Using breeding and quantitative genetics to understand the C_4 pathway

Conor J. C. Simpson[1*], Gregory Reeves[1*], Anoop Tripathi[1], Pallavi Singh[1], Julian M. Hibberd[1†]

[1]University of Cambridge, Department of Plant Sciences, Downing Street, Cambridge, CB2 3EA, United Kingdom

*These authors contributed equally to this work

[†]Correspondence: jmh65@cam.ac.uk

Keywords: C_4 photosynthesis, natural variation, hybridization, mapping population designs

One sentence summary: Assessing intra- and interspecific variation and hybridization to identify the genetic basis of C_4 traits via quantitative genetics

Highlight: Forward genetics is a mainstay of biology, but has been used less-routinely to understand the complex C_4 trait. Here we review literature relevant to this approach and consider when and how it could be applied to better understand C_4 photosynthesis.

Abstract

Reducing photorespiration in C_3 crops could significantly increase rates of photosynthesis and yield. One method to achieve this would be to integrate C_4 photosynthesis into C_3 species. This objective is challenging as it involves engineering incompletely understood traits into C_3 leaves including complex changes to their biochemistry, cell biology and anatomy. Quantitative genetics and selective breeding offer under-explored routes to identify regulators of these processes. We first review examples of natural intraspecific variation in C_4 photosynthesis as well as the potential for hybridization between C_3 and C_4 species. We then discuss how quantitative genetic approaches including artificial selection and genome-wide association could be used to better understand the C_4 syndrome and in so doing guide the engineering of the C_4 pathway into C_3 crops.

Introduction

Photosynthetic plants provide humanity's food, many textiles, building materials and represent the source of numerous medicines and fuels. Understanding how improvements in photosynthesis could be achieved therefore has the potential to impact on many aspects of human life. Photosynthesis requires the enzyme Ribulose 1,5 Bisphosphate Carboxylase Oxygenase (RuBisCO) to fix atmospheric carbon dioxide (CO_2) into 3-phosphoglycerate (Calvin and Benson, 1948). Species that only use RuBisCO for carbon fixation are known as "C_3" plants, as 3-phosphoglycerate contains three carbon atoms. RuBisCO however is also able to react with oxygen in addition to CO_2. This oxygenation reaction produces the toxic molecule 2-phosphoglycolate, which must be metabolised and recycled via the photorespiratory cycle. Photorespiration leads to loss of carbon fixed by RuBisCO and release of ammonia from amino acids at the expense of both ATP and reducing power (Bowes *et al.*, 1971). Rates of photorespiration typically increase at higher temperatures because under these conditions the oxygenation reaction of RuBisCO is favoured (Portis and Parry, 2007), but photorespiration can also increase during periods of drought when stomatal closure limits CO_2 supply to the RuBisCO active site. In extreme conditions, photorespiratory rates can use approximately 25% of photosynthetic outputs (Sharkey, 1988).

Land plants have evolved two carbon concentrating mechanisms to reduce photorespiration. These are termed Crassulacean Acid Metabolism (CAM) and C_4 photosynthesis. Whilst in both cases rates of photorespiration are reduced because compared with the C_3 state, approximately tenfold higher concentrations of CO_2 are supplied to RuBisCO, CAM and C_4 species use temporal and spatial systems respectively. It is estimated that the C_4 pathway has evolved independently from C_3 ancestors at least 60 times to yield numerous phenotypes that concentrate CO_2 around RuBisCO (Sage *et al.*, 2011). In all cases, in the C_4 leaf RuBisCO dependent fixation of CO_2 takes place in a specific compartment supplied with high concentrations of CO_2 such that the oxygenase activity of RuBisCO is almost completely abolished (Fig. 1A). In most C_4 species photosynthesis is compartmented between two cell-types so that they are unified by a general pathway in which CO_2 is converted to bicarbonate (HCO_3^-) by Carbonic Anhydrase in mesophyll cells, and then combined with the 3-carbon molecule phospho*enol*pyruvate by the enzyme Phospho*enol*yruvate Carboxylase

(PEPC) into the 4-carbon molecule oxaloacetate (Fig. 1A). Oxaloacetate is then reduced to either malate or transaminated to aspartate. After diffusing to an adjacent cell layer such as the bundle or mestome sheath, malate or aspartate are decarboxylated such that high concentrations of CO_2 accumulate around RuBisCO and so allow high rates of carboxylation (Fig. 1A). Finally, in species that use NAD-ME or NADP-ME to release CO_2 around RuBisCO, the 3-carbon molecule produced from decarboxylation is regenerated to phospho*enol*pyruvate (PEP) in mesophyll cells by Pyruvate,Orthophosphate Dikinase (PPDK) to continue the cycle (Fig. 1A).

Traits underpinning C_4 photosynthesis vary widely between species (Edwards and Voznesenskaya, 2011; Furbank, 2011; Sage and Stata, 2015; Sedelnikova *et al.*, 2018). This interspecific variation in C_4 traits includes differences in leaf anatomy, cell biology and biochemistry as well as the patterns of gene expression that determine these characteristics. For example, the cell-types and arrangement of veins used by C_4 species varies between lineages that have independently evolved the pathway (Fig. 1B). At least nine anatomical types have been described in the grasses (Poaceae) (Edwards and Voznesenskaya, 2011). Examples of this variation include in the number of layers of mestome and/or bundle sheath cells, and whether RuBisCO is compartmented into the bundle or the mestome sheath. Although much of this variation associated with C_4 photosynthesis is found in lineages that are separated by deep evolutionary time, Kranz anatomy also differs in species within families including the Amaranthaceae (Kadereit *et al.*, 2003; Muhaidat *et al.*, 2007; Sage, 2016), Asteraceae (Peter and Katinas, 2003), Cleomaceae (Koteyeva *et al.*, 2011), Portulaceae (Voznesenskaya *et al.*, 2017) and Poaceae (Ohsugi and Murata, 1985; Edwards and Voznesenskaya, 2011). Of the approximately 8,100 C_4 species defined to date, six operate the C_4 pathway in a single cell (Fig. 1B). In these single-celled C_4 species the pathway is distributed between separate populations of chloroplasts such that the cell biology of these species has been modified compared with the C_3 state. However, modifications to the cell biology of C_4 leaves is not restricted to these single-cell species. In C_4 species that separate photosynthesis between two cell types, plasmodesmatal frequency is increased compared with the C_3 state (Botha, 1992; Danila *et al.*, 2016). Some lineages contain suberin in the bundle sheath cell wall whilst others do not (Mertz and Brutnell, 2014) and whilst some C_4 lineages arrange chloroplasts in bundle sheath

89 cells centripetally, others do this centrifugally with respect to the veins (Edwards and
90 Voznesenskaya, 2011).

91 Lastly, soon after the discovery of C_4 photosynthesis, differences in the biochemistry of the
92 pathway were discovered among C_4 species (Hatch *et al.* 1975). These different pathways were
93 termed C_4 "sub-types" due to the fact that decarboxylation is associated with three separate C_4 acid
94 decarboxylases, NADP-dependent Malic Enzyme (NADP-ME), NAD-dependent Malic Enzyme
95 (NAD-ME), and Phospho*enol*pyruvate Carboxykinase (PEPCK). Although there is growing support
96 for the notion that species can modify the extent to which each C_4 acid decarboxylases is engaged
97 (Omoto *et al.*, 2012; Sharwood *et al.*, 2014; Sales *et al.*, 2018), the differences in biochemistry
98 associated with the sub-types exemplify the fact that the C_4 pathway is a convergent phenomenon,
99 and that its operation varies between species.

100 The differences in leaf anatomy, cell biology and biochemistry between independent C_4 lineages
101 have frequently been summarised (Edwards and Voznesenskaya, 2011; Sage, 2016). In contrast,
102 there have been fewer recent attempts to synthesise the literature relating to forced hybridisations
103 between C_3 and C_4 species. Studies have included somatic hybridisations of phylogenetically distant
104 C_3 and C_4 plants, as well as sexual hybridisations of congeneric species. Whilst these wide
105 hybridisations have provided insight into the extent to which C_4 traits can be maintained and inherited
106 in C_3 species, a growing body of evidence documents variation in C_4 traits within a species. We
107 summarise examples of this work and suggest that there are opportunities to use quantitative trait
108 mapping to better understand the C_4 pathway. Not only could these classical approaches provide
109 insight into the evolution and genetic basis of C_4 photosynthesis, they may also inform efforts to
110 engineer more efficient C_3 crops.

111

112 **Somatic hybridization of C_3 and C_4 species**

113 Approaches such as protoplast fusion allow somatic or asexual hybridisation. Protoplasts from
114 somatic cells from separate species are fused and regenerated into hybrid plants (Carlson *et al.*,
115 1972; Evans, 1983). In many cases, asexual hybridisation can form fertile hybrids between species
116 that are considered sexually incompatible. Attempts to form hybrids via somatic hybridisation of C_3

rice (*Oryza sativa*) and other C_4 grasses have been moderately successful. Terada *et al.* (1987) produced somatic hybrids between rice and C_4 *Echinochloa oryzicola* that were morphologically different from either parent. Some contained 60 chromosomes which corresponded to the full hybrid complement but plants developed necrosis and died before forming roots. Moreover, rice and C_4 *Panicum maximum* (now *Megathyrsus maximus*) were successfully fused to form hybrids with abnormal floral structures with lowered fertility (Xin *et al*., 1997). In all, 28 hybrids flowered but only five set fertile seed. To our knowledge, this work has never been repeated.

