpanzees and humans’, Nature Genetics. doi: 10.1038/ng1529.
305. Qiao, H. et al. (2014) ‘Antagonistic roles of ubiquitin ligase HEI10 and SUMO ligase RNF212 regulate meiotic recombination’, Nature Genetics. Nature Research, 46(2), pp. 194–199. doi: 10.1038/ng.2858.
306. Ramesh, M. A., Malik, S.-B. and Logsdon, J. M. (2005) ‘A Phylogenomic Inventory of Meiotic Genes: Evidence for Sex in Giardia and an Early Eukaryotic Origin of Meiosis’, Current Biology. Cell Press, 15(2), pp. 185–191. doi: 10.1016/J.CUB.2005.01.003.
307. Rao, H. B. D. P. et al. (2017) ‘A SUMO-ubiquitin relay recruits proteasomes to chromosome axes to regulate meiotic recombination.’, Science (New York, N.Y.). American Association for the Advancement of Science, 355(6323), pp. 403–407. doi: 10.1126/science.aaf6407.
308. Rauwolf, U. et al. (2011) ‘Uncoupling of sexual reproduction from homologous recombination in homozygous Oenothera species’, Heredity 2011 107:1. Nature Publishing Group, 107(1), p. 87. doi: 10.1038/hdy.2010.171.
309. Rees, H. (1961) ‘Genotypic control of chromosome form and behaviour’, The Botanical Review. Springer-Verlag, 27(2), pp. 288–318. doi: 10.1007/BF02860084.
310. Rees, H. and Thompson, J. B. (1956) ‘Genotypic control of chromosome behaviour in rye. III. Chiasma frequency in homozygotes and heterozygotes’, Heredity, 10(3), pp. 409–424. doi: 10.1038/hdy.1956.34.
311. Reynolds, A. et al. (2013) ‘RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis.’, Nature genetics. NIH Public Access, 45(3), pp. 269–78. doi: 10.1038/ng.2541.
312. RHOADES, M. (1968). Studies on the cytological basis of crossing-over. Peacok, W. J. and Brock, R. D. (eds). Repli- cation and Recombination of Genetic Material. Australian Academy of Science, Canberra, pp. 229—241.
313. Rice, W. R. (2002) ‘Experimental tests of the adaptive significance of sexual recombination’, Nature Reviews Genetics. Nature Publishing Group, 3(4), pp. 241–251. doi: 10.1038/nrg760.
314. Rick, C. M. (1969) ‘Controlled Introgression of Chromosomes of SOLANUM PENNELLII into LYCOPERSICON ESCULENTUM: Segregation and Recombination.’, Genetics. Genetics Society of America, 62(4), pp. 753–68. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17248457 (Accessed: 1 August 2016).
315. Rieseberg, L. H. (2001) ‘Chromosomal rearrangements and speciation’, Trends in Ecology & Evolution, 16(7), pp. 351–358. doi: 10.1016/S0169-5347(01)02187-5.

316. Roberts, M. A. et al. (1999) ‘Induction and Characterization of Ph1 Wheat Mutants’, Genetics, 153, pp. 1909–1918. Available at: http://www.genetics.org/content/genetics/153/4/1909.full.pdf (Accessed: 30 September 2017).
317. Rodgers-Melnick, E. et al. (2015) ‘Recombination in diverse maize is stable, predictable, and associated with genetic load’, Proceedings of the National Academy of Sciences, 112(12), pp. 3823–2828. doi: 10.1073/pnas.1413864112.
318. Roeder, G. S. (1995) ‘Review Sex and the single cell: Meiosis in yeast’, 92, pp. 10450–10456. Available at: http://www.pnas.org/content/pnas/92/23/10450.full.pdf (Accessed: 21 April 2018).
319. Roeder, G. S. (1997) ‘Meiotic chromosomes: it takes two to tango’, Genes and Development, 11, pp. 2600–2621. Available at: http://genesdev.cshlp.org/content/11/20/2600.full.pdf (Accessed: 22 August 2017).
320. Romanienko, P. J. and Camerini-Otero, R. D. (2000) ‘The Mouse Spo11 Gene Is Required for Meiotic Chromosome Synapsis’, Molecular Cell, 6(5), pp. 975–987. doi: 10.1016/S1097-2765(00)00097-6.
321. Ronceret, A. et al. (2009) ‘PHS1 regulates meiotic recombination and homologous chromosome pairing by controlling the transport of RAD50 to the nucleus.’, Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 106(47), pp. 20121–6. doi: 10.1073/pnas.0906273106.
