

Genome Diversity and Evolution in Canine Transmissible Venereal Tumour



Andrea Strakova

**King's College
Department of Veterinary Medicine
University of Cambridge**

This dissertation is submitted for the degree of Doctor of Philosophy

September 2017

I dedicate this Thesis to my grandmother Jana, a veterinarian, who had hoped to be able to read it.

ABSTRACT

The canine transmissible venereal tumour (CTVT) is a contagious cancer that is naturally transmitted between dogs by the allogeneic transfer of living cancer cells during coitus. CTVT first arose several thousand years ago and has been reported in dog populations worldwide. The goals of this Thesis were (1) to gain further understanding of CTVT distribution patterns and prevalence around the world, (2) to use genetics to trace the historical spread of CTVT and (3) to map the genetic as well as phenotypic diversity of CTVT tumours around the world.

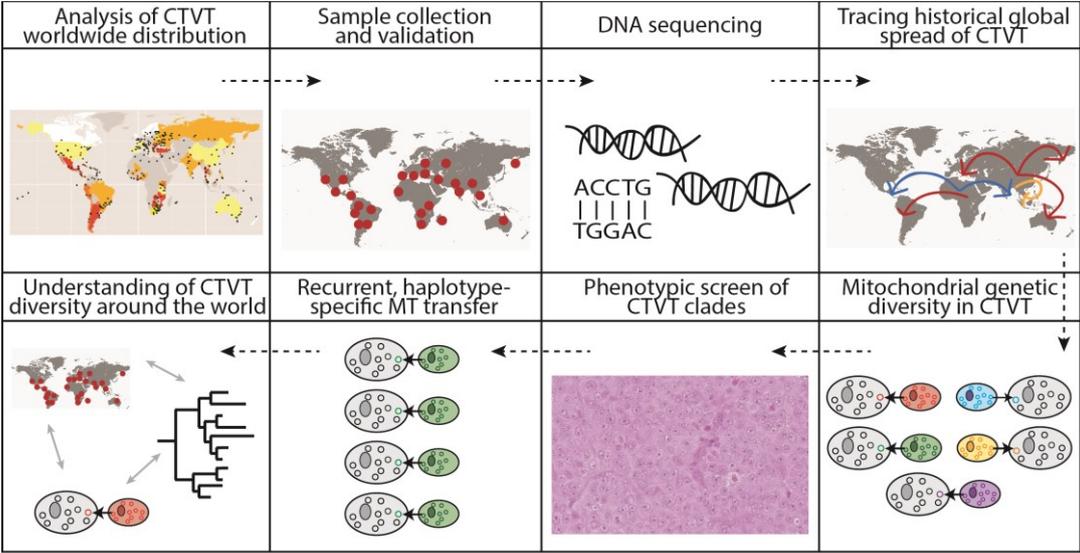
To understand the distribution patterns of CTVT, I obtained information from 645 veterinarians and animal health workers in 109 countries, and generated a snapshot of the locations in which this disease is found. Additionally, as preparation for further genetic analysis, I collected samples from over one thousand CTVT cases from more than 50 countries, optimised methods for high-throughput DNA extraction and quantification and optimised a qPCR-based assay for CTVT diagnosis and host contamination detection.

With the goal of tracing the historical spread of CTVT and learning about the genetic diversity of this disease, I sequenced complete mitochondrial genomes of 449 CTVT tumours and their matched hosts. The analysis of the CTVT mitochondrial diversity revealed that CTVT has captured mitochondrial DNA (mtDNA) through horizontal transfer events at least five times during the history of the lineage, delineating five tumour clades. CTVT appears to have spread rapidly around the world within the last 2,000 years, perhaps transported by dogs travelling along historic maritime trade routes. This work indicated that negative selection has operated to prevent accumulation of deleterious mutations in captured mtDNA, and that recombination has caused occasional mtDNA re-assortment. A histology-based screen of CTVT clades did not show any significant phenotypic differences between groups.

In order to determine how the five mtDNA clades relate to each other, I analysed data from 539 CTVT exomes. This revealed that a single canine mtDNA haplogroup has recurrently and recently undergone multiple horizontal transfer events. Analysis of this haplotype highlighted a number of candidate genetic variants which may be conferring a selective advantage to this haplotype in CTVT, possibly by influencing mtDNA transcription or replication.

Overall, genetic and phenotypic analysis of CTVT tumours from across the globe has broadened our understanding of CTVT diversity, and provided important insights into the biology of a unique transmissible cancer.

GRAPHICAL ABSTRACT



ACKNOWLEDGEMENTS

It is difficult to describe my gratitude to my PhD supervisor, Dr Elizabeth Murchison – thank you for your continuous support, advice, encouragement, excitement, inspiration and friendship. Thank you for this incredible opportunity to work on such an extra-ordinary topic; I have truly enjoyed every moment of it! Thank you for supporting my passion, and for the freedom to explore and discover. Without your expertise and brilliant ideas, I wouldn't have this fantastic project that has allowed me to discover the true magic of research.

I am incredibly grateful to all the veterinary collaborators who have contributed samples or other information to my research (listed in the Table of Collaborators) – without your help, this project wouldn't have been possible. Thank you for your dedication! The samples that you collected and night-time emails coming from the other side of the world were a fuel for my excitement. Thank you for your friendships and clinical veterinary inspiration. And thank you, Mir, for all your photos!

I would like to thank all the members of the Transmissible Cancer Group, past or present - Máire Ní Leathlobhair, Adrian Baez-Ortega, Tracy Wang, Kevin Gori, Max Stammnitz, Young Mi Kwon, Isobelle Bolton, Alexander Sampson, Karina Ferreira de Castro, Haleema Sadia, Stella Zhang, Jocelyn Bisson, Jose Tubio, Tim Coorens – thank you for your contributions, collaborations, advice, help, fabulous science chats, friendships and company in the lab (or in the field)! It has been a real privilege to work with you all and I wouldn't have been able to finish my project without you!

Importantly, I would like to thank Dr Katherine Hughes and Dr Hannah Bender for their invaluable histopathology (and other) advice and discussions. Many thanks go to Dr Ian Mickleburgh for his assistance with cloning experiments. I would also like to express my gratitude to Dr David Wedge, my advisor, and Dr Mike Hobart, my second advisor, for their insightful discussions and comments. My thanks go to all the people who assisted with questionnaire design and analysis, specifically Dr Michelle Morters and Dr Simon Frost. I am very grateful to the Wellcome Trust Sanger Institute staff - Cancer Genome Groups, IT support, Core Sequencing Facilities and Sample Management team. To all the support staff at the Department of Veterinary Medicine – thank you for answering my questions. And I am extremely grateful to countless others who contributed along the way – thank you, I appreciate every bit of your assistance and support.

I am extremely grateful to the Wellcome Trust, which provided funding for this project; without the support I would have never been able to carry out this work. Thank you to King's College, to the Department of Veterinary Medicine and to the University of Cambridge – it has been a true privilege to study here.

Finally, to my family – thank you for always being there!

TABLE OF COLLABORATORS

Individual providing assistance	Organisation	Contribution
Elizabeth P Murchison	University of Cambridge	PhD supervisor
David C Wedge	Wellcome Trust Sanger Institute, University of Oxford	PhD advisor and statistical advice
Katherine Hughes	University of Cambridge	Veterinary histopathology consultation
Hannah Bender	Murdoch University	Veterinary histopathology consultation
Inigo Martincorena	Wellcome Trust Sanger Institute	Advice on methods for analysing selection
German Tischler	Wellcome Trust Sanger Institute	Probe design filtering
Simon Frost	University of Cambridge	Questionnaire analysis advice
Young Seok Ju	Wellcome Trust Sanger Institute	Discussions and advice
Ludmil Alexandrov	Wellcome Trust Sanger Institute	Discussions and advice
Mike Hobart	University of Cambridge	Discussions and advice
Members of Transmissible Cancer Group	Position	Contribution
Máire Ní Leathlobhair	PhD student	Collaborative analysis of CTVT mtDNA sequencing data as discussed throughout the text
Adrian Baez-Ortega	PhD student	Collaborative analysis of CTVT exome sequencing data as discussed throughout the text
Isobelle Bolton	Part II pathology veterinary student	Histopathology analysis under my supervision
Alexander Sampson	Part II pathology veterinary student	Sample collection and CTVT genetic analysis under my supervision
Stella Zhang	Part II genetics student	Sample processing
Karina Ferreira de Castro	Visiting student	Sample collection and processing
Haleema Sadia	Visiting student	Sample collection and processing
Jocelyn L Bisson	Veterinary elective project student	Sample collection and processing
Tracy Wang	Laboratory manager	Sample collection and processing
Kevin Gori	Postdoctoral research associate	Discussions and advice
Jose Tubio	Postdoctoral research associate	Discussions and advice

Veterinary collaborators collecting samples under direction of AS	Organisation	Contribution
Ilona Airikkala-Otter	Worldwide Veterinary Service Centre, India	Sample collection
Janice L Allen	Animal Management in Rural and Remote Indigenous Communities (AMRRIC), Australia	Sample collection
Karen M Allum	World Vets, United States	Sample collection
Leontine Bansse-Issa	Stichting Dierenbescherming, Suriname	Sample collection
Cristobal Briceno	Universidad de Chile, Chile	Sample collection
Maria V Brignone	Universidad de Buenos Aires, Argentina	Sample collection
Artemio Castillo Domracheva	Corozal Veterinary Hospital, University of Panama	Sample collection
Anne Corrigan	St. George's University, Grenada	Sample collection
Hugh R Cran	The Nakuru District Veterinary Scheme Ltd, Kenya	Sample collection
Jane T Crawford	Animal Medical Centre, Belize	Sample collection
Stephen M Cutter	AMRRIC, Australia	Sample collection
Eric Davis	International Animal Welfare Training Institute, United States	Sample collection
Laura Delgadillo Keenan	Veterinary clinic Sr. Dog's, Mexico	Sample collection
Edward M Donelan	AMRRIC, Australia	Sample collection
Ibikunle A Faramade	National Veterinary Research Institute, Nigeria	Sample collection
Erika R Flores	International Fund for Animal Welfare, Mexico	Sample collection
Eleni Fotopoulou	Intermunicipal Stray Animals Care Centre, Greece	Sample collection
Skye N Fruean	Animal Protection Society of Samoa	Sample collection
Ricardo Gaitan	Animal Care Services, United States	Sample collection
Fanny Gallardo-Arrieta	University of Zulia, Venezuela	Sample collection
Olga Glebova	Veterinary clinic BIOCONTROL, Russia	Sample collection
Nona Goletsou	Veterinary clinic, Greece	Sample collection
Rodrigo F Häfelin Manrique	Veterinary clinic El Roble, Chile	Sample collection

Joaquim JGP Henriques	OnevetGroup, Portugal	Sample collection
David Hyeroba	Entebbe Wildlife Sanctuary, Uganda	Sample collection
Natalia Ignatenko	Veterinary clinic Zoovetservis, Ukraine	Sample collection and questionnaire distribution
Debra Kamstock	Colorado State University, United States	Sample collection
Yacoub Kane	EISMV, Senegal	Sample collection
Cathy King and World Vets staff and volunteers	World Vets, United States	Sample collection
Cesar A Koci	Universidad Catolica de Cordoba, Argentina	Sample collection
Debbie Koenig	World Vets, United States	Sample collection and questionnaire distribution
Adriana Krupa	Universitair Diergeneeskundig Centrum Utrecht, Netherlands	Sample collection
Steve Kruzeniski	World Vets, Nicaragua/United States	Sample collection
Marta Lanza-Perea	St. George's University, Grenada	Sample collection
Remo Lobetti	Bryanston Veterinary Hospital, South Africa	Sample collection
Adriana M Lopez Quintana	Veterinary Clinic Lopez Quintana, Uruguay	Sample collection
Thibault Losfelt	Clinique Veterinaire de Grand Fond, Reunion, France	Sample collection
Gabriele Marino	University of Messina, Italy	Sample collection
Simón Martínez Castañeda	Universidad Autonoma del Estado de Mexico, Mexico	Sample collection
Mayra F Martínez-López	Universidad de las Americas, Ecuador	Sample collection
Michael Meyer	Touray & Meyer Vet Clinic, Gambia	Sample collection
Scott Moroff		Sample collection
Berna Nakanwagi	The Kampala Veterinary Surgery, Uganda	Sample collection
Andrigo B De Nardi	Sao Paulo State University, Brazil	Sample collection
Frances E Nargi	MIT Lincoln Laboratory, United States	Sample collection
Karter Neal	Santa Cruz Veterinary Clinic, United States	Sample collection
Winifred Neunzig	World Vets, United States	Sample collection
Sally Nixon	Vets Beyond Borders, Australia	Sample collection
Marsden M Onsare	Aniworld veterinary clinic, Kenya	Sample collection
Antonio Ortega-Pacheco	Autonomous University of Yucatan, Mexico	Sample collection

Francisco Javier Pedraza Ordonez	Universidad de Caldas, Colombia	Sample collection
Maria C Peleteiro	University of Lisbon, Portugal	Sample collection
Katherine Polak	Soi Dog Foundation, Thailand	Sample collection
Ruth J Pye	Vets Beyond Borders, Australia	Sample collection
Juan Carlos Ramirez Ante	Universidad de Caldas, Colombia	Sample collection
Oscar Rarieya	Veterinary clinic, Kenya	Sample collection
John F Reece	Help in Suffering, India	Sample collection
Jose Rojas Gutierrez	Veterinary clinic Dr Jose Rojas, Chile	Sample collection
Maya Rybko	Veterinary clinic, Ukraine	Sample collection
Rodrigo dos Santos Horta	Federal University of Minas Gerais, Brazil	Sample collection
Sheila K Schmeling	Corozal Veterinary Clinic, Belize	Sample collection
Olga Shamanova	Veterinary clinic Vetmaster, Russia	Sample collection
Alan Sherlock/Thinlay N Bhutia	Sikkim Anti Rabies & Animal Health, India	Sample collection
Richard K Ssuna	Lilongwe SPCA, Malawi	Sample collection
Audrey E Steenland-Smit	Stichting Dierenbescherming, Suriname	Sample collection
Alla J Svitich	State Hospital of Veterinary Medicine, Ukraine	Sample collection
Irina Swarisch	Veterinary clinic Canine Heart, Ukraine	Sample collection
Lester Tapia	World Vets, Nicaragua/United States	Sample collection
Ismail Thoya Ngoka	Kenya Society for Protection and Care of Animals, Kenya	Sample collection
Bogdan A Vițălaru	Faculty of Veterinary Medicine, Bucharest, Romania	Sample collection
Anna P de Vos	Ladybrand Animal Clinic, South Africa	Sample collection
Johan P de Vos	Veterinary Oncology Referral Centre De Ottenhorst, Netherlands	Sample collection
Sevil Vural	Ankara University, Turkey	Sample collection
David Walker		Sample collection
Oliver Walkinton	Vets Beyond Borders, Australia	Sample collection
Alvaro S Wehrle-Martinez	National University of Asuncion, Paraguay	Sample collection
Mirjam G van der Wel	Animal Anti Cruelty League, South Africa	Sample collection
Sophie Widdowson	Lilongwe SPCA, Malawi	Sample collection