There have also been attempts to form hybrids between wheat and C_4 grasses. A cell suspension of Trititrigia (a perennial hybrid of *T. durum* and *Thinopyrum intermedium*) was hybridised with maize (Wang *et al*., 1993; Wang and Niizeki, 1994). Plants that regenerated were aneuploids carrying incomplete sets of chromosomes from both species. Although the progeny were not full hybrids, this study demonstrated that after asexual hybridization maize and *Triticum* chromosomes were not eliminated during successive cell divisions despite the uniparental genome elimination that occurred when both species are hybridised sexually (Laurie and Bennett, 1986, 1989; Laurie *et al*., 1990). Szarka *et al.* (2002) fused a cell suspension of an albino maize mutant with wheat protoplasts. Plants that regenerated resembled maize but were green indicating that photosynthesis from wheat rescued the albino phenotype in maize. Cytological observations showed the plants had all parental chromosomes, but no morphological traits associated with C_4 photosynthesis were detected and although the plants produced male and female flowers, all were sterile (Szarka *et al.* 2002). Independently, Xu *et al.* (2003) reported wheat-maize hybrids that contained nuclear and mitochondrial genomes of both species but plastid DNA only from wheat. These somatic hybrids resembled wheat and although many flowered they were all sterile. This may have been due, at least in part, to the fact that the wheat and maize cell suspension cultures had chromosomal aberrations prior to fusion. Thus, taken as a whole, work on asexual hybridisation of C_3 and C_4 cereals indicates that chromosomes of both photosynthetic types are stable in fused cells. However, in reports such as those from Xu *et al*. (2003) and Szarka *et al.* (2002), plants were not viable after transfer from tissue culture. In contrast, sexual hybridisation of closely related C_3 and C_4 species has in some

144 cases allowed production of fertile plants and their progeny assessed over multiple generations. We
145 address this next.
146
147 **Sexual hybridisation of C_3 and C_4 species**
148 A number of taxa containing either congeneric C_3 and C_4 species or C_3, C_3-C_4 intermediates and
149 C_4 species have been successfully hybridised (Fig. 2A&B). Although the outcome of these analysis
150 varied, whilst wholesale transfer of C_4 traits have not been reported in some instances specific traits
151 were introgressed into a C_3 background. For example, crosses between C_4 *Atriplex rosea* and C_3
152 *Atriplex prostrata* (formerly *A. patula* ssp. *hastata* and *A. triangularis* respectively), C_3 *A. rosea* and
153 C_3 *A. glabriuscula* have been made (Björkman *et al.*, 1969; Nobs *et al.*, 1970). Populations derived
154 from such crosses were progressed and C_4-like characteristics assessed (Björkmann *et al.* 1971).
155 Among 200 F_3 individuals screened for CO_2 compensation point, 178 individuals showed values
156 similar to the C_3 parent, 19 showed intermediate phenotypes and 3 were similar to the C_4 parent
157 (Björkman *et al.*, 1969). Thus, in a small number of individuals it appears that crossing was able to
158 integrate loci associated with compensation point. When F_1 derived from a C_4 *A. rosea* by C_3 *A.*
159 *patula* hybridisation were backcrossed to C_4 *A. rosea* these BC_1 offspring segregated for either C_4
160 or C_3 photosynthesis, with only two individuals showing C_4 photosynthesis (Rikiishi *et al.*, 1988),
161 suggesting dominance towards a C_3 state in this hybrid combination. In these reports above, no F_1
162 individual, nor any within segregating F_2 and F_3 populations showed a full transfer of C_4
163 photosynthesis. More recently F_2 individuals derived from a resynthesized C_4 *A. rosea* by C_3 *A.*
164 *prostrata* cross showed large variation in leaf anatomy and nearly intermediate CO_2 compensation
165 points but individuals in the F_3 generation seemed to revert to C_3-like values (Oakley *et al.*, 2014).
166 Hybrids have also been made between C_3 and C_4-like species of *Flaveria* (Apel *et al.*, 1988; Cameron
167 *et al.*, 1989) and C_3-C_4 intermediate and C_4 *Flaveria* species (Brown *et al.*, 1986, 1992). Significant
168 F_1 sterility was encountered (Brown and Bouton, 1993) but F_2 were obtained and although they
169 possessed continuous variation with regard to C_4 leaf anatomy and carbon isotope discrimination
170 characteristics, it was skewed away from the mid-parental mean towards a C_3 or C_3-C_4 phenotype.

171 This would indicate dominance deviation towards a C_3 phenotype despite the presence of genes
172 that allow C_4 photosynthesis. In F_1 hybrids derived from a C_3 by C_4-like *Flaveria* cross, enzyme
173 activities of PEPC, PPDK and NADP-ME were skewed towards those associated with C_3
174 photosynthesis, but C_4-like activities were reported for NADP-Malate Dehydrogenase (Holaday *et*
175 *al.*, 1988) indicating that incomplete dominance for certain genes may exist while others show
176 dominant activity patterns. In summary, although many C_3 by C_4 hybrids in the dicotyledons showed
177 reduced fertility and limited penetrance of C_4 traits, these studies also indicate that aspects of C_4
178 photosynthesis are heritable in a C_3 background. As many other closely related C_3 and C_4 species
179 exist (Fig. 2C) it is possible that additional stable hybrids could be generated that exhibit increased
180 genomic stability and/or better trait segregation between the C_3, C_3-C_4, and C_4 types. Hybrids
181 between different C_4 decarboxylation subtypes may also be possible. Closely related species such
182 as *Blepharis cilaris* and *Blepharis attenuata* that use NAD-ME and NADP-ME respectively have been
183 described (Akhani *et al.*, 2008). To our knowledge whilst no hybrids have been reported in *Blepharis,*
184 natural hybrids between *Cynodon dactylon* (NAD-ME) and *Chloris* sp. (PEPCK) display intermediate
185 activities of NAD-ME and PEPCK (Prendergast, 1987).

186 C_3-C_4 hybrids have been generated in the grasses by two broad approaches. First, as with
187 dicotyledons, congeners using either C_3 or C_3-C_4 photosynthesis have been crossed. Second, much
188 wider crosses of distantly related species have been performed. Examples of crosses within a genus
189 include C_3 and C_3-C_4 intermediate *Steinchisma* (formally *Panicum*) species from the Poaceae
190 (Bouton *et al.*, 1986; Brown *et al.*, 1986; Sternberg *et al.*, 1986). F_2 and F_5 individuals derived from
191 hybridisation of *Steinchisma milioides* (C_3-C_4) and *Steinchisma laxum* (C_3), or *S. spathellosum* (C_3-
192 C_4) and *S. boliviense* (C_3) exhibited intermediate leaf morphologies, CO_2 compensation points and
193 $\delta^{13}C$ values. Also within the Poaceae C_3 and C_4 accessions of *Alloteropsis semialata* have been
194 hybridised producing plants with intermediate anatomical traits as well as C_4 gene expression
195 (Bianconi *et al.*, 2021). Thus, in these hybridisations, some traits important for C_4 photosynthesis
196 could be introduced into an otherwise C_3 leaf. A variety of attempts at wide hybridisation have also
197 been reported. For example, although maize pollen germinates and fertilises the ovule of wheat to
198 form zygotes containing a full haploid set of each parental genome (Laurie & Bennett 1986), these

199 hybrids were unstable and after three rounds of mitotic cell divisions during embryogenesis all maize
200 chromosomes were lost (Laurie & Bennett 1986, 1989). In contrast, after hybridisation of oat and
201 pearl millet (*Pennisetum glaucum*) (Gernand *et al.* 2005; Ishii *et al.* 2010), some oat embryos
202 contained all pearl millet chromosomes and embryo rescue allowed hybrids possessing the haploid
203 genomes of both species to be obtained (Ishii *et al.*, 2013). It appears that the pearl millet
204 chromosomes had incorporated centromeric oat histones (Ishii *et al.*, 2015) but these haploid oat-
205 millet F_1 hybrids developed necrosis and died. This may have been caused by incompatibility
206 between the species or non-ideal tissue culture conditions. Crosses between wheat and grain pearl
207 millet (*Pennisetum americanum*) or oat and maize both allowed individual chromosomes from one
208 species to be incorporated into the other. In the case of wheat and grain pearl millet from 958
209 hybridisations one wheat plant carrying an additional pearl millet chromosome was identified (Ahmad
210 & Comeau 1990). Although this chromosome was maintained until flowering, it was not detected in
211 the next generation. Thus, wheat-pearl millet hybrids may be more stable than wheat-maize hybrids
212 but problems maintaining chromosomes from both parents still appear to exist. Unlike wheat-maize
213 hybrids, maize chromosomes have successfully been integrated into oat. This allowed the synthesis
214 of so called oat-maize chromosome addition lines that stably inherit single chromosome pairs from
215 maize (Kynast *et al.*, 2001, 2004). As with the pearl millet-oat crosses (Ishii *et al.*, 2015) stability of
216 the oat-maize addition lines appears to be mediated by incorporation of centromeric oat histones
217 into the maize chromosomes such that proper chromosomal segregation can take place during
218 mitosis (Jin *et al.*, 2004; Wang *et al.*, 2014). In some maize-oat lines, C_4 characteristics such as
219 abundant transcripts of *PEPC* or C_4-like bundle sheath cell size and vein spacing were detected
220 (Tolley *et al.*, 2012).

221 In summary, the findings based on wide hybridization of maize and oat indicate that breeding
222 offers a possible route to incorporate some C_4 traits into C_3 crops without prior knowledge of the
223 underlying genetics. Although additional parental combinations may exist that allow greater trait
224 stability in progeny this approach has not yet allowed loci controlling C_4 traits to be identified. In
225 contrast, quantitative variation in C_4 characteristics within a C_4 species would allow trait mapping
226 and there is increasing evidence that this could be informative.

227

228 **Intraspecific variation in C_4 photosynthesis**

229 As PEPC discriminates less than RuBisCO against the ^{13}C isotope, a stronger C_4 cycle leads to
230 lower incorporation of ^{13}C into tissue and so less negative $\delta^{13}C$ values (Leary, 1988). Intraspecific
231 variation in $\delta^{13}C$ has been reported in maize and *Gynandropsis gynandra* (Voznesenskaya *et al.*,
232 2007; Kolbe and Cousins, 2018; Kolbe *et al.*, 2018; Reeves *et al.*, 2018; Twohey III *et al.*, 2019). To
233 our knowledge, the extent to which this variation in C_4 efficiency is caused by differences in Kranz
234 anatomy, cell biology or C_4 biochemistry has not been determined but as summarised next, variation
235 in some of these traits within a species has been reported. This includes variation in vein density in
236 maize (Yabiku and Ueno, 2017; Kolbe and Cousins, 2018) as well as bundle sheath cell size in
237 *Alloteropsis semialata* (Lundgren *et al.*, 2016) and *G. gynandra* (Reeves *et al.*, 2018). Thus, natural
238 variation in Kranz anatomy is found within species of C_4 monocotyledons and dicotyledons.
239 Statistical modelling suggests evolution of enlarged bundle sheath cells and vein density were
240 amongst the first changes to occur during the transition from C_3 to C_4 photosynthesis (Williams *et*
241 *al.*, 2013) and phylogenetic reconstructions reveal that these changes likely happened in response
242 to reduced water availability (Edwards and Smith, 2010). As bundle sheath cell size and vein density
243 were found to be correlated with water use efficiency in maize (Yabiku and Ueno, 2017) and *G.*
244 *gynandra* (Reeves et al., 2018) it is possible that analysis of C_4 accessions adapted to different water
245 availabilities will allow additional examples of intraspecific variation in Kranz anatomy to be identified.
246 While bundle sheath cells are always greener in C_4 compared with C_3 species, the proportion of
247 leaf tissue allocated to bundle sheath compared with the mesophyll cells can be caused by either
248 increased bundle sheath cell size or vein density (Sedelnikova *et al.*, 2018). Interestingly, within *G.*
249 *gynandra* these characteristics co-vary and correlate negatively with one another (Reeves *et al.*,
250 2018). In addition to variation in Kranz anatomy in a species, there is also evidence that the cell
251 biology of C_4 leaves can differ. For example, some accessions of *Panicum coloratum* possess a
252 suberised bundle sheath whilst others do not (Ohsugi & Murata 1985). There is also variation in
253 chloroplast organisation with some accessions arranging chloroplasts centrifugally and others
254 centripetally compared with veins (Ohsugi & Murata 1985). Interestingly, *Cynodon dactylon* an NAD-

ME subtype with centripetal chloroplasts and a suberised bundle sheath, hybridises naturally with *Chloris* that uses PEPCK as the primary C_4 acid decarboxylase, has centrifugally arranged chloroplasts and no suberisation of the bundle sheath (Prendergast 1987). F_1s demonstrated intermmediacy for these traits (Prendergast, 1987). Thus, these species offer an interesting system to study regulators of bundle sheath cell biology.