322. Ross, K. J. et al. (1997) ‘Cytological characterization of four meiotic mutants of Arabidopsis isolated from T-DNA-transformed lines’, Chromosome Research, 5(8), pp. 551–559. doi: 10.1023/A:1018497804129.
323. Roze, D. and Barton, N. H. (2006) ‘The Hill-Robertson effect and the evolution of recombination.’, Genetics, 173(3), pp. 1793–811. doi: 10.1534/genetics.106.058586.
324. Rybnikov, S. R., Frenkel, Z. M. and Korol, A. B. (2017) ‘What drives the evolution of condition-dependent recombination in diploids? Some insights from simulation modelling.’, Philosophical transactions of the Royal Society of London. Series B, Biological sciences. The Royal Society, 372(1736), p. 20160460. doi: 10.1098/rstb.2016.0460.
325. Saintenac, C. et al. (2009) ‘Detailed recombination studies along chromosome 3B provide new insights on crossover distribution in wheat (Triticum aestivum L.).’, Genetics, 181(2), pp. 393–403. doi: 10.1534/genetics.108.097469.
326. Salathé, M., Kouyos, R. D. and Bonhoeffer, S. (2009) ‘On the causes of selection for recombination underlying the red queen hypothesis.’, The American naturalist. NIH Public Access, 174 Suppl 1(Suppl 1), pp. S31-42. doi: 10.1086/599085.
327. Saleem, M., Lamb, B. C. and Nevo, E. (2001) ‘Inherited Differences in Crossing Over and Gene Conversion Frequencies Between Wild Strains of Sordaria fimicola From “Evolution Canyon”’, Genetics, 159(4).
328. SÄLL, T. (2008) ‘Genetic control of recombination in barley: Variation in linkage between marker genes’, Hereditas. Blackwell Publishing Ltd, 112(2), pp. 171–178. doi: 10.1111/j.1601-5223.1990.tb00054.x.
329. Salomé, P. A. et al. (2012) ‘The recombination landscape in Arabidopsis thaliana F2 populations’, Heredity. Nature Publishing Group, 108(4), pp. 447–455. doi: 10.1038/hdy.2011.95.
330. Sanchez-Moran, E. et al. (2002) ‘Variation in Chiasma Frequency Among Eight Accessions of Arabidopsis thaliana’, Genetics, 162(3), pp. 1415–1422. Available at: http://www.genetics.org/content/162/3/1415.short (Accessed: 22 January 2015).

331. Sanchez-Moran, E. et al. (2007) ‘ASY1 mediates AtDMC1-dependent interhomolog recombination during meiosis in Arabidopsis’, Genes and Development, 21, pp. 2220–2233. doi: 10.1101/gad.439007.
332. Sanchez-Moran, E. et al. (2008) ‘ASY1 coordinates early events in the plant meiotic recombination pathway.’, Cytogenetic and genome research. Karger Publishers, 120(3–4), pp. 302–12. doi: 10.1159/000121079.
333. Sandor, C. et al. (2012) ‘Genetic Variants in REC8, RNF212, and PRDM9 Influence Male Recombination in Cattle’, PLoS Genetics. Edited by K. Paigen. Public Library of Science, 8(7), p. e1002854. doi: 10.1371/journal.pgen.1002854.
334. Sasaki, M., Lange, J. and Keeney, S. (2010) ‘Genome destabilization by homologous recombination in the germ line.’, Nature reviews. Molecular cell biology. NIH Public Access, 11(3), pp. 182–95. doi: 10.1038/nrm2849.
335. Schnable, P. S., Hsia, A.-P. and Nikolau, B. J. (1998) ‘Genetic recombination in plants’, Current Opinion in Plant Biology, 1(2), pp. 123–129. doi: 10.1016/S1369-5266(98)80013-7.
336. Schultes, N. P. and Szostak, J. W. (1991) ‘A poly(dA.dT) tract is a component of the recombination initiation site at the ARG4 locus in Saccharomyces cerevisiae.’, Molecular and cellular biology. American Society for Microbiology, 11(1), pp. 322–8. doi: 10.1128/MCB.11.1.322.
337. Schwacha, A. et al. (1997) ‘Interhomolog bias during meiotic recombination: meiotic functions promote a highly differentiated interhomolog-only pathway.’, Cell. Elsevier, 90(6), pp. 1123–35. doi: 10.1016/S0092-8674(00)80378-5.