Other collaborators providing useful information or assistance with sample collection		
Animal Balance project	Anna Kuznetsova	Clare Rebbeck
Cecilia Amaral	Olakunle AbdulRasaq Lawal	Ceseltina Semedo
Ekaterina Batrakova	Mihran Lazyan	Santiago Soto
Rafaela Bortolotti Viera	Thabo Lerotholi	Ivan Stoikov
Matthew Breen	Marco Lima-Maigua	Lester Tapia
Nikki Brown	Jimmy Loayza-Feijoo	Mirela Tinucci Costa
Austin Burt	Kati Loeffler	Adam Wagner
Cinzia Cantacessi	Mayra Lopez-Bucheli	David Walker
Patrick Chinnery	Margarita Mancero-Albuja	Guo-Dong Wang
Fernando Constantino Casas	Cynthia Marchiori Bueno	Robin Weiss
	Talita M Morata Raposo	Kevin Xie
John Cooper	Luis Martinez-Lopez	Ting-Ting Yin
Amici Cannis Cotacachi	Alfredo Martinez-Meza	Yaping Zhang
Lytvynenko Dmytro	Edward Migneco	Veterinary workers of Pet Centre (UVAS, Lahore, Pakistan)
Adela Espinoza	Michelle Morters	
Ariberto Fassati	Edmore Mushoriwa	
David Hanzlcek	Claudio Murgia	Staff at the WVS
Rafael Ricardo Huppés	Alvira Murison Swartz	International Training Centre in Ooty (India)
Matilde Jimenez-Coello	Edwin Ortiz-Rodríguez	
Patrick Kelly	Lisa Pellegrini	Staff of Veterinary clinic “El Roble”
Andrew King	Gerry Polton	
Tatiana Korytina	Freddy Proano-Perez	Students from St. George’s University (Grenada)
Clara Lee Arnold, Kelli Blank, Joseph Morgan, Amanda Zehnder Thielman and the team at Hopkins Belize Humane Society	Jane, Miriam and the team at the Animal Medical Centre, Belize Clare Bryant Hannah V Siddle	Jeremy Skepper, Lyn Carter and the team at the Cambridge Advanced Imaging Centre
Wellcome Trust Sanger Institute Staff		Contribution
Lucy Hildyard, Sarah Sims and the DNA pipelines		Sample and DNA pipelines management
Core Sequencing Facility		DNA sequencing
IT groups		IT support
Members of the Cancer Genome Group		Data management and support, general advice

TABLE OF CONTENTS

ABSTRACT	i
GRAPHICAL ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
TABLE OF COLLABORATORS	v
TABLE OF CONTENTS	x
LIST OF FIGURES	xix
LIST OF TABLES	xxii
LIST OF APPENDICES	xxiv
LIST OF ABBREVIATIONS AND SYMBOLS	xxv
PREFACE	xxvi

CHAPTER 1

CANINE TRANSMISSIBLE VENEREAL TUMOUR AS THE CANCER WHICH SURVIVED 1

1.1 CANCER EVOLUTION	2
1.1.1 Cancer as a disease	2
1.1.2 Clonal evolution of cancer	4
1.1.3 Germline and somatic variation	5
1.1.4 Evolutionary selection in cancer	6
1.2 CANINE TRANSMISSIBLE VENEREAL TUMOUR	7
1.2.1 Transmissible cancers in nature	7
1.2.2 Clinical features of CTVT	8
1.2.2.1 <i>Gross clinical features</i>	8
1.2.2.2 <i>Histopathological features</i>	11
1.2.2.3 <i>Cytological features</i>	11
1.2.3 Diagnosis and treatment of CTVT	13
1.2.3.1 <i>Diagnosis</i>	13
1.2.3.2 <i>Treatment</i>	14
1.2.4 Mode of CTVT transmission	14
1.2.5 CTVT growth phases, experimental transmission studies and immune system interaction	16
1.2.5.1 <i>Experimental transmission studies</i>	16
1.2.5.2 <i>Adaptations for immune system escape</i>	16
1.2.6 Worldwide distribution of CTVT	17

1.2.7 Confirmation of the clonal nature of CTVT	18
1.2.7.1 Transplantation studies	18
1.2.7.2 Cytogenetics	19
1.2.7.3 Genetics	19
1.2.7.3.1 LINE-MYC rearrangement	19
1.2.7.3.2 Microsatellites and the Major Histocompatibility Complex	19
1.2.7.3.3 Genome	20
1.2.8 Age, origin and clonal evolution of CTVT	20
1.2.8.1 Age and origin of CTVT	20
1.2.9 CTVT genome	22
1.2.10 Diversity amongst CTVT tumours	23
1.3 CANINE GENETICS	25
1.3.1 Dog domestication and global dispersal	25
1.3.2 Breed formation	25
1.4 MITOCHONDRIAL GENOME	27
1.4.1 Structure and function of the mitochondrial genome	27
1.4.2 Mitochondrial DNA homoplasmy and heteroplasmy	29
1.4.3 Mitochondrial DNA mutations in cancer	30
1.4.4 Canine mitochondrial DNA	31
1.4.5 Mitochondrial DNA in CTVT	33
1.5 GOALS OF THIS THESIS	34

CHAPTER 2

WORLDWIDE DISTRIBUTION AND PREVALENCE OF CANINE TRANSMISSIBLE VENEREAL TUMOUR	39
2.1 INTRODUCTION	40
2.1.1 Previous knowledge of CTVT worldwide distribution	40
2.1.2 Goals of my work	40
2.2 MATERIALS AND METHODS	42
2.2.1 Questionnaire design and distribution	42
2.2.2 Questionnaire data analysis	42
2.3 RESULTS	44
2.3.1 Historical worldwide CTVT distribution	44

2.3.2 Current worldwide CTVT distribution	48
2.3.3 Status of dogs infected with CTVT	52
2.3.4 Natural transmission of CTVT to wild canids	53
2.3.5 Factors influencing CTVT prevalence	54
2.3.6 Metastasis and treatment of CTVT	55
2.4 DISCUSSION	58
2.4.1 Disappearance of CTVT from the United Kingdom	58
2.4.2 Implications of the study	59
2.4.3 Challenges and limitations of the study	60
2.5 CONCLUSION	61

CHAPTER 3

CANINE TRANSMISSIBLE VENEREAL TUMOUR SAMPLE COLLECTION AND VALIDATION	63
3.1 INTRODUCTION	64
3.1.1 CTVT sample collection	64
3.1.2 CTVT validation	65
3.1.3 Goals of this Chapter	67
3.2 MATERIALS AND METHODS	68
3.2.1 Sample collection	68
3.2.1.1 Collaborators performing sample collection	68
3.2.1.2 Sample collection kit	68
3.2.1.3 Sample collection procedure	69
3.2.1.4 Sample shipment	70
3.2.1.5 Sample storage and initial processing	70
3.2.2 DNA extraction and quantification	71
3.2.2.1 Protocol optimisation	71
3.2.2.2 Variation in DNA yield between tissue types	71
3.2.2.3 DNA extraction and quantification protocols	73
3.2.2.3.1 DNA extraction from tissues	73
3.2.2.3.2 DNA extraction from blood	74
3.2.2.3.3 DNA quantification using pico green	76
3.2.2.3.4 DNA quantification using Qubit	76

3.2.2.3.5 <i>Whole genome amplification (WGA) protocol</i>	77
3.2.3 CTVT diagnosis and estimation of host contamination levels	77
3.2.3.1 <i>CTVT diagnosis</i>	77
3.2.3.2 <i>Estimation of host contamination levels</i>	78
3.2.3.3 <i>Analysis of qPCR results</i>	79
3.2.4 Processing and validation of histology samples	80
3.3 RESULTS	81
3.3.1 Summary of genetic samples processed	81
3.3.2 Genetic samples collected	81
3.3.3 CTVT diagnostic validation	82
3.3.4 Set of validated CTVT samples	83
3.3.5 Host contamination levels in confirmed CTVT samples	87
3.3.6 Histology samples processed	88
3.3.7 Dogs with the complete set of samples	88
3.4 DISCUSSION	89
3.4.1 Summary of sample collection process	89
3.4.2 CTVT diagnosis through <i>LINE-MYC</i> amplification	89
3.4.2.1 <i>Interpretation of LINE-MYC-negative tumours</i>	90
3.4.2.2 <i>Interpretation of LINE-MYC positive hosts</i>	91
3.4.3 Clinical data summary of CTVT samples collected	91
3.4.4 CTVT samples in my future work	93

CHAPTER 4

MITOCHONDRIAL GENETIC DIVERSITY, SELECTION AND RECOMBINATION IN CANINE TRANSMISSIBLE VENEREAL TUMOUR	95
4.1 INTRODUCTION	96
4.1.1 Mitochondria in CTVT	96
4.1.2 Genetic changes in mtDNA	97
4.1.3 MtDNA recombination	98
4.1.4 Genetic diversity in CTVT	98
4.1.5 Linking phenotypic diversity to genetic changes in CTVT	99
4.1.6 Goals of this Chapter	99

4.2 MATERIALS AND METHODS	101
4.2.1 Sample collection, DNA extraction and CTVT diagnosis	101
4.2.2 Library preparation and DNA sequencing	101
4.2.3 Sequencing validation and quality assessment	102
4.2.3.1 <i>Copy number plots</i>	103
4.2.4 MtDNA copy number calculations	103
4.2.5 Substitution calling	105
4.2.5.1 <i>Sequence alignment</i>	105
4.2.5.2 <i>Extraction and filtering</i>	105
4.2.5.3 <i>Variant allele fraction (VAF) plots</i>	106
4.2.5.4 <i>Post-processing</i>	108
4.2.5.4.1 <i>Somatic substitutions in tumours with matched hosts</i>	108
4.2.5.4.2 <i>Somatic substitutions in tumours without matched hosts</i>	108
4.2.5.4.3 <i>Germline substitutions in hosts</i>	109
4.2.5.5 <i>Additional quality checks and validation</i>	109
4.2.5.6 <i>Host contamination</i>	111
4.2.5.7 <i>Recurrent mutations and back-mutations</i>	111
4.2.5.8 <i>Predicted functional consequences of substitutions</i>	111
4.2.5.9 <i>Extracting substitution variants from publicly available dog sequences</i>	111
4.2.6 Indel calling	112
4.2.6.1 <i>Extraction and filtering</i>	112
4.2.6.2 <i>Post processing</i>	113
4.2.6.2.1 <i>Somatic indels</i>	113
4.2.6.2.2 <i>Germline indels</i>	113
4.2.6.3 <i>Recurrent indels</i>	113
4.2.6.4 <i>Variant Allele Fraction (VAF) calculation for indels</i>	114
4.2.6.5 <i>Predicted functional consequences of indels</i>	114
4.2.7 Phylogenetic analyses	114
4.2.7.1 <i>Phylogenetic trees</i>	114
4.2.7.2 <i>Confirmation of individual horizontal transfer events</i>	115
4.2.7.2.1 <i>Classification of tumour substitutions</i>	115
4.2.7.2.2 <i>Definition of a CTVT clade</i>	118
4.2.7.3 <i>Estimated timing of clade divergence</i>	118
4.2.7.3.1 <i>Timing based on nuclear DNA</i>	119
4.2.7.3.2 <i>Timing based on number of cell divisions</i>	119
4.2.7.3.3 <i>Timing based on number of mutations per year</i>	119
4.2.8 Haplotype analysis	119
4.2.8.1 <i>Haplotype nomenclature</i>	119

4.2.8.1.1 <i>Host haplotypes</i>	119
4.2.8.1.2 <i>Tumour haplotypes</i>	120
4.2.8.2 <i>Reconstructed donor haplotypes</i>	120
4.2.9 Mitochondrial recombination analysis	121
4.2.9.1 <i>Automated mtDNA recombination analysis</i>	121
4.2.9.2 <i>Analysis of phylogenetic outliers</i>	121
4.2.9.3 <i>Haplotype cloning and sequencing</i>	121
4.2.9.4 <i>PacBio long-read sequencing and analysis</i>	122
4.2.10 Selection analyses	123
4.2.10.1 <i>Variant allele fraction</i>	123
4.2.10.1.1 <i>Substitutions</i>	123
4.2.10.1.2 <i>Indels</i>	124
4.2.10.2 <i>dN/dS</i>	124
4.2.11 Nuclear copies of mtDNA (NuMT) analysis	124
4.2.11.1 <i>NuMTs in CanFam3.1</i>	124
4.2.11.2 <i>Somatically acquired NuMTs</i>	125
4.2.12 Histopathology analysis	125
4.2.12.1 <i>Histopathology screen design</i>	125
4.2.12.2 <i>Scoring criteria</i>	126
4.2.12.2.1 <i>Selecting a field</i>	126
4.2.12.2.2 <i>Scoring mitotic index</i>	126
4.2.12.2.3 <i>Scoring percentage of lymphocyte infiltration in parenchyma</i>	128
4.2.12.2.4 <i>Scoring percentage of lymphocyte infiltration at interface</i>	128
4.2.12.2.5 <i>Scoring haemorrhage</i>	128
4.2.12.2.6 <i>Scoring necrosis</i>	129
4.2.12.3 <i>Statistical analysis</i>	130
4.2.12.3.1 <i>Random effects</i>	131
4.3 RESULTS	133
4.3.1 <i>Samples analysed in this Chapter</i>	133
4.3.2 <i>Sequencing of mtDNA</i>	133
4.3.3 <i>Number of mtDNA horizontal transfer events in CTVT</i>	134
4.3.4 <i>Relative time since mtDNA horizontal transfer events</i>	137
4.3.5 <i>Phylogeography of CTVT</i>	139
4.3.6 <i>Negative selection in CTVT mtDNA</i>	142
4.3.7 <i>Ancient mtDNA recombination in CTVT</i>	143
4.3.8 <i>Recent mtDNA recombination in CTVT</i>	145
4.3.9 <i>Histopathology analysis of CTVT clades</i>	147

4.3.9.1 <i>Mitotic index and lymphocyte density</i>	147
4.3.9.2 <i>Haemorrhage and necrosis</i>	150
4.4 DISCUSSION	151
4.4.1 Historical spread of CTVT	151
4.4.2 Mechanisms of mtDNA horizontal transfer	153
4.4.3 Frequency of mtDNA horizontal transfer	154
4.4.4 MtDNA recombination	155
4.4.5 Negative selection in CTVT	155
4.4.6 Phenotypic differences between CTVT clades	157
4.4.7 Host genetic variation	158
4.4.8 Future directions and next steps	158