To our knowledge, there are no clear examples of quantitative variation in the extent to which accessions of an individual C_4 species use the various C_4 acid decarboxylases. However, there are two reasons to consider this likely. First, in twenty-six founder lines of a maize multi-parent population, variation in the activities of C_4 enzymes has been reported (McMullen *et al.*, 2009; Kolbe *et al.*, 2018). As the founders show differences in enzyme activity it is likely that lines of the mapping population possess similar variation. Accessions of *A. semialata* (Dunning *et al.* 2017) and *G. gynandra* (Reeves et al., 2018) demonstrate differences in transcript abundance and so it appears likely that these species will also demonstrate variation in activity of C_4 acid decarboxylases. Second, the extent to which the different C_4 acid decarboxylases are engaged can vary with the environment. For example, in *G. gynandra* and maize increased abundance of transcripts encoding C_4 enzymes did not correlate with photosynthetic efficiency (Reeves *et al.* 2018; Kolbe & Cousins 2018) but in *G. gynandra* they were associated with increased water use efficiency. Additionally, the PEPCK sub-type is considered more efficient under lower levels of light since it theoretically requires fewer quanta of light per CO_2 molecule fixed (Furbank, 2011; Yin and Struik, 2020). Consistent with this, sugarcane (*Saccharum offiniarum*) and maize which predominantly use NADP-ME showed lower and higher activities of NADP-ME and PEPCK respectively after either shade or salt stress (Omoto *et al.*, 2012; Sharwood *et al.*, 2014; Sales *et al.*, 2018). Increased CO_2 leakage from bundle sheath cells has also been reported, and it has been proposed that this is caused by increased use of cytosolic PEPCK compared with the chloroplastic NADP-ME (Sales *et al.,* 2018). If populations of these species have become reproductively isolated in habitats with distinct light supplies, differences in sub-type preference may have evolved. Thus, C_4 traits ranging from discrimination against $\delta^{13}C$, C_4 leaf anatomy, bundle sheath cell biology, and C_4 transcript abundance have been documented within a species. In each case breeding and quantitative genetics offer an opportunity to identify loci

283 controlling these traits. Within this context, we next assess opportunities associated with quantitative
284 genetics to better understand C_4 photosynthesis.
285
286 **Quantitative genetics and C_4 photosynthesis**
287 Quantitative genetics allow traits exhibiting continuous variation to be linked to genomic regions
288 termed Quantitative Trait Loci (QTL). Advances in high-throughput phenotyping relevant to
289 photosynthetic performance (reviewed by Choudhury *et al.* 2019; van Bezouw *et al.* 2019) means
290 quantitative genetics now offers a path to dissect the genetics underlying photosynthesis.
291 Traditional QTL mapping requires a linkage map (or genetic map) to order loci. Using a population
292 derived from two parents that differ in a trait of interest, associations between the trait and molecular
293 markers can identify genes in close proximity to the trait (Mauricio, 2001). Advantages of QTL
294 mapping are that limited knowledge of the genome is necessary and producing bi-parental
295 populations is relatively rapid (Fig. 3A). Recombinant Inbred Lines (RILs) can be produced for
296 example from a segregating F_2 generation through rounds of self-fertilization and so generate an
297 immortalized population that can be genotyped once but phenotyped repeatedly. This is especially
298 useful for heritability estimates and mapping QTL in different environments or years (Broman, 2005).
299 Due to considerable differences in the biochemistry and physiology of C_3 and C_4 plants, if mapping
300 populations derived from C_3 and C_4 parents of *Atriplex*, *Alloteropsis* or *Flaveria* were generated, QTL
301 mapping could likely associate genes with a wide variety of C_4 phenotypes. *Alloteropsis semialata*
302 could be of particular interest here because of the presence of both C_3 and C_4 sub-species that
303 hybridise to produce offspring with intermediate characteristics (Bianconi *et al.*, 2021). As self-
304 fertilisation is also possible a population of RILs could be designed specifically for the investigation
305 of C_4 traits. High-throughput phenotyping combined with the convoluted neural network Mask R-
306 CNN (He *et al.*, 2017) has been used for QTL mapping of C_4-relevant traits in biparental populations.
307 This allowed rapid assessment of thousands of images and identification of QTL for stomatal traits
308 such as size and density (Xie *et al.*, 2021).
309 Although QTL mapping is used extensively, its power is limited if the trait is responsive to the
310 environment and so has low heritability. The heritability of many C_4 traits remains poorly understood

311 but there is growing evidence that variation in CO_2 fixation processes and leaf anatomy exist (Table
312 1) and so estimates of heritability of such C_4 traits should be possible. Given the complexity of
313 photosynthesis, its ability to respond to the environment and temporal variation in its efficiency, it is
314 highly likely that low heritability traits will be encountered (Flood *et al.*, 2016). Although traits with low
315 heritability can be investigated using highly controlled environments, highly inbred populations in
316 combination with high-density marker systems are necessary to capture the multiple small-effect
317 QTL contributing to the low-heritability trait of interest. An alternative approach involves Genome
318 Wide Association Studies (GWAS) or Linkage Disequilibrium (LD) mapping, which identifies markers
319 such as single nucleotide polymorphisms (SNP) that are in LD with the phenotype of interest (Tam
320 *et al.*, 2019). GWAS does not require a segregating population but rather uses many diverse
321 accessions that represent thousands of years of recombination to capture multiple alleles allowing
322 marker groups (haplotypes) to be identified in close association with causal loci. Additionally, it has
323 the advantage of being feasible for obligate out-crossers. In order to work successfully GWAS
324 requires many markers since it relies on LD decay (Mackay and Powell, 2007) and as pedigrees are
325 unknown, physical maps are also needed. Although population structure increases the number of
326 false positives derived from GWAS (Korte and Farlow, 2013) this is increasingly being overcome by
327 statistical modelling (Cortes *et al.* 2021). GWAS has identified QTL associated with photosynthetic
328 performance during chilling in maize (Strigens *et al.*, 2013) and sorghum (Ortiz *et al.*, 2017). More
329 recently, a sorghum diversity panel of 756 African accessions was described (Faye *et al.*, 2021) and
330 a diverse 869-line panel (Valluru *et al.*, 2019) subjected to GWAS to identify genes controlling
331 stomatal conductance and water use efficiency (Ferguson *et al.*, 2021; Pignon *et al.*, 2021). The
332 latter two studies used transcriptome data to allow transcriptome-wide association as well as GWAS
333 (reviewed by Wainberg *et al.* 2019) to increase the likelihood of identifying candidate genes.
334 Association mapping has also been used to study the light-dependent reactions of photosynthesis
335 (van Bezouw *et al.* 2019) but to our knowledge QTL determining differences in C_4 carbon fixation or
336 Kranz anatomy have not yet been identified. The sorghum and maize mapping panels present an
337 avenue through which targeted phenotyping of C_4-specific traits could be used to identify genes
338 responsible for the C_4 syndrome. For example, if a gene controlling bundle sheath cell size was

identified through mapping in maize or sorghum this could then be introduced in a C_3 crop such as rice to determine whether this allowed engineering of this trait.

Association mapping can be combined with specific breeding pedigrees to capture multiple recombination events, account for population structure, and so allow higher resolution mapping. These include Nested-Association Mapping (NAM) and Multi-parent Advanced Generation Inter-Crossing (MAGIC) population designs. Both address issues with GWAS and capture more allelic variation than bi-parental populations. Whilst allelic diversity is reduced in these multiparent designs compared with GWAS, linkage mapping as well as association mapping are possible and this is particularly useful when a physical map is not available (Broman *et al.*, 2018). Thus, NAM and MAGIC are currently particularly relevant for C_4 photosynthesis because although annotated genome sequences are being developed for *e.g.*, *Alloteropsis* sp., *Flaveria* sp., and *G. gynandra*, complete and well-annotated genomes for many C_4 model species have not yet been developed. The NAM design involves crossing one recurrent parent with many other accessions. Progeny from each cross is initially bulked and then self-fertilized for multiple generations leading to multiple RIL families (one family per unique founder) that then constitute the final NAM population (Yu *et al.*, 2008; McMullen *et al.*, 2009). At least two NAM populations exist for maize (Yu *et al.*, 2008; Chen *et al.*, 2019) and as mentioned above, significant variation for $\delta^{13}C$ as well as CA, PEPC and RuBisCO activities have been reported in the founder lines (Zhang *et al.*, 2015; Kolbe *et al.*, 2018; Twohey III *et al.*, 2019). Despite this, QTL for these traits have to our knowledge not yet been determined. A sorghum NAM population has been used in conjunction with an association panel to identify QTL for grain filling (Tao *et al.*, 2020). NAM populations offer the chance to study an extremely divergent line, such as a pre-domesticated species in the background of a stable population. This has been done with teosinte and maize as the recurrent parent (Chen *et al.*, 2019). Given the noted differences in maize and teosinte photosynthetic capacity (Yabiku and Ueno, 2017) this offers an interesting resource to map traits that differ between these species.