338. Schwacha, A. and Kleckner, N. (1995) ‘Identification of double holliday junctions as intermediates in meiotic recombination’, Cell, 83(5), pp. 783–791. doi: 10.1016/0092-8674(95)90191-4.
339. Schwander, T., Libbrecht, R. and Keller, L. (2014) ‘Supergenes and Complex Phenotypes’, Current Biology, 24(7), pp. R288–R294. doi: 10.1016/j.cub.2014.01.056.
340. Schweighofer, A., Hirt, H. and Meskiene, I. (2004) ‘Plant PP2C phosphatases: Emerging functions in stress signaling’, Trends in Plant Science, 9(5), pp. 236–243. doi: 10.1016/j.tplants.2004.03.007.
341. Sebastian, R. L. et al. (2000) ‘An integrated AFLP and RFLP Brassica oleracea linkage map from two morphologically distinct doubled-haploid mapping populations’, TAG Theoretical and Applied Genetics. Springer-Verlag, 100(1), pp. 75–81. doi: 10.1007/s001220050011.
342. Segal (2009) ‘What controls nucleosome positions?’, Trends in Genetics. Elsevier Current Trends, 25(8), pp. 335–343. doi: 10.1016/J.TIG.2009.06.002.
343. Séguéla-Arnaud, M. et al. (2015) ‘Multiple mechanisms limit meiotic crossovers: TOP3α and two BLM homologs antagonize crossovers in parallel to FANCM.’, Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 112(15), pp. 4713–8. doi: 10.1073/pnas.1423107112.
344. Séguéla-Arnaud, M. et al. (2016) ‘RMI1 and TOP3α limit meiotic CO formation through their C-terminal domains’, Nucleic Acids Research. Oxford University Press, 287(4), p. gkw1210. doi: 10.1093/nar/gkw1210.
345. Serra, H. et al. (bioRxiv) ‘Massive crossover elevation via combination of HEI10 and recq4a recq4b during Arabidopsis meiosis’. doi: 10.1101/159764.
346. Sessions, A. et al. (2002) ‘A high-throughput Arabidopsis reverse genetics system.’, The Plant cell. American Society of Plant Biologists, 14(12), pp. 2985–94. doi: 10.1105/tpc.004630.

347. Shapiro, H. M. (2003) Parameters and Probes, Practical Flow Cytometry. doi: 10.1002/0471722731.ch7.
348. Sharbel, T., Haubold, B. and Mitchell-Olds, T. (2000) 'Genetic isolation by distance in Arabidopsis thaliana: Biogeography and postglacial colonization of Europe', Molecular Ecology, (January 2001), pp. 2109–2118. doi: 10.1046/j.1365-294X.2000.01122.x.
349. Sherman, J. D. and Stack, S. M. (1995) 'Two-Dimensional Spreads of Synaptonemal Complexes from Solanaceous Plants. VI. High-Resolution Recombination Nodule Map for Tomato (Lycofiersicon esculentum)', Genetics, 141, pp. 683–708. Available at: http://www.genetics.org/content/genetics/141/2/683.full.pdf (Accessed: 5 September 2017).
350. Shilo, S. et al. (2015) 'DNA Crossover Motifs Associated with Epigenetic Modifications Delineate Open Chromatin Regions in Arabidopsis', The Plant Cell, 27(9), p. tpc.15.00391. doi: 10.1105/tpc.15.00391.
351. Sidhu, G. K. et al. (2015) 'Recombination patterns in maize reveal limits to crossover homeostasis.', Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 112(52), pp. 15982–7. doi: 10.1073/pnas.1514265112.
352. Simon, M. et al. (2008) 'Quantitative trait loci mapping in five new large recombinant inbred line populations of Arabidopsis thaliana genotyped with consensus single-nucleotide polymorphism markers.', Genetics, 178(4), pp. 2253–64. doi: 10.1534/genetics.107.083899.
353. Singh, N. D. et al. (2015) 'EVOLUTION. Fruit flies diversify their offspring in response to parasite infection.', Science (New York, N.Y.). American Association for the Advancement of Science, 349(6249), pp. 747–50. doi: 10.1126/science.aab1768.
354. Slotkin, R. K. and Martienssen, R. (2007) 'Transposable elements and the epigenetic regulation of the genome', Nature Reviews Genetics. Nature Publishing Group, 8(4), pp. 272–285. doi: 10.1038/nrg2072.