CHAPTER 5

RECURRENT AND RECENT HAPLOTYPE-SPECIFIC MITOCHONDRIAL DNA HORIZONTAL TRANSFER IN CANINE TRANSMISSIBLE VENEREAL TUMOUR	161
5.1 INTRODUCTION	162
5.1.1 MtDNA transfer in CTVT and other cancers	162
5.1.2 Positive selection of mtDNA haplotypes	163
5.1.3 Goals of my work	164
5.2 MATERIALS AND METHODS	167
5.2.1 Sample collection and DNA extraction	167
5.2.2 CanFam3.1. pull down kit design	167
5.2.3 DNA sequencing	168
5.2.4 Performance of exome pull-down and sequencing	168
5.2.5 Sample validation	169
5.2.6 Substitution calling	170
5.2.7 Indel calling	170
5.2.8 Phylogenetic analyses	170
5.2.9 Analysis of A1d1 haplogroup	171
5.2.9.1 <i>Donor haplotypes</i>	171
5.2.9.2 <i>A1d1 haplogroup frequency and distribution</i>	171
5.2.9.3 <i>A1d1 haplogroup copy number</i>	171

5.2.9.4 <i>A1d1 haplogroup long-term heteroplasmy analysis</i>	172
5.2.9.5 <i>A1d1 haplogroup substitutions</i>	172
5.2.9.6 <i>A1d1a haplogroup indels</i>	173
5.2.9.7 <i>D-loop repeat polymerase chain reaction</i>	173
5.2.9.8 <i>mtDNA recombination analysis</i>	174
5.2.9.8.1 <i>Illumina data analysis</i>	174
5.2.9.8.2 <i>Long read sequencing</i>	174
5.2.10 <i>Somatic tumour substitutions and timing analysis</i>	174
5.2.11 <i>Phenotypic analyses of tumours in A1d1 haplogroup</i>	175
5.2.11.1 <i>Histopathology scoring</i>	175
5.3 RESULTS	176
5.3.1 <i>Samples analysed in this Chapter</i>	176
5.3.2 <i>Recurrent and recent haplotype-specific mtDNA horizontal transfer</i>	176
5.3.3 <i>Analysis of the A1d1 haplogroup</i>	180
5.3.3.1 <i>Identification of donor haplotypes</i>	180
5.3.3.2 <i>Worldwide distribution</i>	182
5.3.3.3 <i>mtDNA copy number</i>	183
5.3.3.3.1 <i>mtDNA copy number in hosts</i>	183
5.3.3.3.2 <i>mtDNA copy number in tumours</i>	184
5.3.3.4 <i>Long-term heteroplasmy</i>	185
5.3.3.5 <i>Genetic features</i>	185
5.3.3.5.1 <i>Substitutions</i>	186
5.3.3.5.2 <i>Indels</i>	187
5.3.3.5.3 <i>Copy number variation of 10bp repeat unit in the mtDNA control region</i>	188
5.3.3.5.4 <i>mtDNA recombination</i>	189
5.3.4 <i>Somatic variant analysis</i>	191
5.3.4.1 <i>Timing analysis using somatic substitutions</i>	192
5.3.5 <i>Phenotypic analysis</i>	193
5.3.5.1 <i>Histopathology analysis</i>	193
5.4 DISCUSSION AND FUTURE DIRECTIONS	195
5.4.1 <i>Recurrent and recent haplotype specific mtDNA horizontal transfer</i>	195
5.4.2 <i>Hypotheses and possible link to Vincristine treatment</i>	198
5.4.3 <i>Exploratory analyses and future directions</i>	202
5.4.3.1 <i>Clinical study of vincristine treatment response</i>	203
5.4.3.2 <i>Transmission electron microscopy analysis of A1d1 mitochondria</i>	206
5.4.3.3 <i>CTVT management during treatment</i>	207
5.4.3.4 <i>Future directions</i>	208

CHAPTER 6

FINAL DISCUSSION AND FUTURE STUDIES	211
6.1 THESIS SUMMARY	212
6.2 IMPLICATIONS OF THIS RESEARCH	215
6.2.1 Study of a transmissible cancer	215
6.2.2 Relevance of my PhD work	216
6.3 LIMITATIONS OF THIS RESEARCH	219
6.4 FUTURE DIRECTIONS	221
6.5 CONCLUSIONS	224
REFERENCES	225
APPENDIX	248

LIST OF FIGURES

CHAPTER 1

- Figure 1-1 Model of clonal evolution of cancer.
- Figure 1-2 Germline and somatic variation.
- Figure 1-3 Clinical presentations of canine transmissible venereal tumour (CTVT).
- Figure 1-4 Extra genital presentations of canine transmissible venereal tumour.
- Figure 1-5 Histology image of canine transmissible venereal tumour cells.
- Figure 1-6 Cytology image of canine transmissible venereal tumour cells.
- Figure 1-7 Map of CTVT worldwide distribution published before this PhD work was commenced.
- Figure 1-8 Transmission of canine transmissible venereal tumour (CTVT) as a clonal cell lineage between unrelated dogs.
- Figure 1-9 Genomic copy number for CTVT tumours 24T (Australia) and 79T (Brazil), as published in (Murchison et al. 2014).
- Figure 1-10 Illustration of mtDNA heteroplasmy.
- Figure 1-11 Summary of processes by which a heteroplasmic mtDNA variant may become homoplasmic.
- Figure 1-12 Phylogenetic tree of dog (unlabelled) and wolf (open squares) haplotypes, obtained from (Savolainen et al. 2002).
- Figure 1-13 Representation of maximum parsimony phylogeny of CTVT mitochondria adapted from (Rebbeck et al. 2011).

CHAPTER 2

- Figure 2-1 Global distribution of published reports of CTVT.
- Figure 2-2 Worldwide CTVT distribution and prevalence.
- Figure 2-3 Map indicating CTVT prevalence estimated by each respondent.
- Figure 2-4 Higher magnification map showing detailed distribution of CTVT prevalence estimates in Europe.
- Figure 2-5 Average reported CTVT prevalence by continent.
- Figure 2-6 Status of dogs with CTVT.
- Figure 2-7 Relationship between spaying/neutering and CTVT prevalence.
- Figure 2-8 Factors associated with CTVT prevalence.
- Figure 2-9 Socio-economic and climatic factors associated with CTVT prevalence.
- Figure 2-10 CTVT metastasis.
- Figure 2-11 CTVT treatment.
- Figure 2-12 Disappearance of CTVT from the United Kingdom during the twentieth century.

CHAPTER 3

- Figure 3-1 Schematic diagram of (A) *LINE-MYC* rearrangement and (B) *CDKN2A* homozygous deletion in CTVT.
- Figure 3-2 Summary of goals and workflow for Chapter 3.
- Figure 3-3 Sample collection kit sent to collaborators.
- Figure 3-4 CTVT samples after being assigned a database number and prepared for long-term storage.
- Figure 3-5 Kernel density estimate graph of DNA yield for different tissue types.
- Figure 3-6 The Qiagen DNAeasy DNA extraction protocol.
- Figure 3-7 Example standard curves for *CDKN2A*, *LINE-MYC* and *ACTB* based on two-fold serial dilutions.

Figure 3-8 Summary of tumour samples collected and processed at each step of the diagnostic pipeline.

Figure 3-9 Relative DNA input of *LINE-MYC* for 1036 tumours and 834 hosts normalised to *ACTB*.

Figure 3-10 Map showing locations of confirmed CTVT tumours.

Figure 3-11 Relative *CDKN2A* input for 873 tumours and 652 hosts normalised to *ACTB*.

Figure 3-12 Summary of histology samples available at each stage of the validation process.

Figure 3-13 Summary of individuals with the complete set of validated samples.

CHAPTER 4

Figure 4-1 Horizontal transfer of mitochondrial genomes in CTVT.

Figure 4-2 Mitochondrial genome sequencing coverage in 338 host samples and 449 CTVT tumour samples.

Figure 4-3 Copy number plots showing depth across the whole genome for 10kb windows.

Figure 4-4 Variant allele fraction plot for (A) a host matched to (B) a CTVT tumour.

Figure 4-5 Example of VAF plot for tumour without matched host.

Figure 4-6 Diagram illustrating the presence of germline, somatic and potential somatic substitutions in CTVT tumours.

Figure 4-7 Examples of mitotic figures to be counted in the mitotic figure scoring system.

Figure 4-8 Examples of haemorrhage within the tumour parenchyma.

Figure 4-9 Examples of necrosis.

Figure 4-10 Command to utilise GLMER model in R.

Figure 4-11 Geographical locations of CTVT samples analysed in this Chapter.

Figure 4-12 mtDNA copy number in CTVT and host samples.

Figure 4-13 CTVT has acquired mtDNA by horizontal transfer at least five times and CTVT mtDNA clades 1-5 all arose from dog mtDNA clade A.

Figure 4-14 Reconstructed donor haplotypes for CTVT mtDNA clades 1 to 5.

Figure 4-15 Number of somatic substitution mutations per CTVT tumour.

Figure 4-16 Geographical distribution of CTVT clades.

Figure 4-17 Simplified representation of maximum likelihood phylogenetic trees for each clade.

Figure 4-18 Cumulative distribution functions for variant allele fraction (VAF) for tumour (A) substitutions and (B) indels.

Figure 4-19 dN/dS for somatic nonsense and missense substitutions.

Figure 4-20 Ancient mtDNA recombination in CTVT.

Figure 4-21 Section of the phylogenetic tree showing sample 559T with its abnormal branch length.

Figure 4-22 Variant allele fraction plot representing sample 559T, identified as a phylogenetic outlier on the mtDNA tree.

Figure 4-23 Recombination detected in tumour 559T (Nicaragua).

Figure 4-24 Comparison of histopathological parameters between clade 1 and clade 2.

Figure 4-25 A model for how CTVT might have spread around the world.

CHAPTER 5

Figure 5-1 Diagram explaining how the analysis of exome sequencing data will allow me to understand the relationships between five CTVT mitochondrial clades analysed in detail in Chapter 4.

Figure 5-2 Proportion of exome with at least 30X coverage in 509 host samples and 551 CTVT tumour samples.

Figure 5-3 Coverage depth across protein coding regions (exons).

Figure 5-4 Recurrent and recent haplotype specific mtDNA horizontal transfer.

Figure 5-5 Worldwide distribution of CTVT mtDNA clades 5, 8-18.

Figure 5-6 Schematic diagram of the A1d1 haplogroup phylogenetic tree adapted from Fregel et al.

Figure 5-7 Distribution of A1d1 haplogroup in the current normal canine population around the world.

Figure 5-8 mtDNA copy number of A1d1 haplogroup versus other haplogroups in normal ovarian and testicular tissue.

Figure 5-9 Tumour mtDNA copy number adjusted for host contamination in different CTVT mitochondrial clades.

Figure 5-10 Predicted functional consequences of variants shared by all CTVT samples in clades 5, 8-18 with donor haplotypes in haplogroup A1d1 or sub-haplogroup A1d1a.

Figure 5-11 Graphical illustration of indel common to all haplotypes in A1d1a haplogroup.

Figure 5-12 Copy number variation of the 10bp repeat unit within the canine mtDNA control region.

Figure 5-13 MtDNA haplotypes detected using long-read sequencing in 559T and 1315T.

Figure 5-14 Suggested scenarios explaining relationship between CTVT tumours 559T and 1315T.

Figure 5-15 Number of somatic variants acquired by each clade after the horizontal transfer event.

Figure 5-16 Histopathology analysis of CTVT clades 5, 8-18 in comparison with other CTVT clades.

Figure 5-17 Phylogenetic structure of A1d1 haplogroup.

Figure 5-18 Schematic comparison between the human and canine mtDNA control region.

Figure 5-19 Diagrams representing the different hypotheses which may explain the recurrent, recent and A1d1 haplogroup-specific mtDNA horizontal transfer.

Figure 5-20 Clinical study of vincristine treatment response in Belize.

Figure 5-21 Transmission electron microscope image of a CTVT tumour sample.

Figure 5-22 Summary of future experiments which could be performed to understand the phenomenon of A1d1 haplogroup specific recurrent and recent mtDNA horizontal transfer event.

CHAPTER 6

Figure 6-1 Summary of approaches used in my PhD thesis to understand evolution and genetic diversity of CTVT around the world.

LIST OF TABLES

CHAPTER 2

Table 2-1 Historical reports confirming presence of CTVT in the United Kingdom in the 19th and 20th centuries.

CHAPTER 3

Table 3-1 Summary of DNA yield obtained from different tissue types.

Table 3-2 Primers used in the CTVT diagnostic quantitative PCR reaction.

Table 3-3 Input reagent volumes and concentrations for CTVT diagnostic quantitative PCR reaction.

Table 3-4 Cycling conditions for CTVT diagnostic quantitative PCR reaction.

Table 3-5 Primers used in the quantitative PCR reaction to estimate host contamination levels.

Table 3-6 Summary of tumour locations for individuals with suspected CTVT tumours.

Table 3-7 Summary of tumour locations for individuals with confirmed CTVT tumours.

Table 3-8 Summary of clinical data for individuals with confirmed CTVT tumours.

Table 3-9 Summary of breed information for individuals with confirmed CTVT tumours.

CHAPTER 4

Table 4-1 List of substitutions excluded from the analysis as a result of additional quality checks.

Table 4-2 List of substitutions that had substantial support and were rescued after being discarded due to proximity to an indel.

Table 4-3 Somatic indels discarded from the analysis as miscalls.

Table 4-4 Classification of tumour substitutions.

Table 4-5 Primers used for PCR reaction spanning a region between MT positions 4450-5024.

Table 4-6 Input reagent volumes and concentrations for PCR reaction spanning a region between MT positions 4450-5024.

Table 4-7 Cycling conditions for PCR reaction spanning a region between MT positions 4450-5024.

Table 4-8 PacBio sequencing reads used for phasing of individual haplotypes in tumour 559T.

Table 4-9 Summary of random effects and number of cases within each category.

Table 4-10 Summary of random effects and reasoning for why they were included in the screen.

Table 4-11 Timing analysis - part 1.

Table 4-12 Timing analysis - part 2.

Table 4-13 GLMER analysis implemented in R.

Table 4-14 Median and interquartile range (IQR) for all histopathology parameters.

Table 4-15 Standard deviation for all histopathology parameters.

Table 4-16 Presence of haemorrhage and necrosis in clade 1 and clade 2.

CHAPTER 5

Table 5-1 Primers used to assess length variation of mtDNA control region caused by polymorphisms in copy number of a 10bp unit.

Table 5-2 Polymerase chain reaction conditions used to assess length variation of mtDNA control region caused by polymorphisms in copy number of a 10bp unit.

Table 5-3 Sub-haplogroups of the A1d1 haplogroup.

Table 5-4 Locations and number of samples in clades 5, 8-18 and their phylogenetically closest haplotype present in the current dog population.

Table 5-5 Long-term low-level heteroplasmy of A1d1 haplogroup.

Table 5-6 Substitutions shared by all mtDNA haplotypes in haplogroups A1d1, A1d1a, A1d1a1.

Table 5-7 Indels shared by all mtDNA haplotypes in haplogroups A1d1, A1d1a, A1d1a1.