The MAGIC design also relies on homozygous founder lines that differ in traits of interest. Inter-crossing for multiple generations allows segregating populations to be formed consisting of lines that capture the founder genomes in unique recombinants (Fig. 3B). Such segregating lines then undergo

self-fertilization for several generations to generate RILs that capture multiple allele combinations from the various parents (Cavanagh *et al.*, 2008). With MAGIC, haplotype diversity is not limited by the use of a single recurrent parent (Ladejobi *et al.*, 2016) and although the MAGIC design requires large amounts of hybridization and significant time to produce the final population (Huang *et al.*, 2015; Pascual *et al.*, 2015; Ongom and Ejeta, 2017; Mahan *et al.*, 2018) simplified strategies can be implemented (Stadlmeier *et al.* 2018). In the context of C_4 photosynthesis, MAGIC RILs are available for maize and sorghum (Dell'Acqua *et al.*, 2015; Ongom and Ejeta, 2017; Mahan *et al.*, 2018; Butrón *et al.*, 2019). Additionally, transcriptome data exist for the founders of one maize MAGIC population (Dell'Acqua *et al.*, 2015) and ninety-four of the MAGIC RILs (Baute *et al.*, 2016). Should these RILs possess variation in activity of C_4 enzymes or components of Kranz anatomy, QTL could be identified. To our knowledge, there is currently no MAGIC population available for a C_4 dicotyledon, nor a mapping panel designed explicitly to map C_4 photosynthetic traits. As variation in C_4 traits has been reported in *A. semialata* and *G. gynandra* (Lundgren *et al.*, 2016; Reeves *et al.*, 2018) and they can be crossed (Sogbohossou *et al.*, 2018; Bianconi *et al.*, 2020) mapping resources in these species would be useful.

Once a QTL is identified using any of the above population types, fine mapping enables causative genes to be identified (Hormozdiari *et al.*, 2014; Tam *et al.*, 2019). Parsing C_4 photosynthesis into individual components such genes controlling C_4 enzyme activity or bundle sheath cell size (Dunning *et al.*, 2017) are identified by different phenotyping techniques, combined with fine mapping could identify additional genes required for C_4 photosynthesis. Exploiting the high degree of natural variation among C_3 and C_4 species will enable genome-wide associations to help map critical photosynthesis regulators. Furthermore, inferences into the inheritance of C_4 components such as cell specific gene expression can be parsed even without proper segregation or recombination in C_3 and C_4 hybrids (Fig. 4). While such methods cannot identify QTL, they can at least establish broad modes of inheritance (Charlesworth and Willis, 2009). For example, sterile F_1 populations derived from C_3 and C_4 parents that show altered transcript abundance or cellular localization of C_4 enzymes, can provide insight into whether genes are controlled in *cis*, *trans*, or a combination of both mechanisms, and whether these mechanisms are functioning in an activating or repressive manner

(Fig. 4). This technique has been deployed in F_1 hybrids derived from a cross between the C_3-C_4 intermediate *Moricandia arvensis* and the C_3 *M. moricandiodes* to show *cis*-regulation dominates control of photosynthetic and anatomical phenotypes (Lin *et al.*, 2021). Information from such studies could inform mapping strategies and marker placement for associations.

In summary, in order to modify C_3 leaves to perform C_4 photosynthesis, an improved understanding of C_4 anatomy, cell biology and biochemistry is needed. Wide hybridization either by sexual or asexual means to recombine interspecific variation found in C_3 and C_4 species or intraspecific photosynthetic variation in C_4 species, combined with mapping populations and high-throughput phenotyping should facilitate a better understanding of C_4 photosynthesis. Quantitative genetics then offer robust methods to better understand the regulatory mechanisms behind these traits. Applying these techniques therefore promise to enhance photosynthetic efficiency of C_3 and C_4 crops and so contribute to a more robust world agriculture in the future.

Acknowledgements

S.C.J.S. was supported by a BBSRC PhD studentship; G.R by a Gates Cambridge Trust PhD Student Fellowship; G.R. P.S. and A.T. by ERC Grant 694733 Revolution and BB/P003117/1 awarded to J.M.H.; and G.R. and A.T. by a Ceres Agri-Tech Fund award. All authors contributed to the analysis of literature and writing of this review.

413 **References**

414 **Akhani H, Ghasemkhani M, Chuong SDX, Edwards GE**. 2008. Occurrence and forms of Kranz
415 anatomy in photosynthetic organs and characterization of NAD-ME subtype C_4 photosynthesis in
416 *Blepharis ciliaris* (L.) B. L. Burtt (Acanthaceae). Journal of Experimental Botany **59**, 1755–1765.
417 **Apel P, Bauwe H, Bassüner B, Maass I**. 1988. Photosynthetic properties of *Flaveria cronquistii*,
418 *F. palmeri*, and hybrids between them. Biochemie und Physiologie der Pflanzen **183**, 291–299.
419 **Baute J, Herman D, Coppens F, Block J De, Slabbinck B, Acqua MD, Pè ME, Maere S,**
420 **Nelissen H, Inzé D**. 2016. Combined large-scale phenotyping and transcriptomics in maize
421 reveals a robust growth regulatory network. Plant Physiology **170**, 1848–1867.
422 **van Bezouw RFHM, Keurentjes JJB, Harbinson J, Aarts MGM**. 2019. Converging phenomics
423 and genomics to study natural variation in plant photosynthetic efficiency. Plant Journal **97**, 112–
424 133.
425 **Bianconi ME, Dunning LT, Curran E V, *et al.*** 2020. Contrasted histories of organelle and nuclear
426 genomes underlying physiological diversification in a grass species. Proceedings of the Royal
427 Society B **287**.
428 **Bianconi ME, Sotelo G, Curran E V, Milenkovic V, Samaritani E, Dunning LT, Osborne CP,**
429 **Christin P**. 2021. Upregulation of C_4 characteristics does not consistently improve photosynthetic
430 performance in intraspecific hybrids of a grass. bioRxiv doi: 10.1101/2021.08.10.455822 [Preprint]
431 **Björkman O, Gauhl E, Nobs M**. 1969. Comparative studies of *Atriplex* species with and without ß-
432 carboxylation photosynthesis. Carnegie Institution of Washington Yearbook **68**, 620–633.
433 **Björkmann O, Nobs MA, Berry J**. 1971. Further studies on hybrids beteen C_3 and C_4 species of
434 *Atriplex*. Carnegie Instituite of Washington Annual Report **70**, 507–511.
435 **Botha CEJ**. 1992. Plasmodesmatal distribution, structure and frequency in relation to assimilation
436 in C_3 and C_4 grasses in southern Africa. Planta **187**, 348–358.
437 **Bouton JH, Brown RH, Evans PT, Jernstedt JA**. 1986. Photosynthesis, leaf anatomy, and
438 morphology of progeny from hybrids between C_3 and C_3/C_4 *Panicum* species. Plant Physiology **80**,
439 487–492.
440 **Bowes G, Ogren WL, Hageman RH**. 1971. Phosphoglycolate production catalzed by Ribulose

441 Diphosphate Carboxylase. Biochemical and Biophysical Research Communications **45**, 716–722.
442 **Broman KW**. 2005. The genomes of recombinant inbred lines. Genetics.
443 **Broman KW, Gatti DM, Simecek P, Furlotte NA, Prins P, Sen Ś, Yandell BS, Churchill GA**.
444 2018. R/qtl2: Software for mapping quantitative trait loci with high-dimensional data and for
445 mapping quantitative trait loci with high-dimensional data and multiparent populations high-
446 dimensional data and multiparent populations. Genetics **211**, 495–502.
447 **Brown RH, Bassett CL, Cameron RG, Evans PT, Bouton JH, Black CC, Sternberg LO, Deniro**
448 **MJ**. 1986. Photosynthesis of F_1 hybrids between C_4 and C_3-C_4 species of *Flaveria*. Plant
449 Physiology **82**, 211–217.
450 **Brown HR, Bouton JH**. 1993. Interspecific hybrids between photosynthetic types. Annual Review
451 of Plant Physiology **44**, 435–436.
452 **Brown RH, Byrd GT, Black CC**. 1992. Degree of C_4 photosynthesis in C_4 and C_3-C_4 Flaveria
453 species and their hybrids: II. inhibition of apparent photosynthesis by a phosphoenolpyruvate
454 carboxylase inhibitor. Plant Physiology **100**, 947–950.
455 **Butrón A, Santiago R, Cao A, Samayoa LF, Malvar RA**. 2019. QTLs for resistance to *Fusarium*
456 ear rot in a Multiparent Advanced Generation Intercross (MAGIC) maize population. Plant Disease
457 **109**, 897–904.
458 **Calvin M, Benson AA**. 1948. The path of carbon in photosynthesis. **107**, 476–480.
459 **Cameron RG, Bassett CL, Bouton JH, Brown RH**. 1989. Transfer of C_4 photosynthetic
460 characters through hybridization of *Flaveria* species. Plant Physiology **90**, 1538–1545.
461 **Carlson PS, Smith HH, Dearing R**. 1972. Parasexual interspecific plant hybridization.
462 Proceedings of the National Academy of Sciences **69**, 2292–2294.
463 **Cavanagh C, Morell M, Mackay I, Powell W**. 2008. From mutations to MAGIC: resources for
464 gene discovery, validation and delivery in crop plants. Current Opinion in Plant Biology **11**, 215–
465 221.
466 **Charlesworth D, Willis JH**. 2009. The genetics of inbreeding depression. Nature Reviews
467 Genetics **10**, 783–796.
468 **Chen Q, Yang CJ, York AM, *et al.*** 2019. TeoNAM: A nested association mapping population for

domestication and agronomic trait analysis in maize. Genetics **213**, 1065–1078.

Choudhury S Das, Samal A, Awada T. 2019. Leveraging image analysis for high-throughput plant phenotyping. **10**, 1–8.

Cortes LT, Zhang Z, Yu J. 2021. Status and prospects of genome-wide association studies in plants. **14**, 1–17.

Danila FR, Quick WP, White RG, Furbank RT. 2016. The metabolite pathway between Bundle Sheath and Mesophyll : Quantification of Plasmodesmata in Leaves of C_3 and C_4 Monocots. The Plant Cell **28**, 1461–1471.

Dell'Acqua M, Gatti DM, Pea G, *et al.* 2015. Genetic properties of the MAGIC maize population: A new platform for high definition QTL mapping in *Zea mays*. Genome Biology **16**, 1–23.

Dunning LT, Lundgren MR, Moreno-Villena JJ, Namaganda M, Edwards EJ, Nosil P, Osborne CP, Christin PA. 2017. Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives. Evolution **71**, 1541–1555.

Edwards EJ, Smith SA. 2010. Phylogenetic analyses reveal the shady history of C_4 grasses. Proceedings of the National Academy of Sciences of the United States of America **107**, 2532–2537.