355. Smagulova, F. et al. (2011) 'Genome-wide analysis reveals novel molecular features of mouse recombination hotspots', Nature. Nature Research, 472(7343), pp. 375–378. doi: 10.1038/nature09869.
356. Smith, J. M. and Haigh, J. (1974) 'The hitch-hiking effect of a favourable gene', Genet. Res., Camb, 23, pp. 23–35. Available at: https://courses.app.ist.ac.at/fileadmin/user_upload/BIO-515-S17/MaynardSmith___Haigh_1974_GenetResCam.pdf (Accessed: 19 April 2018).
357. Smukowski, C. S. and Noor, M. A. F. (2011) 'Recombination rate variation in closely related species', Heredity. Nature Publishing Group, 107(6), pp. 496–508. doi: 10.1038/hdy.2011.44.
358. Snowden (2004) 'hMSH4-hMSH5 Recognizes Holliday Junctions and Forms a Meiosis-Specific Sliding Clamp that Embraces Homologous Chromosomes', Molecular Cell. Cell Press, 15(3), pp. 437–451. doi: 10.1016/J.MOLCEL.2004.06.040.
359. Sommermeyer, V. et al. (2013) 'Spp1, a Member of the Set1 Complex, Promotes Meiotic DSB Formation in Promoters by Tethering Histone H3K4 Methylation Sites to Chromosome Axes', Molecular Cell. Cell Press, 49(1), pp. 43–54. doi: 10.1016/J.MOLCEL.2012.11.008.
360. Stapley, J. et al. (2017) 'Variation in recombination frequency and distribution across eukaryotes: patterns and processes', Philosophical Transactions B Royal Society, 372. Available at: http://dx.doi.org/10.1098/rstb.2016.0455 (Accessed: 13 April 2018).

361. Stebbins, G. L. et al. (1946) ‘Artificial and Natural Hybrids in the Gramineae, Tribe Hordeae I. Elymus, Sitanion, and Agropyron’, American Journal of Botany. Botanical Society of America, Inc., 33(5), p. 338. doi: 10.2307/2437121.
362. Stebbins, G. L. (1950) Variation and Evolution in Plants. Columbia University Press, New York.
363. Stern, C. (1926) ‘An Effect of Temperature and Age on Crossing-Over in the First Chromosome of Drosophila Melanogaster.’, Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 12(8), pp. 530–2. doi: 10.1073/PNAS.12.8.530.
364. Stevison, L. S. et al. (2017) ‘Recombination rate plasticity: revealing mechanisms by design.’, Philosophical transactions of the Royal Society of London. Series B, Biological sciences. The Royal Society, 372(1736), p. 20160459. doi: 10.1098/rstb.2016.0459.
365. DE STORME, N. and GEELEN, D. (2014) ‘The impact of environmental stress on male reproductive development in plants: biological processes and molecular mechanisms’, Plant, Cell & Environment. Wiley/Blackwell (10.1111), 37(1), pp. 1–18. doi: 10.1111/pce.12142.
366. Sudupak, M. A., Bennetzen, J. L. and Hulbert, S. H. (1993) ‘Unequal exchange and meiotic instability of disease-resistance genes in the Rp1 region of maize.’, Genetics, 133(1).
367. Sun, H., Treco, D. and Szostak, J. W. (1991) ‘Extensive 3’-overhanging, single-stranded DNA associated with the meiosis-specific double-strand breaks at the ARG4 recombination initiation site’, Cell, 64(6), pp. 1155–1161. doi: 10.1016/0092-8674(91)90270-9.
368. Sun, W. et al. (2008) ‘The FANCM Ortholog Fml1 Promotes Recombination at Stalled Replication Forks and Limits Crossing Over during DNA Double-Strand Break Repair’, Molecular Cell. Elsevier Inc., 32(1), pp. 118–128. doi: 10.1016/j.molcel.2008.08.024.
369. Suzuki, N. et al. (2014) ‘Abiotic and biotic stress combinations’, New Phytologist, 203. doi: 10.1111/nph.12797.
370. Swarbreck, D. et al. (2008) ‘The Arabidopsis Information Resource (TAIR): gene structure and function annotation.’, Nucleic acids research. Oxford University Press, 36(Database issue), pp. D1009-14. doi: 10.1093/nar/gkm965.