Table 5-8 Average number of somatic and potential somatic substitutions in clades 5, 8-18 and estimated timing of mtDNA horizontal transfer events.

LIST OF APPENDICES

CHAPTER 2

- Additional file 2-1 Questionnaire on canine transmissible venereal tumour worldwide distribution.
- Additional file 2-2 Contemporary and historical reports providing a numerical value for CTVT prevalence.
- Additional file 2-3 Summary of CTVT prevalence data by country.
- Additional file 2-4 Contemporary and historical reports of naturally occurring CTVT cases in the published literature.

CHAPTER 3

- Additional file 3-1 Owner consent form.
- Additional file 3-2 CTVT sample collection information sheet.
- Additional file 3-3 CTVT clinical data collection sheet.
- Additional file 3-4 Sample collection leaflet.

CHAPTER 4

- Additional file 4-1 Confirmation of CTVT diagnosis.
- Additional file 4-2 Sequencing coverage and tumour cell fraction.
- Additional file 4-3 Single point substitution variant lists.
- Additional file 4-4 Sample information.
- Additional file 4-5 Annotation of single point substitutions.
- Additional file 4-6 Publicly available mitochondrial dog genomes used in the study.
- Additional file 4-7 Summary of small insertions and deletions (indels).
- Additional file 4-8 Annotation of insertions and deletions (indels).
- Additional file 4-9 Maximum likelihood phylogenetic trees for CTVT clades 1 to 5.
- Additional file 4-10 Ancient mtDNA recombination in CTVT clade 1.
- Additional file 4-11 CTVT tumour and host mtDNA haplotype lists.
- Additional file 4-12 Maximum likelihood phylogenetic tree of CTVT mtDNA.
- Additional file 4-13 Geographical locations and mtDNA clades for CTVT tumours and hosts.
- Additional file 4-14 Summary of back mutations.
- Additional file 4-15 Substitutions with corresponding VAF (before normalisation) for each of 449 CTVT tumours and 338 CTVT hosts.
- Additional file 4-16 Indels with corresponding VAF (before normalisation) for each of 438 CTVT tumours and 334 CTVT hosts.

CHAPTER 5

- Additional file 5-1 Reconstructed donor haplotypes for CTVT mtDNA clades 5, 8-18.

PUBLICATIONS ASSOCIATED WITH THIS THESIS

- Additional file 6-1 Publications associated with this Thesis.

LIST OF ABBREVIATIONS AND SYMBOLS

Abbreviation	Meaning
BAM	Binary sequence alignment and mapping
bp	Base pairs
BWA	Burrows-Wheeler Alignment
cm	Centimetre
cm³	Cubic centimetre
CTVT	Canine transmissible venereal tumour
DEFRA	Department for Environment, Food & Rural Affairs
°C	Degrees Celsius
D-loop	Displacement-loop
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
EDTA	Ethylene diamine-tetraacetic acid
GLMER	Generalized linear mixed-effects model
indel	Small insertion or deletion
IQR	Interquartile range
kb	Kilobase (1000 base pairs)
LOH	Loss of heterozygosity
ml	Millilitre
mm³	Cubic millimetre
MT	mitochondria
mtDNA	Mitochondrial DNA
n	Sample size
NCBI	National Center for Biotechnology Information
ng/μl	Nanogram/microliter
NuMT	Nuclear mitochondrial DNA fragment
p	p-value
PCR	Polymerase chain reaction
qPCR	Quantitative polymerase chain reaction
ROS	Reactive oxygen species
rpm	Revolutions per minute
s (time)	second
SNP	Single nucleotide polymorphism
μg	Microgram
UK	United Kingdom
μl	Microliter
μm	Micrometre
VAF	Variant allele frequency
%	Percentage

PREFACE

The canine transmissible venereal tumour (CTVT) is a contagious cancer that is naturally transmitted between dogs by the allogeneic transfer of living cancer cells during coitus. CTVT first arose several thousand years ago and has since then spread through dog populations worldwide. My main interests in this Thesis were (1) to gain further understanding of CTVT distribution patterns and prevalence around the world, (2) to use genetics to trace the historical spread of CTVT and (3) to map the genetic as well as phenotypic diversity of CTVT tumours around the world.

The research was performed in the Transmissible Cancer Group at the Department of Veterinary Medicine, under the direction of Dr Elizabeth Murchison and with the support of a large network of veterinary collaborators who contributed essential samples and information used in this study (see Table of Collaborators, page v). All DNA sequencing was performed at the Wellcome Trust Sanger Institute, Hinxton. DNA sequencing data were stored at the Wellcome Trust Sanger Institute, Hinxton.

This dissertation summarises my work performed to understand the genetic and phenotypic diversity of CTVT around the world and to answer the three main goals listed above. The following paragraphs describe work undertaken in each Chapter, together with listing the publications resulting from each study.

Chapter 1 presents a review of canine transmissible venereal tumour (CTVT), the oldest known cancer lineage in nature. Parts of this Chapter have been published as an article in Current Opinion in Genetics and Development. A discussion of canine genetics together with the role of mitochondria in cancer and in CTVT follows. Considering the knowledge of CTVT genetic and phenotypic diversity at the time this PhD work was started, Chapter 1 summarises the goals of this PhD thesis. **Publication: Strakova A and Murchison EP. The cancer which survived: Insights from the genome of an 11,000 year-old cancer. Curr. Opin. Genet. Dev. 30, 49-55 (2015). See Additional file 6-1 for full article text.**

Chapter 2 introduces analysis of worldwide CTVT distribution patterns and includes CTVT prevalence information obtained from 645 veterinarians located in 109 countries. Results from this Chapter were published as an article in BMC Veterinary Research. **Publication: Strakova A, Murchison EP. The changing global distribution and prevalence of canine transmissible venereal tumour. BMC Vet Res. 10, 168 (2014). See Additional file 6-1 for full article text.**

Chapter 3 describes results from CTVT sample collection process, together with design, validation and implementation of a CTVT diagnostic pipeline. Apart from being used to validate tumour samples analysed in this Thesis, the final diagnostic pipeline was also implemented by Karina Ferreira de Castro, a visiting student in the Transmissible Cancer Group, which led to publication of an article in *Veterinary and Comparative Oncology*. **Publication: Castro FK, Strakova A, Tinucci Costa M, Murchison EP. Evaluation of a genetic assay for canine transmissible venereal tumour diagnosis in Brazil. *Vet Comp Oncol*. 15: 615–618 (2016). See Additional file 6-1 for full article text.**

Chapter 4 presents analysis of 449 sequenced complete mitochondrial genomes from a worldwide population of CTVT tumours, with the goal of learning about the genetic diversity of this disease. Investigation into histopathological phenotypic differences between CTVT tumours from different mitochondrial clades is also described. A significant proportion of results presented in this Chapter was published as an article in *eLife*. **Publication: Strakova A, Ni Leathlobhair M, Wang G-D, Yin T-T, Airikkala-Otter I, Allen JL, Allum KM, Bansse-Issa L, Bisson JL, Castillo Domracheva A, de Castro KF, Corrigan AM, Cran HR, Crawford JT, Cutter SM, Delgadillo Keenan L, Donelan EM, Faramade IA, Flores Reynoso E, Fotopoulou E, Fruean SN, Gallardo-Arrieta F, Glebova O, Häfelin Manrique RF, Henriques JJGP, Ignatenko N, Koenig D, Lanza-Perea M, Lobetti R, Lopez Quintana AM, Losfelt T, Marino G, Martincorena I, Martínez Castañeda S, Martínez-López MF, Meyer M, Nakanwagi B, De Nardi AB, Neunzig W, Nixon SJ, Onsare MM, Ortega-Pacheco A, Peleteiro MC, Pye RJ, Reece JF, Rojas Gutierrez J, Sadia H, Schmeling SK, Shamanova O, Ssuna RK, Steenland-Smit AE, Svitich A, Thoya Ngoka I, Vițălaru BA, de Vos AP, de Vos JP, Walkinton O, Wedge DC, Wehrle-Martinez AS, van der Wel MG, Widdowson SAE and Murchison EP. Mitochondrial genetic diversity, selection and recombination in a canine transmissible cancer. *eLife*. (2016); DOI: 10.7554/eLife.14552. See Additional file 6-1 for full article text.**

Chapter 5 discusses use of exome sequencing data to perform analyses to understand the relationships between five CTVT mitochondrial clades presented in Chapter 4. The initial findings are complemented by analysis of a single canine mitochondrial DNA haplogroup that has recurrently and recently undergone multiple horizontal transfer events. Work presented in Chapter 5 is currently being prepared for publication, and thus far has received input from Adrian Baez-Ortega, Máire Ní Leathlobhair, Katherine Hughes, Hannah Bender, Alexander Sampson, Isobelle Bolton, Tracy Wang, Kevin Gori, veterinary collaborators who collected samples and Elizabeth Murchison.

Finally, Chapter 6 summarises the results obtained from studies on worldwide CTVT genetic and phenotypic diversity described in this PhD thesis and draws conclusions on wider implications of this work, together with highlighting areas of future interest.

Declaration of originality

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 Table of Collaborators (see page v) 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 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 specified in the text. The dissertation does not exceed the prescribed word limit of 60 000 words.

1

Canine transmissible venereal tumour as the cancer which survived

Summary

Cancer is a disease of the genome, whereby accumulation of genetic mutations leads to abnormal proliferation of the body's own cells. Although cancer cells are often considered "immortal" and have the potential to survive indefinitely, the existence of a neoplasm is normally limited by the lifespan of its host. Only a few cancers have evolved to survive beyond the lifespan of their original host, and one of them affects the canine population. The canine transmissible venereal tumour (CTVT) is a transmissible cancer that is spread between dogs by the allogeneic transfer of living cancer cells during coitus. CTVT affects dogs around the world and is the oldest and most divergent cancer lineage known in nature. CTVT first emerged as a cancer about 11,000 years ago from the somatic cells of an individual dog, and has subsequently acquired adaptations for cell transmission between hosts and for survival as an allogeneic graft. In this Chapter, I review the clinical, morphological, molecular, immunological and genetic characteristics of CTVT. Although CTVT has propagated through the canine population as a single clone for all of its history, one of the most interesting features of this disease is its diversity around the world and ability to survive over millennia. One of the adaptations underlying the success of this remarkable lineage may be its ability to periodically acquire mitochondrial DNA through horizontal transfer events. In addition to discussing the current knowledge of CTVT, this Chapter summarises the overall goals of this PhD thesis.

1.1 Cancer evolution

1.1.1 Cancer as a disease

Cancer as a disease has been around for thousands of years, and scientific understanding of its biology has changed and evolved over time. Many hypotheses explaining the causes of cancer have been proposed, rejected and subsequently adopted again (Mukherjee 2010, Morange 2011). It wasn't until the beginning of the twentieth century when the first link between abnormality of genetic material and uncontrollable growth in cancer was made by Theodor Boveri, who in 1914 postulated that tumours originated by incorrect chromosome segregation which was passed to daughter cells (Boveri 1914, Balmain 2001).

After the discovery of the deoxyribonucleic acid (DNA), as the molecule of inheritance (Avery et al. 1944, Watson and Crick 1953), new findings have further supported the theory that genetic pathology was responsible for development of cancer. Early observations from cytogenetic studies have shown that particular chromosomal rearrangements are associated with specific cancer types. A pivotal discovery was made in patients with chronic myelogenous leukemia, who shared a translocation between chromosomes 9 and 22, that became known as the 'Philadelphia chromosome' (Rowley 1973).

Further milestone was the finding that a single gene of Rous sarcoma virus was required to transform infected chicken cells into neoplastic cells (Parker et al. 1984) – this study eventually led to the discovery of *proto-oncogenes*, which are normal genes responsible for cellular growth control and cell cycle progression, that when mutated lead to transformation into a neoplastic cell. Studies of heritable and non-heritable retinoblastomas revealed that at least two mutational events were necessary for the development of this cancer (Knudson 1971), with subsequent studies revealing importance of the RB1 gene in inhibition of cell cycle progression (Murphree and Benedict 1984). Genes directly inhibiting oncogenic development were later termed *tumour suppressors*. The causes of genetic alteration in oncogenic transformation were at this stage gradually converging into a consolidated understanding seeing that mutation of specific genes was key to oncogenesis.

Causes of genetic alteration in oncogenic transformation in humans have since been linked to oncogenic viruses (e.g. Epstein-Barr virus) (Moore and Chang 2010), bacteria (e.g. *Helicobacter Pylori*) (Parsonnet 1995) or helminths (Vennervald and Polman 2009). Further described causes of genetic change include environmental exposure, such as chemicals or UV-light (Belpomme et al. 2007), tobacco (Hecht 2003) and others.

As understood now, the term “cancer” encompasses many different types of malignancies, characterised by abnormal cell growth. Years of cancer research have led to a proposal that all tumours share a set of functional abilities enabling the survival, proliferation and dissemination of cancer cells, termed “hallmarks of cancer” (Hanahan and Weinberg 2000, Hanahan and Weinberg 2011, Fouad and Aanei 2017).

Perhaps the most important feature of cancer cells is their ability to sustain proliferative signalling, thus dividing uncontrollably within the tissue, followed by evasion of growth suppressor signals (Hanahan and Weinberg 2000). Under normal circumstances, these uncontrollably dividing cells would undergo apoptosis – which is yet another feature evaded by cancer cells (Lowe and Lin 2000). Telomeres, that protect the ends of chromosomes and normally limit life span of a cell, have been implicated to play a role in cancer cell immortality, thus enabling them to divide indefinitely (Mathon and Lloyd 2001). Tumours, like normal tissue, require access to oxygen and nutrients, acquired in the case of cancer through induction of angiogenesis (formation of new vessels) (Folkman 2002). Ability to invade surrounding tissues is a predisposing feature for metastasis of cancer cells to different parts of the body (Hanahan and Weinberg 2000, Nguyen 2004). More recently, the importance of reprogramming energy metabolism and evasion of immune destruction by the various arms of the immune system have been implicated as additional features of cancer cells (Warburg 1956a, Warburg 1956b, Hanahan and Weinberg 2011).

Cancer is also commonly found in the animal kingdom and it is an important disease in veterinary medicine. Examples include cancers associated with oncogenic viruses in cats, poultry and cattle, tumours associated with environmental hazards in cattle and dogs, or genetic predisposition of certain dog breeds (Misdorp 1996, Bronden et al. 2007). Man’s best friend, the dog, is known to be affected by many types of cancers, and it is well recognised that different dog breeds are at variable risks of developing certain cancers (Dobson 2013).

1.1.2 Clonal evolution of cancer

The acquisition of the above-mentioned abilities of cancer is dependent on accumulation of genetic mutations that lead to abnormal proliferation of body's own cells – the pathway to malignancy is, however, very variable between different cancers. The process of cancer development occurs through an evolutionary selection of genetic changes, which offer the cell a fitness advantage, leading to the progressive transformation from a normal cell into a cancer cell (Foulds 1954, Nowell 1976) (Figure 1-1).