Edwards GE, Voznesenskaya E V. 2011. C_4 phtosynthesis: Kranz forms and single-cell C_4 in terrestrial plants. In: Raghavendra AS, Sage RF, eds. C_4 photosynthesis and related CO_2 concentrating mechanisms. Dordrecht: Springer, 29–61.

Evans DA. 1983. Agricultural applications of plant protoplast fusion. Nature Biotechnology **1**, 253–261.

Faye JM, Maina F, Akata EA, *et al.* 2021. A genomics resource for genetics, physiology, and breeding of West African sorghum. Plant Genome **14**, 1–18.

Ferguson JN, Fernandes SB, Monier B, *et al.* 2021. Machine Leaning Enabled Phenotyping for GWAS and TWAS of WUE Traits in 869 Field-Grown Sorghum Accessions. Plant Physiology kiab346, https://doi.org/10.1093/plphys/kiab346

Flood PJ, Kruijer W, Schnabel SK, Schoor R, Jalink H, Snel JFH, Harbinson J, Aarts MGM. 2016. Phenomics for photosynthesis, growth and reflectance in Arabidopsis thaliana reveals

497 circadian and long-term fluctuations in heritability. Plant Methods **12**, 1–14.
498 **Furbank RT**. 2011. Evolution of the C_4 photosynthetic mechanism: Are there really three C_4 acid
499 decarboxylation types? Journal of Experimental Botany **62**, 3103–3108.
500 **Gernand D, Rutten T, Varshney A, Rubtsova M, Prodanovic S, Brüß C, Kumlehn J, Matzk F,**
501 **Houben A**. 2005. Uniparental chromosome elimination at mitosis and interphase in wheat and
502 pearl millet crosses involves micronucleus formation, progressive heterochromatinization, and DNA
503 fragmentation. Plant Cell **17**, 2431–2438.
504 **Hatch M, Kagawa T, Craig S**. 1975. Subdivision of C_4-pathway species based on differing C_4 acid
505 decarboxylating systems and ultrastructural features. Functional Plant Biology **2**, 111.
506 **He K, Gkioxari G, Dollar P, Girshick R**. 2017. Mask R-CNN. IEEE International Conference on
507 Computer Vision, 2980-2988.
508 **Holaday AS, Brown RH, Bartlett JM, Sandlin EA, Jackson RC**. 1988. Enzymic and
509 photosynthetic characteristics of reciprocal F_1 hybrids of *Flaveria pringlei* (C_3) and *Flaveria brownii*
510 (C_4-like species). Plant Physiology **87**, 484–490.
511 **Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E**. 2014. Identifying causal variants at
512 loci with multiple signals of association. Genetics **198**, 497–508.
513 **Huang BE, Verbyla KL, Verbyla AP, Raghavan C, Singh VK, Gaur P, Leung H, Varshney RK,**
514 **Cavanagh CR**. 2015. MAGIC populations in crops: current status and future prospects. Theoretical
515 and Applied Genetics **128**, 999–1017.
516 **Ishii T, Sunamura N, Matsumoto A, Eltayeb AE, Tsujimoto H**. 2015. Preferential recruitment of
517 the maternal centromere-specific histone H3 (CENH3) in oat (*Avena sativa* L.) × pearl millet
518 (*Pennisetum glaucum* L.) hybrid embryos. Chromosome Research **23**, 709–718.
519 **Ishii T, Tanaka H, Eltayeb AE, Tsujimoto H**. 2013. Wide hybridization between oat and pearl
520 millet belonging to different subfamilies of Poaceae. Plant Reproduction **26**, 25–32.
521 **Ishii T, Ueda T, Tanaka H, Tsujimoto H**. 2010. Chromosome elimination by wide hybridization
522 between Triticeae or oat plant and pearl millet: Pearl millet chromosome dynamics in hybrid
523 embryo cells. Chromosome Research **18**, 821–831.
524 **Jin W, Melo JR, Nagaki K, Talbert PB, Henikoff S, Dawe RK, Jiang J**. 2004. Maize

525 centromeres: Organization and functional adaptation in the genetic background of oat. Plant Cell
526 **16**, 571–581.
527 **Kadereit G, Borsch T, Weising K, Freitag H**. 2003. Phylogeny of Amaranthaceae and
528 Chenopodiaceae and the evolution of C_4 photosynthesis. International Journal of Plant Sciences
529 **164**, 959–986.
530 **Kataria S, Guruprasad KN**. 2012. Intraspecific variations in growth, yield and photosynthesis of
531 sorghum varieties to ambient UV (280-400nm) radiation. Plant Science **196**, 85–92.
532 **Kolbe AR, Cousins AB**. 2018. Mesophyll conductance in *Zea mays* responds transiently to CO_2
533 availability: implications for transpiration efficiency in C_4 crops. New Phytologist **217**, 1463–1474.
534 **Kolbe AR, Studer AJ, Cousins AB**. 2018. Biochemical and transcriptomic analysis of maize
535 diversity to elucidate drivers of leaf carbon isotope composition. Functional Plant Biology **45**, 489–
536 500.
537 **Korte A, Farlow A**. 2013. The advantages and limitations of trait analysis with GWAS: A review.
538 Plant Methods **9**, 1.
539 **Koteyeva NK, Voznesenskaya E V., Roalson EH, Edwards GE**. 2011. Diversity in forms of C_4 in
540 the genus *Cleome* (Cleomaceae). Annals of Botany **107**, 269–283.
541 **Kynast RG, Okagaki RJ, Galatowitsch MW, Granath SR, Jacobs MS, Stec AO, Rines HW,**
542 **Phillips RL**. 2004. Dissecting the maize genome by using chromosome addition and radiation
543 hybrid lines. Proceedings of the National Academy of Sciences **101**, 9921–9926.
544 **Kynast RG, Riera-Lizarazu O, Vales MI, *et al.*** 2001. A complete set of maize individual
545 chromosome additions to the oat genome. Plant Physiology **125**, 1216–1227.
546 **Ladejobi O, Elderfield J, Gardner KA, Gaynor RC, Hickey J, Hibberd JM, Mackay IJ, Bentley**
547 **AR**. 2016. Maximizing the potential of multi-parental crop populations. Applied and Translational
548 Genomics **11**, 9–17.
549 **Laurie DA, Bennett MD**. 1986. Wheat x maize hybridization. Canadian Journal of Genetics and
550 Cytology **28**, 313–316.
551 **Laurie DA, Bennett MD**. 1989. The timing of chromosome elimination in hexaploid wheat x maize
552 crosses. Genome **32**, 953–961.

Laurie DA, O'Donoughue LS, Bennett MD. 1990. Wheat x maize and other wide sexual hybrids: Their potential for genetic manipulation and crop improvement. In: Gustafson JP, ed. Gene Manipulation in Plant Improvement II.95–126.

Leary MHO. 1988. Carbon Isotopes in Photosynthesis. BioScience **38**, 328–336.

Lightfoot E, Przelomska N, Craven M, O Connell TC, He L, Hunt H V., Jones MK. 2016. Intraspecific carbon and nitrogen isotopic variability in foxtail millet (*Setaria italica*). Rapid communications in mass spectrometry : RCM **30**, 1475–1487.

Lin M, Schlüter U, Stich B, Weber APM. 2021. Cis -regulatory divergence underpins the evolution of C_3-C_4 intermediate photosynthesis in Moricandia. bioRxiv doi: 10.1101/2021.05.10.443365 [Preprint]

Lundgren MR, Christin PA, Escobar EG, Ripley BS, Besnard G, Long CM, Hattersley PW, Ellis RP, Leegood RC, Osborne CP. 2016. Evolutionary implications of C_3–C_4 intermediates in the grass Alloteropsis semialata. Plant Cell and Environment **39**, 1874–1885.

Mackay I, Powell W. 2007. Methods for linkage disequilibrium mapping in crops. Trends in Plant Science **12**, 57–63.

Mahan AL, Murray SC, Klein PE. 2018. Four-Parent Maize (FPM) population: Development and phenotypic characterization. Crop Science **58**, 1106–1117.

Mauricio R. 2001. Mapping quantitative trait loci in plants: Uses and caveats for evolutionary biology. Nature Reviews Genetics **2**, 370–381.

McMullen MD, Kresovich S, Villeda HS, *et al.* 2009. Genetic properties of the maize nested association mapping population. Science **325**, 737–740.

Mertz RA, Brutnell TP. 2014. Bundle sheath suberization in grass leaves: Multiple barriers to characterization. Journal of Experimental Botany **65**, 3371–3380.

Muhaidat R, Sage RF, Dengler NG. 2007. Diversity of Kranz Anatomy and Biochemistry in C_4 Eudicots. American Journal of Botany **94**, 362–381.

Nobs MA, Björkmann O, Pearcy RW, Boynton JE. 1970. Hybrids between *Atriplex* species with and without beta-carboxylation photosynthesis. Year Book of the Carnegie Institution of Washington **69**, 617–662.

Oakley JC, Sultmanis S, Stinson CR, Sage TL, Sage RF. 2014. Comparative studies of C_3 and C_4 *Atriplex* hybrids in the genomics era: Physiological assessments. Journal of Experimental Botany **65**, 3637–3647.

Ohsugi R, Murata T. 1985. C_4 photosynthetic characteristics of *Panicum* species in the Dichotomiflora group. Japan Agricultural Research Quarterly **19**, 125–131.

Omoto E, Taniguchi M, Miyake H. 2012. Adaptation responses in C_4 photosynthesis of maize under salinity. Journal of Plant Physiology **169**, 469–477.

Ongom PO, Ejeta G. 2017. Mating Design and Genetic Structure of a Multi-Parent Advanced Generation Intercross (MAGIC) Population of Sorghum (Sorghum bicolor L.) Moench). G3: Genes, Genomes, Genetics **8**, 331–341.

Ortiz D, Hu J, Salas Fernandez MG. 2017. Genetic architecture of photosynthesis in *Sorghum bicolor* under non-stress and cold stress conditions. Journal of Experimental Botany **68**, 4545–4557.

Pascual L, Desplat N, Huang BE, Desgroux A, Bruguier L, Bouchet JP, Le QH, Chauchard B, Verschave P, Causse M. 2015. Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. Plant Biotechnology Journal **13**, 565–577.

Peter G, Katinas L. 2003. A new type of Kranz anatomy in Asteraceae. Australian Journal of Botany **51**, 217–226.