371. Symonds, V. V. et al. (2005) ‘Mapping quantitative trait loci in multiple populations of Arabidopsis thaliana identifies natural allelic variation for trichome density.’, Genetics, 169(3), pp. 1649–58. doi: 10.1534/genetics.104.031948.
372. Takebayashi, N. and Morrell, P. L. (2001) ‘Is Self-Fertilization an Evolutionary Dead End? Revisiting an Old Hypothesis with Genetic Theories and a Macroevolutionary Approach’, American Journal of Botany. Botanical Society of America, 88(7), p. 1143. doi: 10.2307/3558325.
373. Tang, S. et al. (2015) ‘Pervasive and essential roles of the top3-rmi1 decatenase orchestrate recombination and facilitate chromosome segregation in meiosis’, Molecular Cell. Elsevier Inc., 57(4), pp. 607–621. doi: 10.1016/j.molcel.2015.01.021.
374. Tenaillon, M. I. et al. (2002) ‘Patterns of Diversity and Recombination Along Chromosome 1 of Maize (Zea mays ssp. mays L.)’, Genetics, 162(3). Available at: http://www.genetics.org/content/162/3/1401.full#ref-15 (Accessed: 5 September 2017).
375. Tessadori, F. et al. (2009) ‘Phytochrome B and histone deacetylase 6 control light-induced chromatin compaction in Arabidopsis thaliana.’, PLoS genetics, 5(9), p. e1000638. doi: 10.1371/journal.pgen.1000638.
376. Thompson, M. J. and Jiggins, C. D. (2014) ‘Supergenes and their role in evolution’, Heredity. Nature Publishing Group, 113(1), pp. 1–8. doi: 10.1038/hdy.2014.20.

377. Timmermans, M. C. P. et al. (1997) 'Region-Specific Cis- and Trans-Acting Factors Contribute to Genetic Variability in Meiotic Recombination in Maize', Genetics, 146(3). Available at: http://www.genetics.org/content/146/3/1101.short (Accessed: 4 August 2017).
378. Ungerer, M. C. et al. (2003) 'Genotype-Environment Interactions at Quantitative Trait Loci Affecting Inflorescence Development in Arabidopsis thaliana', Genetics, 165(1). Available at: http://www.genetics.org/content/165/1/353.short (Accessed: 5 August 2017).
379. de Vicente, M. C. and Tanksley, S. D. (1991) 'Genome-wide reduction in recombination of backcross progeny derived from male versus female gametes in an interspecific cross of tomato', Theoretical and Applied Genetics. Springer-Verlag, 83(2), pp. 173–178. doi: 10.1007/BF00226248.
380. Villeneuve (2001) 'Whence Meiosis?', Cell. Cell Press, 106(6), pp. 647–650. doi: 10.1016/S0092-8674(01)00500-1.
381. Vincenten, N. et al. (2015) 'The kinetochore prevents centromere-proximal crossover recombination during meiosis.', eLife. eLife Sciences Publications Limited, 4, p. e10850. doi: 10.7554/eLife.10850.
382. Wada, K. C. and Takeno, K. (2010) 'Stress-induced flowering', Plant Signaling & Behavior. Taylor & Francis, 5(8), pp. 944–947. doi: 10.4161/psb.5.8.11826.
383. Ward, J. O. et al. (2007) 'Mutation in Mouse Hei10, an E3 Ubiquitin Ligase, Disrupts Meiotic Crossing Over', PLoS Genetics. Public Library of Science, 3(8), p. e139. doi: 10.1371/journal.pgen.0030139.
384. Webster, M. T. and Hurst, L. D. (2012) 'Direct and indirect consequences of meiotic recombination: implications for genome evolution', Trends in Genetics, 28(3), pp. 101–109. doi: 10.1016/j.tig.2011.11.002.
385. Weinig, C. et al. (2002) 'Novel Loci Control Variation in Reproductive Timing in Arabidopsis thaliana in Natural Environments', Genetics, 162. Available at: http://www.genetics.org/content/genetics/162/4/1875.full.pdf (Accessed: 30 March 2018).
386. Westergaard, M. and Wettstein, D. V (1972) 'The Synaptinemal Complex', Annual Review of Genetics. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA, 6(1), pp. 71–110. doi: 10.1146/annurev.ge.06.120172.000443.
387. White, M. J. and Morley, F. H. (1955) 'Effects of Pericentric Rearrangements on Recombination in Grasshopper Chromosomes.', Genetics. Genetics Society of America, 40(5), pp. 604–19. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17247575 (Accessed: 30 September 2017).