As an evolutionary process, cancer progression is dependent on the existence of genetic variation. This genetic variation is generated by substitution mutations (single base changes), small or large deletions or insertions, rearrangements, translocations, amplifications or changes in number of copies of DNA segments (Stratton 2011). The effects of acquired genetic instability, followed by sequential selection of genetic subpopulations, form the basis for biological processes recognised to take place during evolution of a tumour (Nowell 1976).

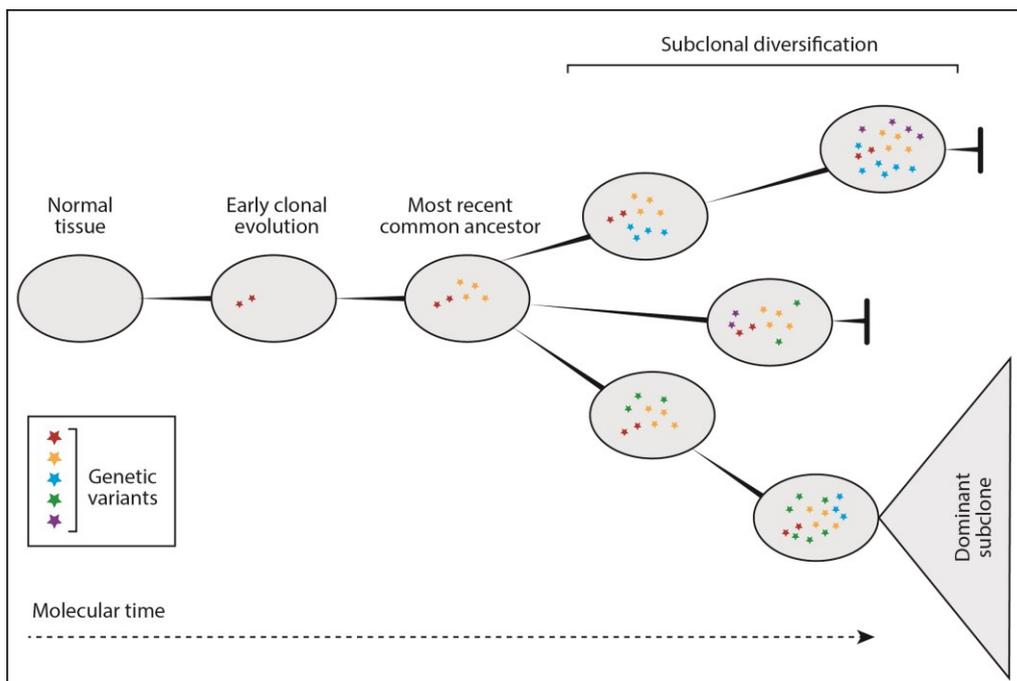


Figure 1-1 Model of clonal evolution of cancer. Subclonal diversification is followed by sequential selection of dominant genetic subpopulations. Figure adapted from (Nik-Zainal et al. 2012b).

The majority of genetic changes can be assigned into two categories, based on their effect on the cell: (1) gain-of-function mutations in proto-oncogenes (e.g. *RAS*, *MYC*), which lead to increased cell growth, proliferation and survival, (2) loss-of-function mutations in tumour suppressor genes (e.g. *TP53*), leading to loss of proliferation and cell survival restraints (Lee and Muller 2010, Hanahan and Weinberg 2011).

In the last two decades, scientific understanding of genetic variation in cancer has rapidly expanded, mainly through the impact of next-generation massively parallel DNA sequencing technology (Mardis and Wilson 2009, Wheeler and Wang 2013, Goodwin et al. 2016). Next-generation sequencing technologies incorporated revolutionary innovations that enabled scientists to sequence genomes of cancers faster and cheaper, thus obtaining a more comprehensive view of genomic information. This fast improving next-generation sequencing technology has illuminated the mutational landscapes of cancers and provided new insights into cancer genome evolution.

1.1.3 Germline and somatic variation

Germline (hereditary) genetic variants are the variants that are inherited by individuals and are fixed in all cells of an individual, but vary between individuals in a population (see Figure 1-2). Germline variation is found in all populations, and results in diverse appearance between individuals. Somatic variants, on the other hand, arise as mutations in somatic cells (i.e. not gametes) of the body. Such mutations are private to the cells of the individual in which they arose, and are not passed on to their offspring.

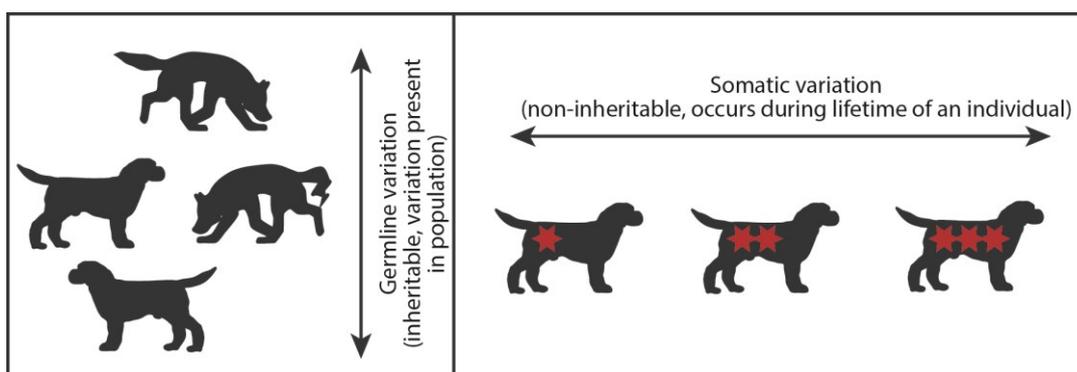


Figure 1-2 Germline and somatic variation. Germline variation is normal variation present in the population. Somatic variation is genetic variation arising in somatic cells during the lifetime of an individual.

1.1.4 Evolutionary selection in cancer

Most somatic mutations have little effect on the cell. However, certain somatic mutations in proto-oncogenes or tumour suppressor genes provide a replicative advantage to the cell, leading to development of a tumour. Mutational processes in cancer cells lead to the occurrence of somatic variants, that may provide an additional selective advantage to the cell, and its progeny may become the dominant subpopulation within the tumour (Nowell 1976) (see Figure 1-1).

The somatic variants which offer a growth advantage at the cellular scale, may be positively selected and act as the 'driver' mutations in the initiation of cancer (Ostrow et al. 2014). On the other hand, if a somatic variant has no effect on the fitness of the cell, it may just be propagated to the next generation alongside driver mutations as a 'passenger' mutation.

The degree to which somatic variants are selected in cancer can be quantified by examining the patterns of substitution. The most commonly used measure is the ratio of the rates of non-synonymous (causing an amino-acid change) to synonymous (not causing an amino-acid change) variants, also referred to as dN/dS (or Ka/Ks) (Nei and Gojobori 1986, Fay and Wu 2003). In the absence of selection, the rate of accumulation of synonymous and non-synonymous variants will be equal ($dN/dS \sim 1$). Negative selection ($dN/dS < 1$) will remove the non-synonymous variants from the pool of somatic variation, while positive selection ($dN/dS > 1$) increases their frequency (Ostrow et al. 2014).

The extent to which somatic mutations provide evolutionary advantage to the cancerous growth is, however, normally limited by the lifespan of the host. In this sense, transmissible cancers have gained the ultimate advantage, as they are able to propagate between hosts and continue to evolve as clonal cell lineages.

1.2 Canine transmissible venereal tumour

1.2.1 Transmissible cancers in nature

Transmissible cancers are a rare phenomenon in nature. Most cancers remain within the body they originated from. Transmissible cancers, however, survive beyond the deaths of their original hosts, through transmission of living cancer cells.

Causes of cancer were largely unknown before the twentieth century (Balmain 2001). In the nineteenth century, the discovery of a canine cancer that could be passed between dogs as an infectious disease (Nowinsky 1876) created controversy both in the scientific field and amongst the public, leading to fear that all cancers may be contagious. Further work, however, revealed that the canine cancer reported by Nowinsky was likely to be an exception, and that most cancers were not directly transmissible. The cancer studied by Nowinsky was likely to have been the canine transmissible venereal tumour (CTVT).

Apart from CTVT, there are only seven other known examples of transmissible cancers in nature – two lineages affecting Tasmanian devils (called DFT1 and DFT2) (Pearse and Swift 2006, Pye et al. 2016) and five distinct lineages of contagious disseminated neoplasia affecting various species of marine bivalves (Metzger et al. 2015, Metzger et al. 2016). Furthermore, a transmissible cancer in a laboratory population of hamsters was transmissible by cannibalism and mosquitoes (Brindley and Banfield 1961, Cooper et al. 1964, Banfield et al. 1965). Cancer cells have also rarely been reported to have spread between two humans within a variety of contexts, including surgical accident, organ transplant, in utero and during experimental treatments (Moore et al. 1957, Gartner et al. 1996, Kauffman et al. 2002, Tolar and Neglia 2003, Sala-Torra et al. 2006, Strauss and Thomas 2010). However, no human cancer that spread serially between more than two individuals has been reported.

Each contagious cancer arose from the cells of a single individual, before being passaged through the population as a clonal lineage. The significance of transmissible cancers lies in the fact that they have evolved into tumours that are autonomous from their original hosts, and became effectively neoplastic cells that behave as parasites.

1.2.2 Clinical features of CTVT

1.2.2.1 *Gross clinical features*

Canine transmissible venereal tumour (CTVT) is a transmissible cancer which usually manifests clinically as tumours associated with external genitalia of both male and female dogs (see Figure 1-3). A disease consistent with CTVT was first described in the veterinary literature at least 200 years ago, as an “ulcerous state” of genital regions “accompanied by fungous excrescence” (Blaine 1810). It has also been historically referred to as Sticker’s sarcoma or transmissible venereal sarcoma.

The early signs of CTVT include discharge from the vulva or prepuce, and licking of affected genital regions (Cohen 1985). In female dogs, the tumour can affect any part of vagina and it may also protrude out of the vulva; in later stages, it often appears as cauliflower-like, pedunculated and multilobular mass (Ajello 1980, Cohen 1985). In male dogs, the tumour is commonly located on the bulbus glandis (base of penis), but can also affect other parts of the penis or the prepuce (Karlson and Mann 1952, Cohen 1985). Different presentations of CTVT tumours are displayed in Figure 1-3. Extra-genital localisations of CTVT have also been described, including tumours affecting the eyes, nasal areas, mouth and skin (Belkin 1959, Ndiritu et al. 1977, Weir et al. 1978, Perez et al. 1994, Guedes et al. 1996, Albanese et al. 2002, Gurel 2002, Chikweto 2013) (see Figure 1-4).

CTVT metastasis has been previously noted in individual case reports (Feldman 1929, Cella 1939, Rust 1949, Belkin 1959, Barron et al. 1963, Prier and Johnson 1964, Sastry et al. 1965, Higgins 1966, Adams and Slaughter 1970, Manning and Martin 1970, Rottcher 1972, Kimeto and Mugeru 1974, Osipov and Golubeva 1976, Idowu 1977, Ajello 1980, Yang 1987, Miller 1990, Moulton 1990, Gandotra 1993, Ferreira et al. 2000, Pereira et al. 2000, Gurel 2002, Abuom 2006, Levy et al. 2006, Park et al. 2006, Batista et al. 2007, Bastan 2008, Mylonakis et al. 2008, Varughese et al. 2012, Chikweto 2013). Cases of metastasis are rare, and they have been more frequently described in immunosuppressed dogs and in puppies (Karlson and Mann 1952, Cohen 1973, Brown et al. 1980, Cohen 1985).

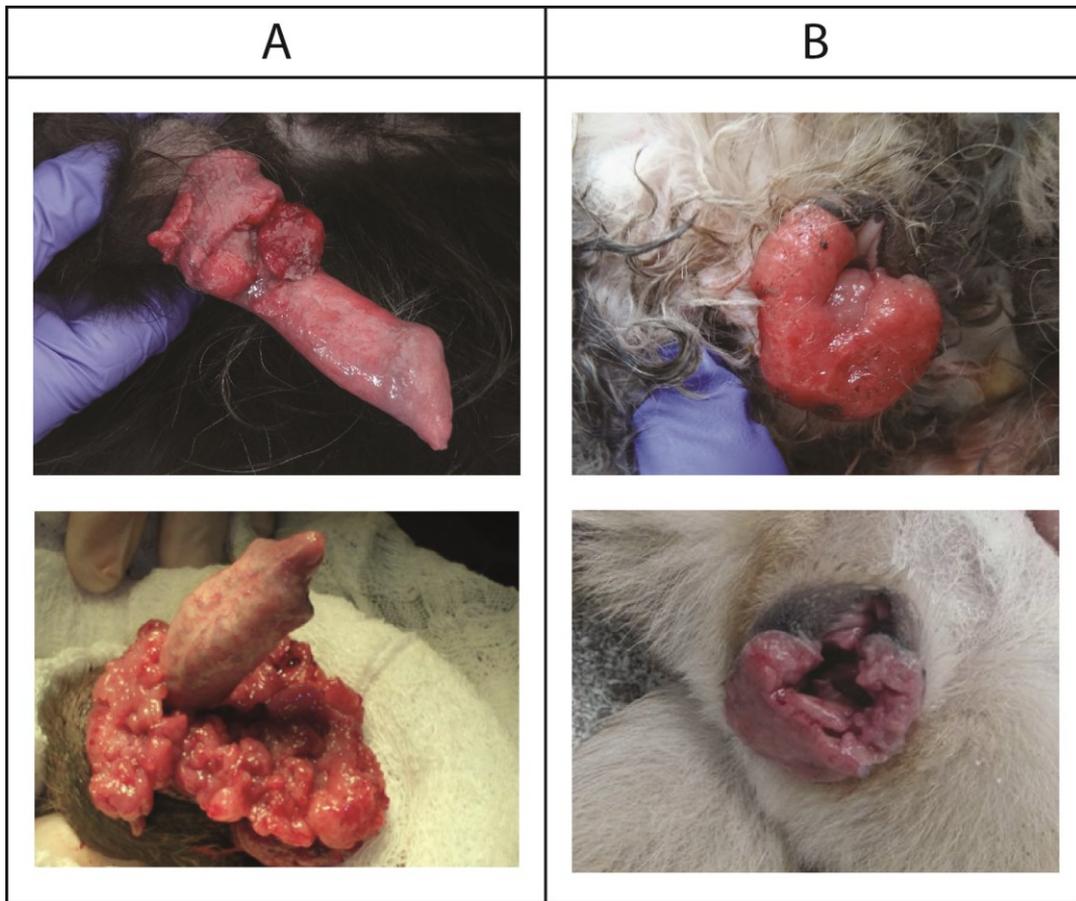


Figure 1-3 Clinical presentations of canine transmissible venereal tumour (CTVT) in (A) male dogs, (B) female dogs. Photos taken by the author (top row) and courtesy of Karina Ferreira de Castro and Mirjam van der Wel (bottom row).