Pignon CP, Fernandes SB, Valluru R, Bandillo N, Lozano R, Buckler E, Gore MA, Long SP, Brown PJ, Leakey ADB. 2021. Phenotyping stomatal closure by thermal imaging for GWAS and TWAS of water use efficiency-related genes. Plant Physiology kiab395 https://doi.org/10.1093/plphys/kiab395

Portis AR, Parry MAJ. 2007. Discoveries in Rubisco (Ribulose 1,5-bisphosphate carboxylase/oxygenase): A historical perspective. Photosynthesis Research **94**, 121–143.

Prendergast HD V. 1987. Structural, biochemical and geographical relationships in australian C_4 grasses.

Reeves G, Singh P, Rossberg TA, Sogbohossou EOD, Schranz ME, Hibberd JM. 2018.

609 Natural variation within a species for traits underpinning C_4 photosynthesis. Plant Physiology **177**,
610 504–512.
611 **Rikiishi K, Oguro H, Samejima M, Sugiyama T, Hinata K**. 1988. C_4-like plants derived from a
612 cross (Atriplex rosea (C_4) x A. patula (C_3)) x A. rosea. Japanese Journal of Breeding **38**, 397–408.
613 **Sage RF**. 2016. A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery:
614 Species number, evolutionary lineages, and hall of fame. Journal of Experimental Botany **67**,
615 4039–4056.
616 **Sage RF, Christin P-A, Edwards EJ**. 2011. The C_4 plant lineages of planet Earth. Journal of
617 Experimental Botany **62**, 3155–3169.
618 **Sage RF, Stata M**. 2015. Photosynthetic diversity meets biodiversity: The C_4 plant example.
619 Journal of Plant Physiology **172**, 104–119.
620 **Sales CRG, Ribeiro R V., Hayashi AH, Marchiori PER, Silva KI, Martins MO, Silveira JAG,**
621 **Silveira NM, Machado EC**. 2018. Flexibility of C_4 decarboxylation and photosynthetic plasticity in
622 sugarcane plants under shading. Environmental and Experimental Botany **149**, 34–42.
623 **Sedelnikova O V., Hughes TE, Langdale JA**. 2018. Understanding the genetic basis of C_4 Kranz
624 anatomy with a view to engineering C_3 crops. Annual Review of Genetics **52**, 249–270.
625 **Sharkey TD**. 1988. Estimating the rate of photorespiration in leaves. Physiologia Plantarum **73**,
626 147–152.
627 **Sharwood RE, Sonawane B V., Ghannoum O**. 2014. Photosynthetic flexibility in maize exposed
628 to salinity and shade. Journal of Experimental Botany **65**, 3715–3724.
629 **Sogbohossou EOD, Achigan-Dako EG, Maundu P, Solberg S, Deguenon EMS, Mumm RH,**
630 **Hale I, Van Deynze A, Schranz ME**. 2018. A roadmap for breeding orphan leafy vegetable
631 species: A case study of *Gynandropsis gynandra* (Cleomaceae). Horticulture Research **5**, 1–15.
632 **Stadlmeier M, Hartl L, Mohler V**. 2018. Usefulness of a multiparent advanced generation
633 intercross population with a greatly reduced mating design for genetic studies in winter wheat.
634 Frontiers in Plant Science **871**, 1–12.
635 **Sternberg LDSL, Deniro MJ, Sloan ME, Black CC**. 1986. Compensation point and isotopic
636 characteristics of C_3/C_4 intermediates and hybrids in Panicum. Plant Physiology **80**, 242–245.

Strigens A, Freitag NM, Gilbert X, Grieder C, Riedelsheimer C, Schrag TA, Messmer R, Melchinger AE. 2013. Association mapping for chilling tolerance in elite flint and dent maize inbred lines evaluated in growth chamber and field experiments. Plant, Cell and Environment **36**, 1871–1887.

Szarka B, Göntér I, Molnár-Láng M, Mórocz S, Dudits D. 2002. Mixing of maize and wheat genomic DNA by somatic hybridization in regenerated sterile maize plants. Theoretical and Applied Genetics **105**, 1–7.

Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. 2019. Benefits and limitations of genome-wide association studies. Nature Reviews Genetics **20**, 467–484.

Tao Y, Zhao X, Wang X, Hathorn A, Hunt C, Cruickshank AW, Erik J, Godwin ID, Mace ES, Jordan DR. 2020. Large-scale GWAS in sorghum reveals common genetic control of grain size among cereals. Plant Biotechnology Journal **18**, 1093–1105.

Terada R, Kyozuka J, Nishibayashi S, Shimamoto K. 1987. Plantlet regeneration from somatic hybrids of rice (*Oryza sativa* L.) and barnyard grass (*Echinochloa oryzicola* Vasing). Molecular & General Genetics **210**, 39–43.

Tolley BJ, Sage TL, Langdale JA, Hibberd JM. 2012. Individual maize chromosomes in the C_3 plant oat can increase bundle sheath cell size and vein density. Plant Physiology **159**, 1418–1427.

Twohey III RJ, Roberts LM, Studer AJ. 2019. Leaf stable carbon isotope composition reflects transpiration efficiency in *Zea mays*. The Plant Journal **97**, 475–484.

Valluru R, Gazave EE, Fernandes SB, Ferguson JN, Lozano R, Hirannaiah P, Zuo T, Brown PJ, Leakey ADB, Gore MA. 2019. Deleterious Mutation Burden and Its Association with Complex Traits in Sorghum (*Sorghum bicolor*). **211**, 1075–1087.

Voznesenskaya E V., Koteyeva NK, Chuong SDX, Ivanova AN, Barroca J, Craven LA, Edwards GE. 2007. Physiological, anatomical and biochemical characterisation of photosynthetic types in genus *Cleome* (Cleomaceae). Functional Plant Biology **34**, 247–267.

Voznesenskaya E V., Koteyeva NK, Edwards GE, Ocampo G. 2017. Unique photosynthetic phenotypes in *Portulaca* (Portulacaceae): C_3-C_4 intermediates and NAD-ME C_4 species with Pilosoid-type Kranz anatomy. Journal of Experimental Botany **68**, 225–239.

Wainberg M, Sinnott-Armstrong N, Mancuso N, *et al.* 2019. Opportunities and challenges for transcriptome-wide association studies. Nature Genetics **51**, 592–599.

Wang TB, Niizeki M. 1994. Somatic hybridization between Zea mays and Triticum sect. tritirigia. In: Bajaj YPS, ed. Biotechnology in Agriculture and Forestry. Berlin, Heidelberg: Springer Berlin Heidelberg, 99–111

Wang TB, Niizeki M, Harada T, Ishikawa R, Qian YQ, Saito K. 1993. Establishment of somatic hybrid cell lines between *Zea mays* L. (maize) and *Triticum* sect, *trititrigia* MacKey (trititrigia). Theoretical and Applied Genetics **86**, 371–376.

Wang K, Wu Y, Zhang W, Dawe RK, Jiang J. 2014. Maize centromeres expand and adopt a uniform size in the genetic background of oat. Genome Research **24**, 107–116.

Williams BP, Johnston IG, Covshoff S, Hibberd JM. 2013. Phenotypic landscape inference reveals multiple evolutionary paths to C_4 photosynthesis. eLife **2**, 1–19.

Xie J, Fernandes SB, Mayfield-Jones D, Erice G, Choi M, E Lipka A, Leakey ADB. 2021. Optical topometry and machine learning to rapidly phenotype stomatal patterning traits for maize QTL mapping. Plant Physiology kiab299, https://doi.org/10.1093/plphys/kiab299

Xin HW, Sun JS, Yan QS, Zhang XQ. 1997. Plant regeneration from asymmetric somatic hybrids of *Oryza sativa* and *Panicum maximum.* Acta Botanica Sinica **39**, 717–724.

Xu C, Xia G, Zhi D, Xiang F, Chen H. 2003. Integration of maize nuclear and mitochondrial DNA into the wheat genome through somatic hybridization. Plant Science **165**, 1001–1008.

Yabiku T, Ueno O. 2017. Variations in physiological, biochemical, and structural traits of photosynthesis and resource use efficiency in maize and teosintes (NADP-ME- type C4). Plant Production Science **20**, 448–558.

Yin X, Struik PC. 2020. Viewpoints: Exploiting differences in the energy budget among C_4 subtypes to improve crop productivity. New Phytologist **229**, 2400–2409.

Yu J, Holland JB, McMullen MD, Buckler ES. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics **178**, 539–551.

Zhang N, Gibon Y, Wallace JG, *et al.* 2015. Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. Plant Physiology **168**, 575–583.

Figures legends

Fig. 1. Natural variation in C_4 biochemistry and anatomy. (A) An overview of C_4 biochemical subtypes. Although all forms of two-celled C_4 photosynthesis involve initial CO_2 fixation to generate four-carbon intermediates in mesophyll cells and diffusion to bundle sheath cells, the method of decarboxylation to create a high-CO_2 environment around RuBisCO varies between C_4 species. Solid and dashed lines show enzymatic and diffusion steps of the C_4 pathway respectively. (B) Examples of leaf anatomies seen in C_4 species. Exemplar species that use each anatomical variant are shown below each type. Many more anatomical types have been described, which suggests multiple leaf morphologies can facilitate the C_4 pathway. Abbreviations: *M, Mesophyll; B, Bundle sheath; VB, Vascular Bundle; CCC, Central Cytoplasmic Compartment; PC, Peripheral Chloroplast; WS, Water Storage cell; ch, chloroplast.*

Fig. 2. Examples of successful as well as potential hybridizations between C_3 and C_4 species. (A) Phylogenetic reconstruction of the orders constituting flowering plants according to APG IV**,** 2016. Orders containing C_4 lineages are shown in bold. (B) Examplar hybridization webs that have resulted in successful F_1 hybrids between C_3, C_4 and C_3-C_4 intermediate photosynthetic types. (C) Taxa that contain closely related C_3, C_4, or C_3-C_4 intermediate species or accessions for which hybridisation has not been reported, but may be possible. These groups are potential systems where C_4 genes could be mapped. Arrows from the phylogenetic tree indicate from which order the plant species originate (B, C).