388. Wijnker, E. et al. (2013) 'The genomic landscape of meiotic crossovers and gene conversions in Arabidopsis thaliana.', eLife, 2, p. e01426. doi: 10.7554/eLife.01426.
389. Wijnker, E. and de Jong, H. (2008) 'Managing meiotic recombination in plant breeding', Trends in Plant Science, 13(12), pp. 640–646. doi: 10.1016/j.tplants.2008.09.004.
390. Williams, C. G., Goodman, M. M. and Stuber, C. W. (1995) 'Comparative recombination distances among Zea mays L. inbreds, wide crosses and interspecific hybrids.', Genetics, 141(4). Available at: http://www.genetics.org/content/141/4/1573.short (Accessed: 4 August 2017).
391. Winckler, W. et al. (2005) 'Comparison of fine-scale recombination rates in humans and chimpanzees.', Science (New York, N.Y.). American Association for the Advancement of Science, 308(5718), pp. 107–11. doi: 10.1126/science.1105322.
392. Wright, K. M. et al. (2015) 'Selection on Meiosis Genes in Diploid and Tetraploid Arabidopsis arenosa', Molecular Biology and Evolution. Oxford University Press, 32(4), pp. 944–955. doi: 10.1093/molbev/msu398.

393. Wright, S. I., Agrawal, N. and Bureau, T. E. (2003) 'Effects of Recombination Rate and Gene Density on Transposable Element Distributions in Arabidopsis thaliana', Genome Research. Cold Spring Harbor Laboratory Press, 13(8), pp. 1897–1903. doi: 10.1101/GR.1281503.
394. Wu, G. et al. (2015) 'Traffic lines: new tools for genetic analysis in Arabidopsis thaliana.', Genetics. Genetics, 200(1), pp. 35–45. doi: 10.1534/genetics.114.173435.
395. Wu, L. and Hickson, I. D. (2003) 'The Bloom's syndrome helicase suppresses crossing over during homologous recombination', Nature, 426(6968), pp. 870–874. doi: 10.1038/nature02253.
396. Wu, T. et al. (1994) 'Meiosis-induced double-strand break sites determined by yeast chromatin structure', Science. American Association for the Advancement of Science, 263(5146), pp. 515–518. doi: 10.1126/science.8290959.
397. Yamada, T. et al. (2004) 'Roles of histone acetylation and chromatin remodeling factor in a meiotic recombination hotspot.', The EMBO journal. EMBO Press, 23(8), pp. 1792–803. doi: 10.1038/sj.emboj.7600138.
398. Yandeau-Nelson, M. D., Nikolau, B. J. and Schnable, P. S. (2006) 'Effects of trans-acting genetic modifiers on meiotic recombination across the a1-sh2 interval of maize.', Genetics, 174(1), pp. 101–12. doi: 10.1534/genetics.105.049270.
399. Yang, S. et al. (2015) 'Parent–progeny sequencing indicates higher mutation rates in heterozygotes', Nature. Nature Publishing Group. doi: 10.1038/nature14649.
400. Yant, L. et al. (2013) 'Meiotic adaptation to genome duplication in Arabidopsis arenosa', Current biology: CB. Elsevier Ltd, 23(21), pp. 2151–2156. doi: 10.1016/j.cub.2013.08.059.
401. Yant, L. and Bomblies, K. (2015) 'Genome management and mismanagement--cell-level opportunities and challenges of whole-genome duplication.', Genes & development. Cold Spring Harbor Laboratory Press, 29(23), pp. 2405–19. doi: 10.1101/gad.271072.115.
402. Yao, H. et al. (2002) 'Molecular characterization of meiotic recombination across the 140-kb multigenic a1-sh2 interval of maize', Proceedings of the National Academy of Sciences, 99(9), pp. 6157–6162. doi: 10.1073/pnas.082562199.
403. Yao, H. and Schnable, P. S. (2005) 'Cis-effects on meiotic recombination across distinct a1-sh2 intervals in a common Zea genetic background.', Genetics, 170(4), pp. 1929–44. doi: 10.1534/genetics.104.034454.
404. Yelina, N. E. et al. (2012) 'Epigenetic Remodeling of Meiotic Crossover Frequency in Arabidopsis thaliana DNA Methyltransferase Mutants', PLoS Genetics, 8(8). doi: 10.1371/journal.pgen.1002844.