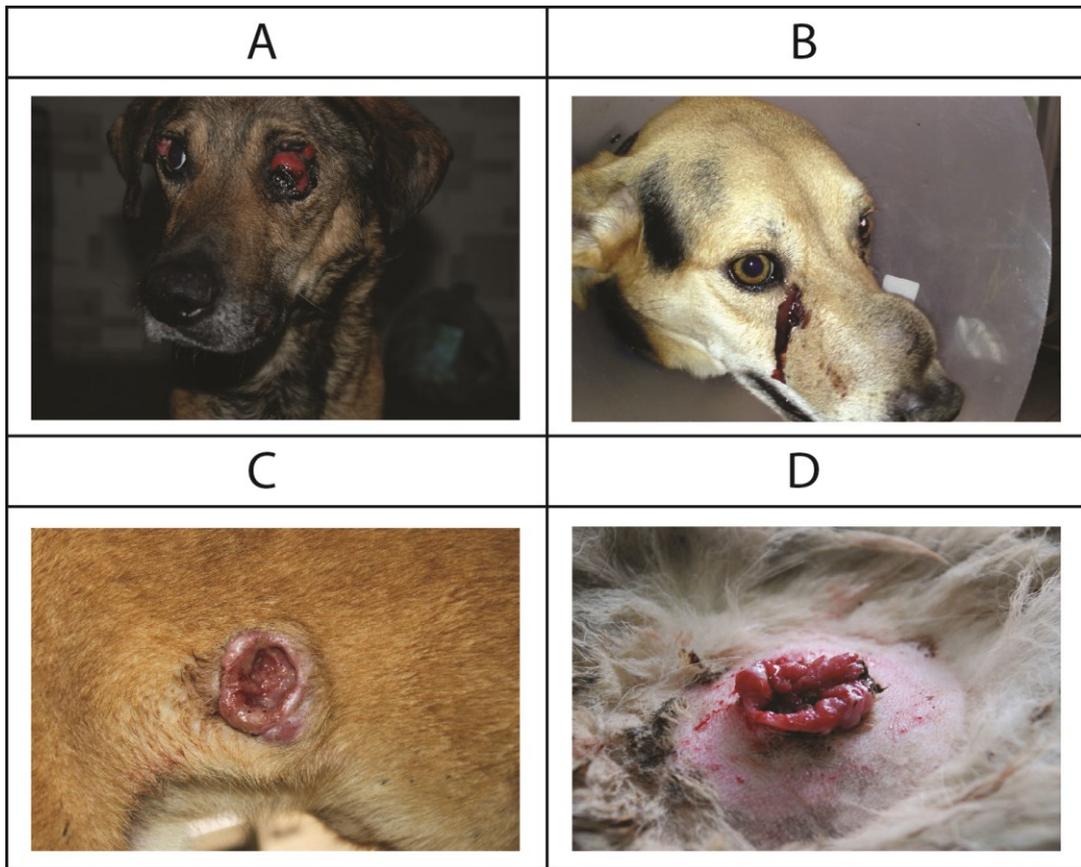


Figure 1-4 Extra genital presentations of canine transmissible venereal tumour on (A) eye (photo courtesy of Natalia Ignatenko), (B) nose (photo courtesy of Karina Ferreira de Castro), (C) skin (photo courtesy of Michael Meyer) and (D) skin (photo courtesy of Natalia Ignatenko).

1.2.2.2 *Histopathological features*

All cancers exist in a microenvironment which is formed by the surrounding stromal tissue and infiltrating host cells (Quail and Joyce 2013). The CTVT microenvironment is characterised by a delicate fibrovascular stroma with variable amounts of host immune cell infiltrate (see Figure 1-5). CTVT cells are usually closely packed and arranged in diffuse masses, clusters, or cords along fibrous connective tissue trabeculae that contain blood vessels (see Figure 1-5) (Cohen 1985).

In haematoxylin-eosin stained histology sections, CTVT cells appear as round or polyhedral with variably distinct cell boundaries, eosinophilic and granular cytoplasm, with a large centrally positioned round-oval nucleus containing coarsely aggregated chromatin and usually a prominent nucleolus (Cohen 1985) (Figure 1-5). The diameter of CTVT cells ranges from 15 to 30µm (Duncan and Prasse 1979). Several mitotic figures can be observed in most high-power fields (Cohen 1985).

Despite extensive studies into histopathological features of CTVT, the cell of origin is debated (Sandusky et al. 1987, Gimeno et al. 1995, Mozos et al. 1996, Marchal et al. 1997, Mukaratirwa et al. 2004). The presence of vimentin, lysozyme and alpha-1-antitrypsin in CTVT tissues indicate that CTVT is most likely derived from the macrophage (histiocytic) lineage (it should, however, be noted that this panel of markers is not exclusive to macrophages; both vimentin and lysozyme are expressed by other mesenchymal cells) (Sandusky et al. 1987, Gimeno et al. 1995, Mozos et al. 1996, Marchal et al. 1997, Mukaratirwa et al. 2004). Reports of CTVT cells being parasitized by *Leishmania infantum* further support the possibility of a macrophage origin (Albanese et al. 2002, Catone et al. 2003).

1.2.2.3 *Cytological features*

Cytologically, CTVT cells are very distinct and therefore, coupled with clinical history, cytology is often used as a diagnostic method (Jackson 1944, Duncan and Prasse 1979). Perhaps the most striking cytological feature of CTVT cells is the presence of clear vacuoles in the cytoplasm of cells (Jackson 1944, Duncan and Prasse 1979, Ganguly et al. 2013) (see Figure 1-6). CTVT cells for cytology analysis may be collected by fine-needle aspirates or impression smears.

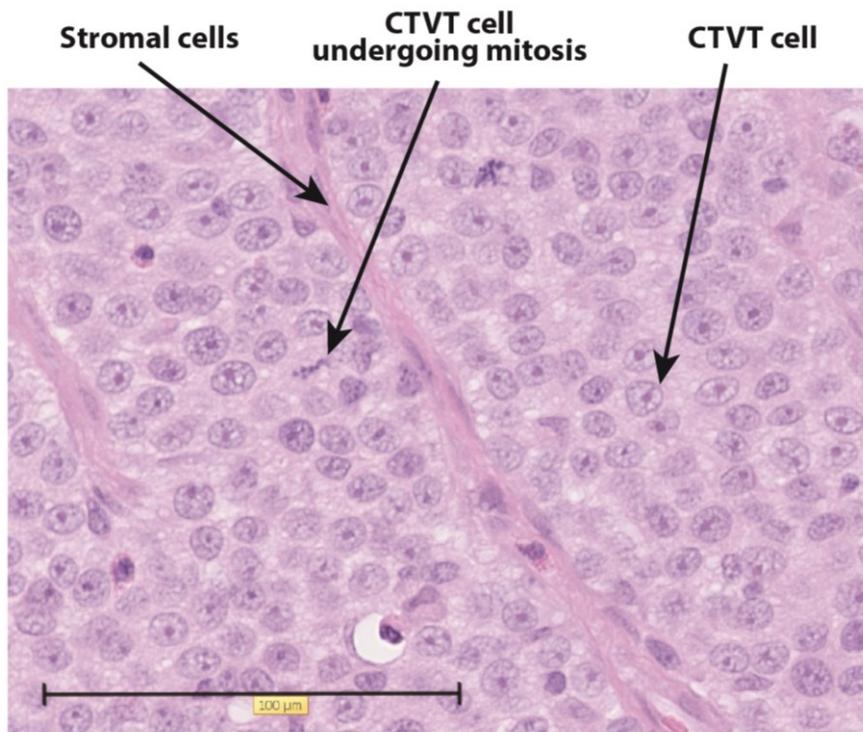


Figure 1-5 Histology image of canine transmissible venereal tumour cells stained with haematoxylin-eosin stain. Magnification 40X.

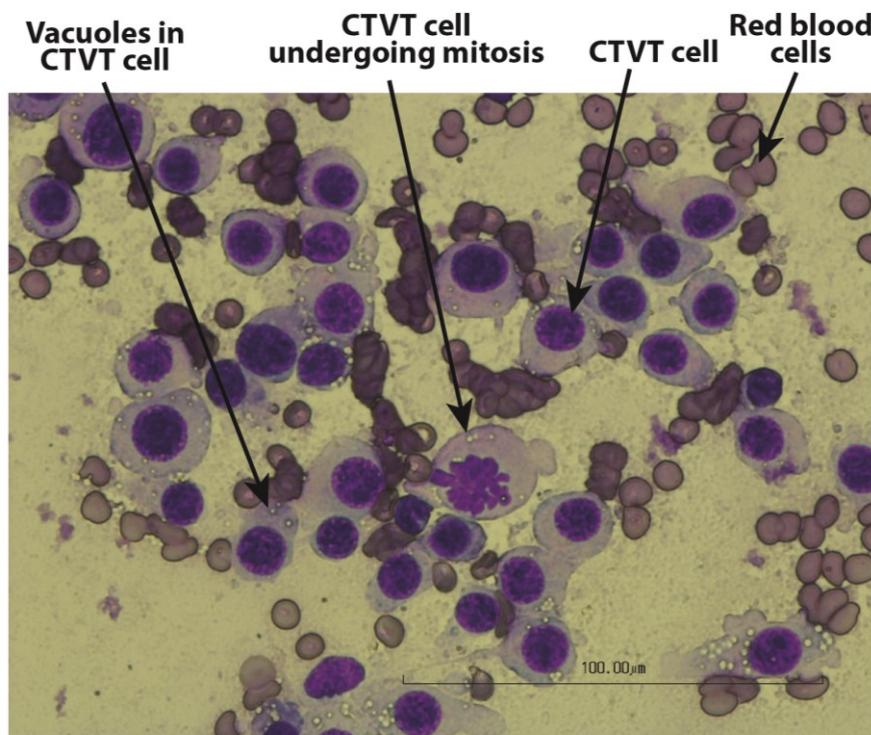


Figure 1-6 Cytology image of canine transmissible venereal tumour cells stained with Diff Quick. Magnification 60X.

1.2.3 Diagnosis and treatment of CTVT

1.2.3.1 *Diagnosis*

Although CTVT has characteristic pathological features, which together with clinical history and physical location of the tumour often provide a reliable diagnosis, many cases are more difficult to diagnose through clinical diagnostic methods only. These cases include extra-genital tumours, which may look similar to cutaneous histiocytoma, tumours on the eye/nose/muzzle or tumours which have an atypical presentation, such as generalised metastases.

To reliably diagnose CTVT cases, the following methods are routinely used in clinical practice or in the laboratory:

(1) Cytology

As described in section 1.2.2.3, CTVT cells have a very distinct appearance on cytology slides (Duncan and Prasse 1979). This method is most commonly used in clinical practices. It is non-invasive, and results can be obtained within a short period of time.

(2) Histopathology

Similarly to above, CTVT cells have a characteristic appearance on haematoxylin-eosin stained sections (see section 1.2.2.2) (Cohen 1985). A biopsy in formalin is required, followed by histopathological processing. If diagnosis is unclear, immunohistochemistry staining can be performed to rule out other round cell neoplasms.

(3) *LINE-MYC* rearrangement

A diagnostic evidence for CTVT is the presence of LINE element insertion near the *MYC* locus in the CTVT nuclear genome (Katzir et al. 1985) (for further information see section 1.2.7.3.1). Molecular biology techniques, such as polymerase chain reaction (PCR), can be used to confirm presence of this marker in tumour samples. The diagnostic test can only be performed on genomic DNA and therefore is not routinely performed in the clinic, although it is currently used in research.

1.2.3.2 *Treatment*

A number of therapeutic approaches are used in the clinic to treat CTVT, as the evidence that naturally occurring CTVT tumours would regress spontaneously is very limited (Smith and Washbourn 1898b, Brown et al. 1980, Cohen 1985, Amber et al. 1990). Therapeutic options to treat CTVT are discussed below:

(1) Chemotherapy

Vincristine is known to be a very effective treatment for CTVT with the majority of cases being responsive (Brown et al. 1980, Calvert et al. 1982, Boscios 1988, Singh et al. 1996, Nak et al. 2005, Said et al. 2009). Other chemotherapeutic drugs used for treatment of CTVT include vinblastine and doxorubicine (Said et al. 2009, Ramadinha et al. 2016).

(2) Surgical removal

Surgical removal of tumours with clearly defined margins is a therapeutic option, however, common tumour recurrence after the surgery has been reported (Karlson and Mann 1952, Idowu 1984, Dass and Sahay 1989).

(3) Radiotherapy

Several reports have shown that CTVT is sensitive to radiotherapy (Osipov and Golubeva 1976, Ajello 1980), however, this method is often not available in clinics where CTVT cases are treated.

1.2.4 Mode of CTVT transmission

Under natural settings, CTVT is usually passed on between individual dogs by the transmission of living cancer cells during coitus. Cancers frequently acquire features that cause cells to depart from a primary tumour and establish new tumours in distant sites of the body via a process of metastasis. CTVT, however, has acquired adaptations for the transmission of cancer cells to new hosts. The family Canidae may have been particularly at risk for the establishment of a sexually-transmitted cancer due to the existence of the long-lasting coital tie that is specific to this group. The coital tie may last for up to 30 minutes, and may lead to injuries to the genital mucosa; such conditions may thus provide an exceptional opportunity for the exchange of cancer cells between individuals (Cohen 1985). Under some conditions, transmission of CTVT appears to be highly efficient – a previous report noted

transmission of CTVT from one infected male dog to eleven out of twelve bitches with which it mated (Smith and Washbourn 1898b).

Reflecting the fact that CTVT is usually transmitted during coitus, the disease occurs most frequently in adult sexually active animals (Karlson and Mann 1952, Brown et al. 1980, Cohen 1985). Although both sexes are susceptible, some reports claimed that that CTVT is more common in males (Osipov and Golubeva 1976, Brown et al. 1980), while others have found the disease to be more common in females (Ajello 1939, Ajello 1980, Gandotra 1993, Singh et al. 1996, Sobral 1998).

CTVT tumours are also occasionally found affecting non-genital regions, most commonly skin, nasal cavity, lymph node, eye and mouth (Belkin 1959, Ndiritu et al. 1977, Weir et al. 1978, Perez et al. 1994, Guedes et al. 1996, Albanese et al. 2002, Gurel 2002, Chikweto 2013). As these sometimes occur without genital involvement (Feldman 1929, Perez et al. 1994, Ginel et al. 1995, Ganguly et al. 2013), this suggests that there may be non-coital routes of CTVT transmission, possibly linked to breaching of the mucosa and involving licking, sniffing or parturition (Cohen 1985).

Direct transmission of cancer cells may select for loss of cell adhesion; indeed, CTVT tumours are typically highly friable (Brown et al. 1980, Thacher and Bradley 1983, Mello Martins et al. 2005). Furthermore, CTVT tumours are delicately encapsulated and bleed readily upon contact (Mello Martins et al. 2005), presumably optimised for the release of CTVT cells during the friction involved in coitus.