Fig. 3. Quantitative genetics in the context of C_4 photosynthesis. (A) A schematic for QTL mapping of leaf anatomical traits. Two homozygous parents, genotyped for four markers, A, B, C and D and differing in vein density are hybridised and advanced to form a bi-parental population that can be used to identify QTL associated with vein density (here located near markers C and D). Numbers show recombination fractions, which are used to position the QTL relative to flanking markers. (B) Population structure of a MAGIC pedigree followed by four generations of inter-crossing and self-

722 fertilization. Progeny contain more genetic variation than that derived from a bi-parental design.
723 Hypothetical plot showing how QTL associated with individually mapped C_4 phenotypes such as
724 gene expression, bundle sheath cell size or gas exchange parameters (*e.g.*, stomatal conductance,
725 CO_2 assimilation, etc) can be mapped with one population.
726
727 **Fig. 4.** Using breeding to understand the molecular basis of C_4 gene regulation. Parental populations
728 that differ in transcript abundance can be due to multiple genetic effects that can be parsed by
729 quantitative genetics. A simplified two loci model where one locus is a *cis*-element and the other an
730 activating *trans*-factor is presented to illustrate how the molecular basis underpinning variations in
731 gene expression can be determined by inheritance of gene expression in F_1 hybrids. If expression
732 of a gene is controlled by changes in *cis*-regulation between parents, offspring exhibit additive
733 expression patterns. If variation in expression is due to changes in *trans* between parents then
734 offspring exhibit dominance deviation towards one parent. Lastly, if differences in gene expression
735 between parents is due to both *cis* and *trans*- offspring demonstrate heterosis or overdominance.

1

Species	Varying Trait	Reference
Alloteropsis semialata (C_4 accessions)	• Abundance of *PEPC* and *PEPCK* transcripts	Dunning *et al.* (2017)
	• PEPC content • Carbon isotope discrimination • Mesophyll cell size • Bundle sheath cell size • Leaf physiology	Lundgren *et al.* (2016)
Gynandropsis gynandra	• C_4 transcript abundance, physiology and leaf morphology	Reeves *et al.* (2018)
Panicum coloratum	• Chloroplast location • Bundle sheath suberization	Ohsugi & Murata (1985)
Setaria italica	• Carbon isotope • "Differing intensities of green"	Lightfoot *et al.* (2016)
Sorghum bicolor	• Net assimilation rate	Kataria & Guruprasad (2012)
Zea mays	• *CA* transcript abundance	Zhang *et al.* (2015)

Zea mays	• CA, PEPC, and RuBisCO activity • Net assimilation rate • Interveinal distance • Mesophyll thickness • Max assimilation rate	Kolbe & Cousins (2018)
	• CA, PEPC and RuBisCO activity • C_4 transcript abundance • Carbon isotope	Kolbe *et al.* (2018)
	• Vein density • Gas exchange traits • PEPC, NADP-ME, PEPCK and RuBisCO activity	(Yabiku and Ueno, 2017)

Table 1. Summary of publications documenting intraspecific variation in traits relevant to C_4 photosynthesis associated traits. Abbreviations are as follows: CA = Carbonic Anhydrase, NADP-ME; NADP-dependent Malic Enzyme; PEPC = Phospho*enol*pyruvate Carboxylase; PEPCK = Phospho*enol*pyruvate Carboxykinase.

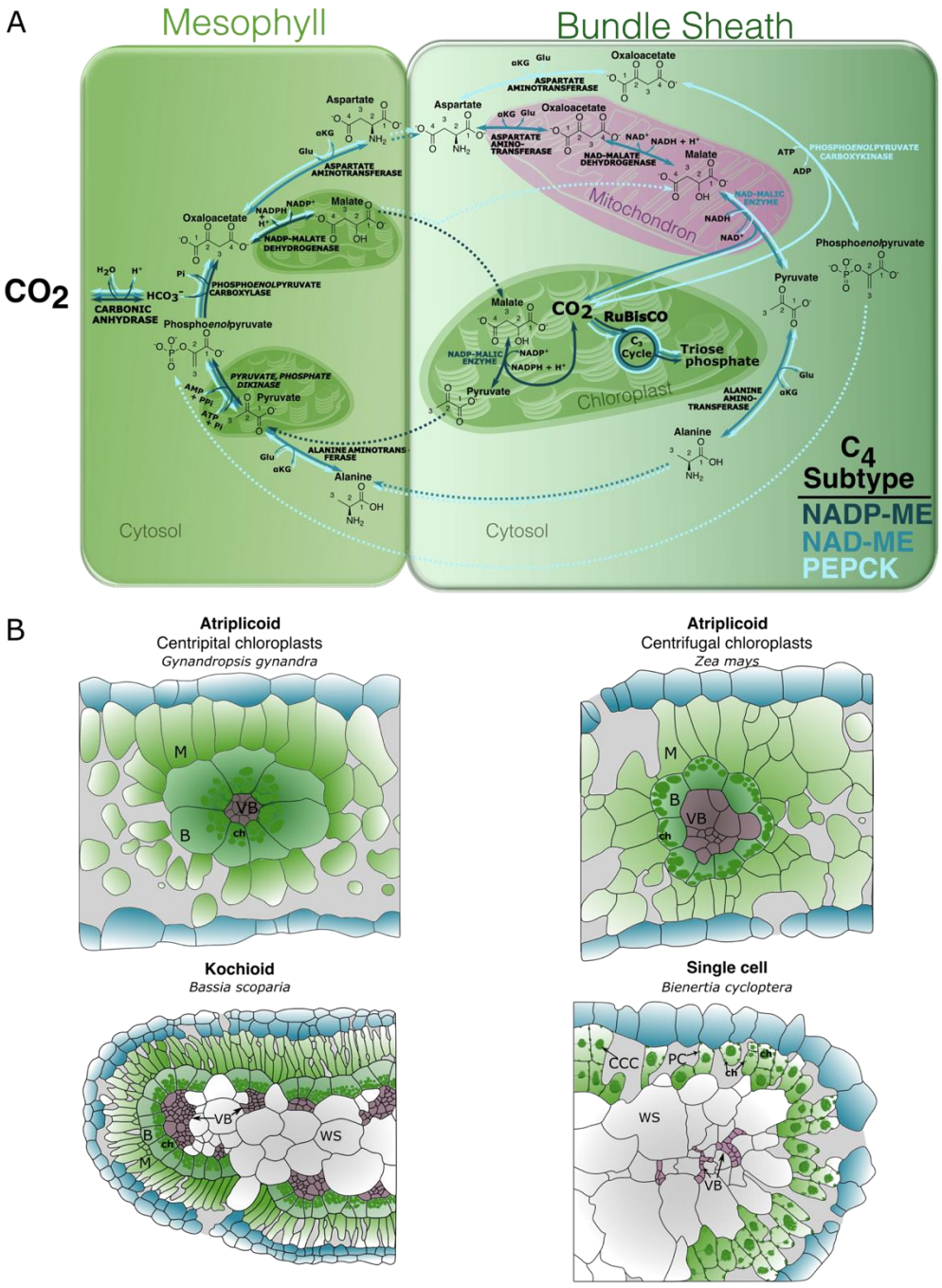

Fig. 1. Natural variation in C_4 biochemistry and anatomy. (A) An overview of C_4 biochemical subtypes. Although all forms of two-celled C_4 photosynthesis involve initial CO_2 fixation to generate four-carbon intermediates in mesophyll cells and diffusion to bundle sheath cells, the method of decarboxylation to create a high-CO_2 environment around RuBisCO varies between C_4 species. Solid and dashed lines show enzymatic and diffusion steps of the C_4 pathway respectively. (B) Examples of leaf anatomies seen in C_4 species. Exemplar species that use each anatomical variant are shown below each type. Many more anatomical types have been described, which suggests

multiple leaf morphologies can facilitate the C_4 pathway. Abbreviations: *M, Mesophyll; B, Bundle sheath; VB, Vascular Bundle; CCC, Central Cytoplasmic Compartment; PC, Peripheral Chloroplast; WS, Water Storage cell; ch, chloroplast.*

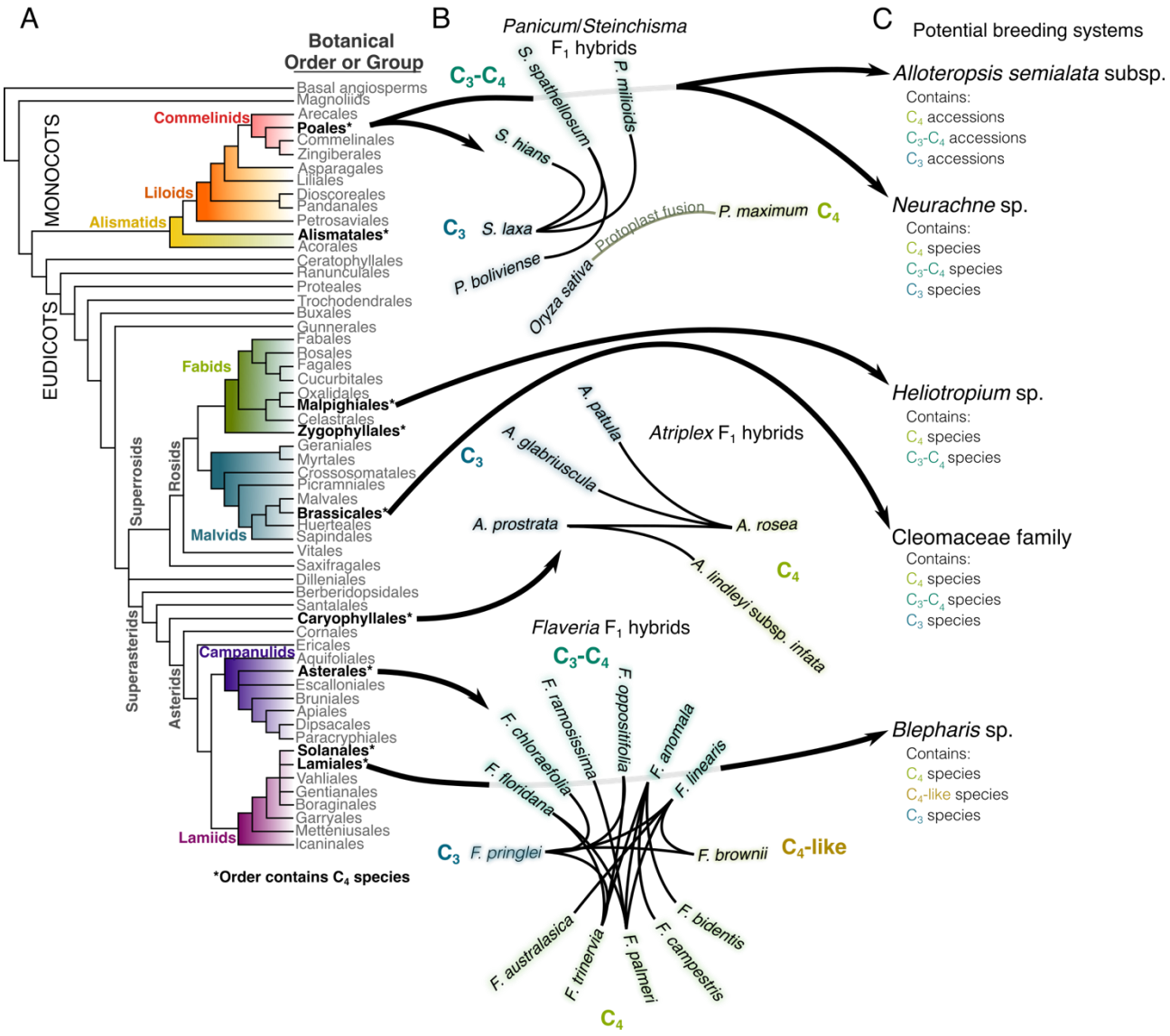

Fig. 2. Examples of successful as well as potential hybridizations between C_3 and C_4 species. (A) Phylogenetic reconstruction of the orders constituting flowering plants according to APG IV**,** 2016. Orders containing C_4 lineages are shown in bold. (B) Exemplar hybridization webs that have resulted in successful F_1 hybrids between C_3, C_4 and C_3-C_4 intermediate photosynthetic types. (C) Taxa that contain closely related C_3, C_4, or C_3-C_4 intermediate species or accessions for which hybridisation has not been reported but may be possible. These groups are potential systems where C_4 genes could be mapped. Arrows from the phylogenetic tree indicate from which order the plant species originate (B, C).