405. Yelina, N. E. et al. (2013) 'High-throughput analysis of meiotic crossover frequency and interference via flow cytometry of fluorescent pollen in Arabidopsis thaliana.', Nature protocols. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 8(11), pp. 2119–34. doi: 10.1038/nprot.2013.131.
406. Yelina, N. E. et al. (2015) 'DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in Arabidopsis.', Genes & development, 29(20), pp. 2183–202. doi: 10.1101/gad.270876.115.
407. Youds, J. L. and Boulton, S. J. (2011) 'The choice in meiosis - defining the factors that influence crossover or non-crossover formation.', Journal of cell science, 124, pp. 501–513. doi: 10.1242/jcs.074427.
408. Zakharyevich, K. et al. (2012) 'Delineation of joint molecule resolution pathways in meiosis identifies a crossover-specific resolvase.', Cell. Springer-Verlag, Heidelberg, 149(2), pp. 334–47. doi: 10.1016/j.cell.2012.03.023.

409. Zarchi, Y. et al. (1972) ‘Chiasmata and the breeding system in wild populations of diploid wheats’, Chromosoma. Springer-Verlag, 38(1), pp. 77–94. doi: 10.1007/BF00319956.
410. Zeng, Z. B. (1993) ‘Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci.’, Proceedings of the National Academy of Sciences, 90(23), pp. 10972–10976. doi: 10.1073/pnas.90.23.10972.
411. Zeng, Z. B. (1994) ‘Precision Mapping of Quantitative Trait Loci’, Genetics, 136, pp. 1457–1468.
412. Zetka, M. C. and Rose, A. M. (1992) ‘The meiotic behavior of an inversion in Caenorhabditis elegans.’, Genetics. Genetics, 131(2), pp. 321–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1644275 (Accessed: 31 July 2016).
413. Zhang, C. et al. (2012) ‘The Arabidopsis thaliana DSB formation (AtDFO) gene is required for meiotic double-strand break formation’, The Plant Journal, 72(2), pp. 271–281. doi: 10.1111/j.1365-313X.2012.05075.x.
414. Zhang, L., Liang, Z., et al. (2014) ‘Crossover Patterning by the Beam-Film Model: Analysis and Implications’, PLoS Genetics, 10(1). doi: 10.1371/journal.pgen.1004042.
415. Zhang, L., Wang, S., et al. (2014) ‘Topoisomerase II mediates meiotic crossover interference’, Nature. Nature Publishing Group, 511, pp. 551–556. doi: 10.1038/nature13442.
416. Zhang, X. et al. (2006) ‘Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method’, Nature Protocols. Nature Publishing Group, 1(2), pp. 641–646. doi: 10.1038/nprot.2006.97.
417. Zhuchenko, A. A., et al. (1986) ‘The relation between genotype adaptivity and reactivity of its recombination characteristics under temperature effect’, *Genetika,* 22(6), 966-974.
418. Zibran, M. F. (2015) ‘Chi-squared test of independence’, Handbook of Biological Statistics, pp. 1–7. Available at: http://www.biostathandbook.com/chiind.html.
419. Zickler, D. et al. (1999) ‘Meiotic chromosomes: integrating structure and function.’, Annual review of genetics. EMBO Press, 33(1), pp. 603–754. doi: 10.1146/annurev.genet.33.1.603.
420. Zickler, D. (2006) ‘From early homologue recognition to synaptonemal complex formation’, Chromosoma. Springer-Verlag, 115(3), pp. 158–174. doi: 10.1007/s00412-006-0048-6.
421. Ziolkowski, P. A. et al. (2015) ‘Juxtaposition of heterozygous and homozygous regions causes reciprocal crossover remodelling via interference during Arabidopsis meiosis.’, eLife. eLife Sciences Publications Limited, 4, p. e03708. doi: 10.7554/eLife.03708.
422. Ziolkowski, P. A. et al. (2017) ‘Natural variation and dosage of the HEI10 meiotic E3 ligase control Arabidopsis crossover recombination.’, Genes & development. Cold Spring Harbor Laboratory Press, 31(3), pp. 306–317. doi: 10.1101/gad.295501.116.
423. Ziolkowski, P. A. and Henderson, I. R. (2017) ‘Interconnections between meiotic recombination and sequence polymorphism in plant genomes’, New Phytologist, 213(3), pp. 1022–1029. doi: 10.1111/nph.14265.