CTVT has been reported to affect many dog breeds, with no apparent breed predisposition for CTVT infection (Karlson and Mann 1952, Brown et al. 1980, Cohen 1985). Although naturally occurring CTVT tumours have only been described in dogs, under experimental settings, CTVT can be transplanted into wild canids, including wolves, coyotes and red foxes (Sticker 1906a, Wade 1908, Dungern 1912, Cockrill and Beasley 1979, Cohen 1985). CTVT can also be induced by subcutaneous inoculation with living tumour cells in immunocompetent, allogeneic dogs.

1.2.5 CTVT growth phases, experimental transmission studies and immune system interaction

1.2.5.1 *Experimental transmission studies*

The patterns of growth progression of CTVT and the importance of the immune system in controlling the disease have been demonstrated by a number of experimental studies. Early observations of CTVT revealed that experimentally transplanted tumours in dogs exhibit progressive growth (progressive phase), frequently followed by immune-mediated spontaneous regression of the tumour two to six months after transplantation (regressive phase) (Wade 1908, Karlson and Mann 1952, Epstein and Bennett 1974). Immunological studies have shown that the passive transfer of sera from post-regression dogs prevented development of tumours transplanted into naive dogs (Powers 1968). Moreover, puppies born to bitches in which the tumour had regressed were able to limit the growth of the injected tumour (Yang and Jones 1973), indicating a transfer of immunity.

In contrast, however, spontaneous regression has not been consistently reported in naturally occurring CTVT (Smith and Washbourn 1898b, Brown et al. 1980, Cohen 1985, Amber et al. 1990). The immune response to CTVT may be influenced by the site of tumour transplantation (experimentally transplanted CTVT tumours are usually injected subcutaneously) and the concurrent presence of injuries and inflammation. It is also possible that there is variation in susceptibility to CTVT within the dog population that influences clinical progression and disease course (Taylor Bennett et al. 1975).

Despite the fact that CTVT cells can be experimentally transplanted to wolves, coyotes and red foxes (Sticker 1906a, Wade 1908, Dungern 1912, Cockrill and Beasley 1979, Cohen 1985), attempts to transmit the tumour to the following immunocompetent animals have not been successful: mice, rats, cats, rabbits, hamsters, chicken, monkeys and mules (Smith and Washbourn 1898a, Sticker 1905, Ajello 1939, Cockrill and Beasley 1979). The tumour has, however, been experimentally transmitted as a xenograft to nude mice (Holmes 1981), X-irradiated mice (Stubbs and Furth 1934) and NOD/SCID mice (Harmelin et al. 2001).

1.2.5.2 *Adaptations for immune system escape*

Although all cancers, including those that remain within a single host, may have acquired adaptations to escape immune destruction, CTVT, as a transmissible cancer, is able to escape the immune system as an allogeneic graft. The highly potent immune response to allogeneic

grafts is primarily mediated by direct allorecognition of foreign major histocompatibility complex (MHC) molecules by the graft recipient's T cells (Rosenberg and Singer 1992). Although MHC molecules are normally expressed by all nucleated cells, CTVT appears to have lost expression of MHC molecules, presumably via a process of immunoselection (Cohen et al. 1984, Yang et al. 1987, Hsiao et al. 2002, Murgia et al. 2006). Similarly, many human cancers modulate MHC molecule expression as a mechanism to escape immune detection (Algarra et al. 1997).

The mammalian immune system has specific mechanisms to detect cells which are not expressing MHC molecules. Natural killer (NK) cells are specialised lymphocytes which become cytotoxic when activated by 'missing self', that is, absent MHC. The mechanisms whereby transmissible cancers escape NK cell killing remain unclear (Brown et al. 2011).

The recruitment of an immunosuppressive tumour microenvironment leading to immune tolerance or anergy may be an important feature in transmissible cancer immune escape (Hsiao et al. 2002, Hsiao et al. 2004, Hsiao et al. 2008, Liu et al. 2008). This suggestion is supported by the observation that CTVT rarely metastasises, thus departing from the established tumour microenvironment, except in immunosuppressed hosts and newborn puppies (Cohen 1973, Yang and Jones 1973).

1.2.6 Worldwide distribution of CTVT

CTVT has been reported in many countries around the world and its distribution is widespread (reviewed in (Stookey 1969, Ndiritu et al. 1977, Das and Das 2000, Eze et al. 2007, Purohit 2009, Stimmelmayer 2010, Ganguly et al. 2013)). The disease has been reported more frequently in tropical and subtropical regions (Rust 1949, Higgins 1966). In the United States, CTVT prevalence has been found to be inversely correlated with latitude and positively correlated with higher mean annual temperature and increased rainfall (Hayes et al. 1983).

Although there have been numerous historical and contemporary reports of CTVT, no studies had previously carried out a systematic survey of CTVT's worldwide distribution. The map displayed in Figure 1-7 shows the most detailed analysis of CTVT distribution published at the time I commenced my PhD. As its main limitation, this analysis did not distinguish countries where CTVT is absent, and countries with no data available (both in green colour).

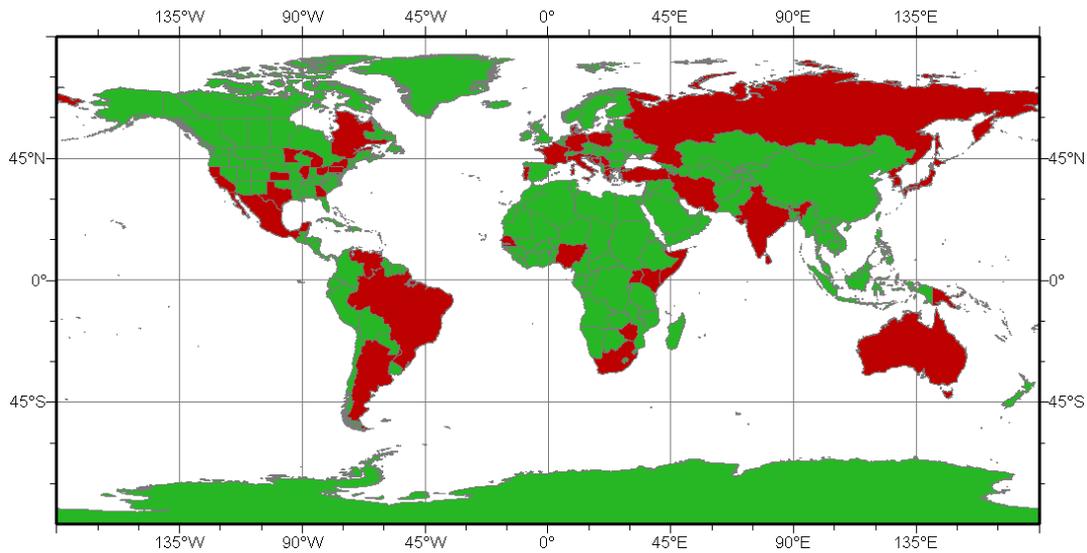


Figure 1-7 Map of CTVT worldwide distribution published before this PhD work was commenced (figure obtained from (Stimmelmayer 2010)). CTVT cases have been reported from countries or states coloured in red. No CTVT cases have been reported or data is not available for countries or states coloured in green. Note that despite its limitations, this is the most detailed analysis of CTVT worldwide distribution published at the time I commenced my PhD.

1.2.7 Confirmation of the clonal nature of CTVT

After years of controversy whether the transmissible agent in CTVT was an oncogenic virus or the cell itself, confirmation of the clonal nature of CTVT was presented by a number of studies over time, through transplantation experiments, cytogenetic studies and genetic analysis, as described further below.

1.2.7.1 Transplantation studies

The first published observations reporting that CTVT could be transplanted subcutaneously, and thus suggesting clonal transmission, were made by Nowinsky in 1876, who performed serial passage of CTVT cells by transplantation between unrelated puppies of different breeds (Nowinsky 1876). Later, further extensive experimental transplantation studies confirmed that CTVT could be transplanted between dogs by the direct transfer of cancer cells or tumour tissue (Wehr 1888, Wehr 1889, Smith and Washbourn 1897, Smith and Washbourn 1898b, Smith and Washbourn 1898a, Sticker 1904, Sticker 1906a, Stubbs and Furth 1934, Karlson and Mann 1952). Despite numerous efforts, experimental transmission using cell-free extracts was performed with a limited success (Ajello 1960, Cohen 1985), indicating that

tumour cells were necessary for transmission, and thus disproving route of transmission through an oncogenic virus.

1.2.7.2 *Cytogenetics*

Cytogenetic studies of tumours from different countries (including France, Jamaica, Japan, Nigeria, Russia, Uganda and the United States) further supported the clonal nature of this disease. The chromosome number in CTVT from different parts of the world was shown to consistently be 57-59 (Sofuni and Makino 1963, Weber et al. 1965, Barski and Cornefert-Jensen 1966, Kakpakova et al. 1968, Thorburn et al. 1968, Murray et al. 1969, Wright et al. 1970, Oshimura et al. 1973, Idowu 1977, Richardson et al. 1987, Fujinaga et al. 1989), compared to constitutive number of 78 in normal dogs. The CTVT karyotype was shared by all CTVT tumours and characterised by extensive and specific chromosomal aberrations, with the similarity highlighted by presence of 15-17 metacentric or submetacentric chromosomes, contrasting to only 2 found in normal dogs (Sofuni and Makino 1963, Weber et al. 1965, Barski and Cornefert-Jensen 1966, Kakpakova et al. 1968, Thorburn et al. 1968, Murray et al. 1969, Wright et al. 1970, Oshimura et al. 1973, Idowu 1977, Richardson et al. 1987, Fujinaga et al. 1989).

1.2.7.3 *Genetics*

1.2.7.3.1 *LINE-MYC rearrangement*

Additional important evidence for clonal transmission of CTVT was provided by the identification of a genomic rearrangement involving the insertion of a repetitive 1.5 kb DNA segment derived from a long interspersed nuclear element (LINE) upstream of the first exon of the *MYC* gene, that was identified in CTVT tumours but not in normal dogs (Katzir et al. 1985, Katzir et al. 1987). Subsequent studies confirmed that this *LINE-MYC* rearrangement is an early and stable marker of CTVT (Amariglio et al. 1991, Choi et al. 1999, Choi and Kim 2002, Liao et al. 2003, Murgia et al. 2006, Park et al. 2006, Rebbeck et al. 2009). The effect of insertion of the truncated LINE element on the *MYC* gene transcription is not clear, although *MYC* expression is detected in CTVT (Katzir et al. 1987).

1.2.7.3.2 *Microsatellites and the Major Histocompatibility Complex*

Clonal transmission of CTVT was further confirmed by molecular studies, which found that the pattern of microsatellite markers and MHC polymorphisms is highly conserved in a set of CTVT tumours from distant parts of the world and distinct from those in corresponding CTVT

hosts, indicating a monophyletic origin for the disease (Murgia et al. 2006, Rebbeck et al. 2009). Murgia et al. have further shown that mitochondrial and MHC variation confirms the clonal nature of the disease, and further divides modern CTVT tumours into two sub-clades with distribution across many countries (Murgia et al. 2006).

1.2.7.3.3 Genome

Next-generation sequencing analysis of two complete CTVT genomes from Australia and Brazil has further highlighted the remarkable similarity between CTVT tumours from different parts of the world, thus providing additional evidence for CTVT being a single clonal cell lineage (Murchison et al. 2014) (see also Figure 1-9).

1.2.8 Age, origin and clonal evolution of CTVT

1.2.8.1 Age and origin of CTVT

After unequivocal evidence showing that CTVT is indeed a single clone (Murgia et al. 2006, Rebbeck et al. 2009), a number of studies have sought to estimate the approximate time when CTVT arose, to ascertain whether the disease is a recently emerged pathogen or whether it has an ancient origin (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014). The oldest known report, published in 1810, of a disease that is likely to be CTVT would argue that it is at least a few hundred years old (Blaine 1810). Microsatellite and phylogenetic analyses of CTVT tumours indicated that CTVT may be between 200 and 2500 years old (Murgia et al. 2006) or between 6,500 and 78,000 years old (Rebbeck et al. 2009), with the most recent common ancestor of current tumours existing only 47-470 years ago (Rebbeck et al. 2009). Later estimates using number of C>T mutations at CpG molecular clock based on human medulloblastoma (43.3 mutations of this signature genome-wide per year) estimated that CTVT may have arisen ~11,000 years ago (Murchison et al. 2014). These findings indicate that CTVT may have arisen around the time of dog domestication, which took place around 10,000-15,000 years ago (Savolainen et al. 2002, Freedman et al. 2014). Overall, CTVT is the oldest known cell lineage in nature.

Genetic studies have been performed to examine the breed of origin of CTVT and features of the original 'founder' dog that gave rise to this cancer lineage (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014). Although early phylogenetic analyses based on major histocompatibility complex, microsatellite typing and sequence analysis of the RPPH1 gene were unable to distinguish between a wolf or an ancient-breed dog origin (Murgia et al. 2006, Rebbeck et al. 2009), analysis using polymorphic SNP loci has indicated that the CTVT founder

animal was likely to have been a dog belonging to one of the ancient breeds (Murchison et al. 2014). Additionally, in this analysis, the CTVT founder animal clustered most closely with Alaskan malamutes and huskies, and comparison to canine phenotypic features indicated that the founder animal was likely to have been of medium or large size with an agouti or solid black coat (Murchison et al. 2014).

Overall, at the time when I started my PhD, CTVT was known to have first emerged from the somatic cells of the CTVT 'founder dog' that lived approximately 11,000 years ago (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014) and to have been transmitted between individual dogs by the allogeneic transfer of living cancer cells, as illustrated in Figure 1-8.