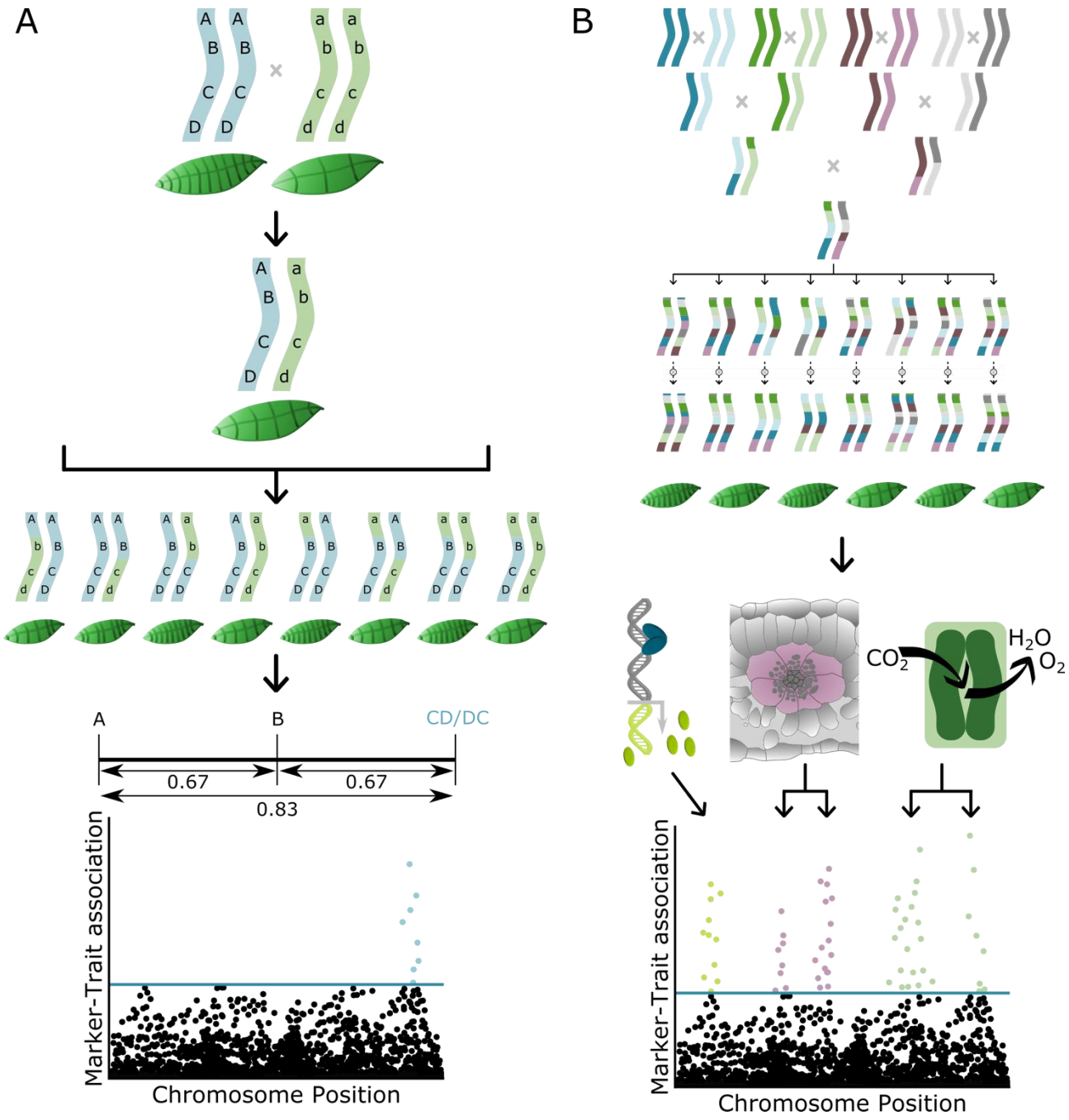

Fig. 3. Quantitative genetics in the context of C_4 photosynthesis. (A) A schematic for QTL mapping of leaf anatomical traits. Two homozygous parents, genotyped for four markers, A, B, C and D and differing in vein density are hybridised and advanced to form a bi-parental population that can be used to identify QTL associated with vein density (here located near markers C and D). Numbers show recombination fractions, which are used to position the QTL relative to flanking markers. (B) Population structure of a MAGIC pedigree followed by four generations of inter-crossing and self-fertilization. Progeny contain more genetic variation than that derived from a bi-parental design. Hypothetical plot showing how QTL associated with individually mapped C_4 phenotypes such as

gene expression, bundle sheath cell size or gas exchange parameters (*e.g.*, stomatal conductance, CO_2 assimilation, etc) can be mapped with one population.

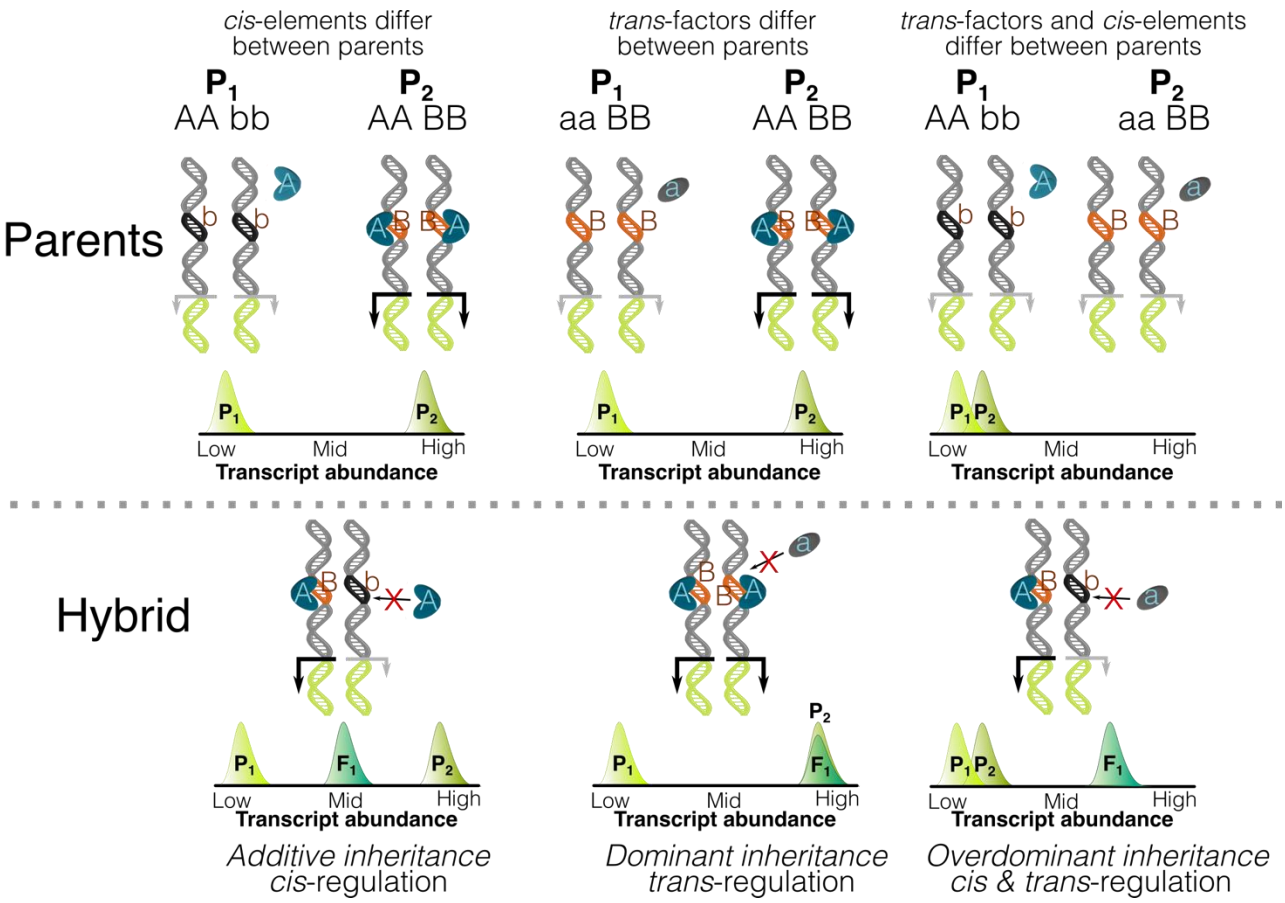

Fig. 4. Using breeding to understand the molecular basis of C_4 gene regulation. Parental populations that differ in transcript abundance can be due to multiple genetic effects that can be parsed by quantitative genetics. A simplified two loci model where one locus is a *cis*-element and the other an activating *trans*-factor is presented to illustrate how the molecular basis underpinning variations in gene expression can be determined by inheritance of gene expression in F_1 hybrids. If expression of a gene is controlled by changes in *cis*-regulation between parents, offspring exhibit additive expression patterns. If variation in expression is due to changes in *trans* between parents then offspring exhibit dominance deviation towards one parent. Lastly, if differences in gene expression between parents is due to both *cis* and *trans*- offspring demonstrate heterosis or overdominance.