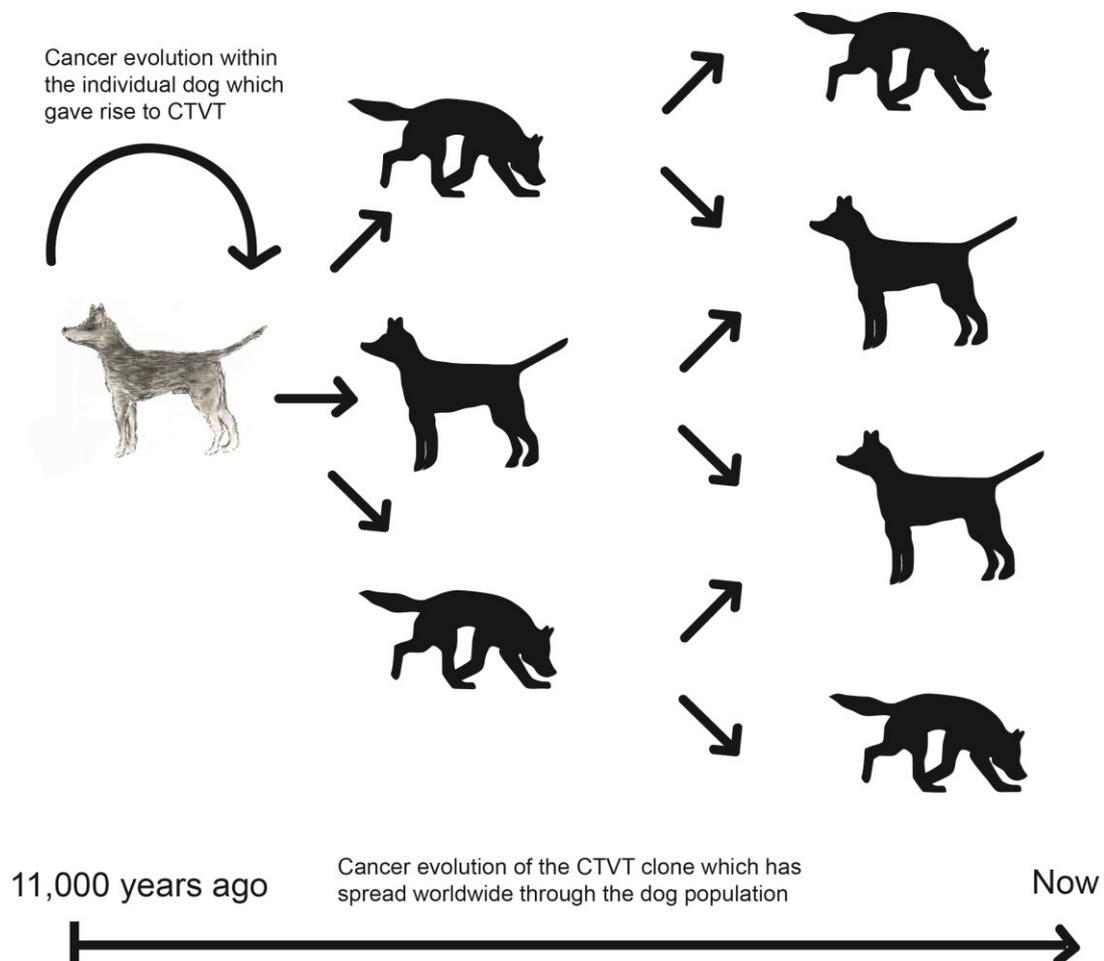


Figure 1-8 Transmission of canine transmissible venereal tumour (CTVT) as a clonal cell lineage between unrelated dogs. CTVT first arose around 11,000 years ago in a 'founder' dog and then spread through the dog population as an infectious disease. Image of founder dog provided by Elizabeth Murchison, credit: Emma Werner.

1.2.9 CTVT genome

Many early studies on CTVT have focused on clinical features, transplantation experiments and treatment options, with only a relatively small number of studies focusing on CTVT genetics (Sofuni and Makino 1963, Weber et al. 1965, Barski and Cornefert-Jensen 1966, Kakpakova et al. 1968, Murgia et al. 2006, Rebbeck et al. 2009, Thomas et al. 2009, Rebbeck et al. 2011). Recently, however, with the advent of next-generation sequencing technologies, the complete genomes of two CTVT tumours from Australia and Brazil have been sequenced (Murchison et al. 2014). Analysis of these genomes revealed that the cancer has acquired approximately 1.9 million somatic substitution mutations as well as thousands of rearrangements, copy-number changes, and retrotransposon insertions (Murchison et al. 2014). Indeed, CTVT has acquired an order of magnitude more somatic mutations than most human cancers, most of which usually have between 1,000 and 10,000 somatic substitution mutations (occasionally they can have many more, even over 100,000) (Stratton 2011). Additionally, more than 10,000 non-synonymous variants have been identified in the CTVT genome and 646 genes have been lost altogether (Murchison et al. 2014). Although a point mutation was identified in the tumour suppressor *TP53*, it was shown not be involved in the clonal origin of CTVT (Choi and Kim 2002, Vazquez-Mota et al. 2008).

A large proportion of the genome is diploid with loss of heterozygosity (LOH) (Rebbeck et al. 2009, Murchison et al. 2014), suggesting that diploidy may be the most stable state for a long-lived lineage. This contrasts with cancers remaining in one individual, which often have significant changes in chromosomal copy number (Thomas et al. 2005, Beroukhim et al. 2010).

Interestingly, however, despite the enormous number of mutations and marked aneuploidy in CTVT, genomic rearrangements and microsatellite alleles observed in CTVT tumours collected from different continents are remarkably similar (Sofuni and Makino 1963, Weber et al. 1965, Murgia et al. 2006, Rebbeck et al. 2009, Thomas et al. 2009, Murchison et al. 2014). It is therefore possible that CTVT has maintained or activated DNA repair mechanisms that safeguard its genome against further mutation and instability. Additionally, given the large mutational burden already carried by CTVT, its genome may be particularly sensitive to further mutation such that negative selection acts to maintain stability. It is interesting that the oldest human cancer lineage, the HeLa cell line, which has continued to survive by passaging in laboratory cell culture for more than sixty years, also appears to have a relatively stable genome in terms of point mutation (Adey et al. 2013). Given that CTVT is the oldest

known somatic cell lineage, we can speculate whether telomere stabilisation may be another possible mechanism used by CTVT to maintain its genome integrity (Chu et al. 2001).

1.2.10 Diversity amongst CTVT tumours

Despite the similarity between CTVT tumours from different parts of the world, many studies have also alluded to the cytogenetic, genetic and phenotypic diversity between CTVT tumours, which arose as the lineage has spread globally and diverged to some extent along the way (Sofuni and Makino 1963, Weber et al. 1965, Cohen 1985, Murgia et al. 2006, Rebbeck et al. 2009, Thomas et al. 2009, Murchison et al. 2014).

Despite the remarkable similarity of CTVT karyotypes from different parts of the world, these same studies have highlighted the small differences that exist between tumours (Sofuni and Makino 1963, Weber et al. 1965, Barski and Cornefert-Jensen 1966, Kakpakova et al. 1968, Thorburn et al. 1968, Murray et al. 1969, Wright et al. 1970, Oshimura et al. 1973, Idowu 1977, Richardson et al. 1987, Fujinaga et al. 1989). The number of chromosomes in CTVT varies from 57 to 59, and the number of metacentric chromosomes ranges from 15 to 17 (Sofuni and Makino 1963, Weber et al. 1965, Barski and Cornefert-Jensen 1966, Kakpakova et al. 1968, Thorburn et al. 1968, Murray et al. 1969, Wright et al. 1970, Oshimura et al. 1973, Idowu 1977, Richardson et al. 1987, Fujinaga et al. 1989), suggesting that CTVT may have diverged into different sub-clones characterised by specific karyotypic changes.

Genetic studies have also implied diversity amongst CTVT tumours worldwide (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014). Murgia et al. have analysed mitochondrial DNA and MHC copy number variation to show that the single CTVT clone has diverged into two subclades early in its evolution, each of which has gained a broad distribution (Murgia et al. 2006). The analysis performed by Rebbeck et al. shows a substantial geographical structure of CTVT tumours, with genotypes shared by locally distributed tumours as well as by tumours found on different continents, indicating a relatively recent intercontinental spread (Rebbeck et al. 2009). The analysis of two complete CTVT genomes from Brazil and Australia further pointed to the genetic variations acquired by CTVT as it has spread around the world (Murchison et al. 2014). Although most somatic mutations were shared between tumours, a small proportion were unique to each tumour (i.e. acquired after divergence of these two lineages) (Murchison et al. 2014). The genomic copy number profile is also to a large extent shared by these two tumours, but at the same time it revealed a small number of differences between the lineages (see Figure 1-9).

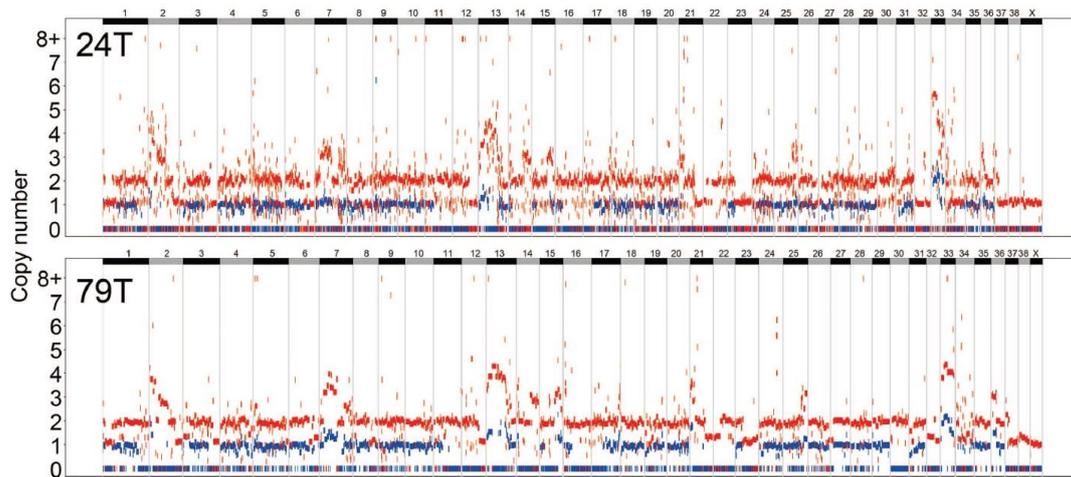


Figure 1-9 Genomic copy number for CTVT tumours 24T (Australia) and 79T (Brazil), as published in (Murchison et al. 2014). Red and blue points represent total copy number and minor copy number (i.e. copy number of the allele present in fewer copies), respectively, calculated by using normalized read counts at each of 2,544,508 single nucleotide polymorphism (SNP) loci. Chromosomes are represented by horizontal alternating black and grey bars. Note the similar genomic copy number profile of the two CTVT tumours.

Phenotypic variation between tumours has also been noted, especially at different stages of tumour growth (see section 1.2.5). Supporting fibrous stroma is usually sparse in tumours that have developed in the host for only several weeks and is more abundant in well-established CTVT masses (Cohen 1985). Regressing tumours are often infiltrated by lymphocytes, plasma cells, and macrophages (Wade 1908, Bloom et al. 1951, Weir et al. 1978, Chandler and Yang 1981).

1.3 Canine genetics

Canine transmissible venereal tumour (CTVT) is found to naturally infect the domestic dog, *Canis familiaris*, although experimental transmission to other members of the Canidae family has also been reported (Sticker 1906a, Wade 1908, Dungern 1912, Cockrill and Beasley 1979, Cohen 1985). This Thesis focuses on the genetic diversity and evolution of CTVT, and therefore evolution of its host species is discussed at this stage.

1.3.1 Dog domestication and global dispersal

Dogs were the first domesticated animals (Larson et al. 2012), and extensive genetic analyses clearly show that domestic dogs arose from grey wolves (*Canis lupus*) (Wayne et al. 1987a, Wayne et al. 1987b, Wayne et al. 1989, Vila et al. 1997, Vila et al. 2005). On the other hand, despite many efforts by archaeologists as well as geneticists, the geography and timing of dog domestication together with the number of domestication events remain debated. The date of divergence between wolves and dogs has been estimated to be at around 15,000 – 30,000 years ago (Vila et al. 1997, Savolainen et al. 2002, Pang et al. 2009, Thalmann et al. 2013, Freedman et al. 2014, Frantz et al. 2016, Wang et al. 2016). Several geographical regions have been suggested as the place of dog domestication, including Europe (Thalmann et al. 2013), East Asia (Savolainen et al. 2002, Wang et al. 2016), the Middle East (Vonholdt et al. 2010) and Central Asia (Pang et al. 2009, Shannon et al. 2015). A recent study, which analysed DNA from a ~4,800 years old dog from Ireland, proposed two domestication events on either side of Eurasia (Frantz et al. 2016). This proposal was however subsequently challenged by a study analysing DNA from two ancient dogs, ~7,000 years old and ~4,700 years old (Botigue et al. 2017).

Although the debate on the origin of dog domestication continues, once domesticated, dogs have rapidly spread around the globe resulting in genetically divergent populations and breeds found in different parts of the world (Leonard et al. 2002, Parker et al. 2004, Ostrander and Wayne 2005) known as ‘basal’ or ‘ancient’ breeds.

1.3.2 Breed formation

It was only within the last few hundred years, however, that selective breeding practices have divided dogs into more than 350 ‘modern’ breeds (Ostrander and Wayne 2005, Dobson 2013). The phenotypic diversity amongst the dog breeds is striking and the diversity in skeletal size and proportion of dogs exceeds that in any other mammalian species (Wayne

1986a, Wayne 1986b, Drake and Klingenberg 2010). Studies have shown that the overall level of diversity in dogs is twice of that in humans (Boyko et al. 2010). The rapid diversification of dog breeds through artificial selection has been termed as one of the 'greatest genetic experiments ever conducted by humans' (Ostrander and Wayne 2005).

In contrast, the genetic diversity within breeds is extremely low, due to restrictions imposed by breeding associations. In many breeds, the founder events included only a few dogs (Ostrander and Wayne 2005) leading to a closed population with minimal gene flow between breeds (Dobson 2013). Estimates suggest that domestication of the wild canid population led to 5% loss of nucleotide diversity, while dog breed formation resulted in a decrease by 37% (Gray et al. 2009, Dobson 2013). This restricted genetic diversity has resulted in unintentional selection of mutations that may be predisposing for a wide range of diseases (Parker et al. 2010).

1.4 Mitochondrial genome

Much of this Thesis focuses on analysis of mitochondrial DNA, and it is therefore appropriate to discuss the structure, functions and genetics of the mitochondrial genome.

1.4.1 Structure and function of the mitochondrial genome

Mitochondria are essential organelles within cells, whose main function is to carry out aerobic respiration and produce ATP (Adenosine Tri-Phosphate) (Hatefi 1985). Each mitochondrion carries a few copies of the circular mitochondrial DNA genome (mtDNA), which is about 15,000-17,000 base pairs long in mammals (Bogenhagen and Clayton 1974, Larsson and Clayton 1995). Each cell possesses numerous mitochondria that are constantly undergoing dynamic processes of fusion and fission (Westermann, 2010). There are on average 10^3 to 10^4 mtDNA copies per cell (Wiesner et al. 1992, Lightowlers et al. 1997, Rooney et al. 2015). MtDNA copy number varies between cell types - for example, in mammalian sperm cells, the copy number of mtDNA is very low (less than 100 copies (Hecht et al. 1984)), whereas in mammalian oocytes the copy number is extremely high (more than 10^5 copies (Michaels et al. 1982, Chinnery et al. 2000)). Mitochondrial DNA is maternally inherited, and although sperm-derived mitochondria enter the oocyte, they are never transmitted to offspring (Giles et al. 1980, Sato and Sato 2013).

The mitochondrial genome encodes 13 essential polypeptides, which all form parts of the mitochondrial respiratory chain that is essential for aerobic respiration. It also encodes RNA machinery used for mitochondrial protein synthesis, which includes 2 rRNAs and 22 tRNAs (Lightowlers et al. 1997, Taylor and Turnbull 2005). Additional subunits of respiratory chain complexes are encoded by the nuclear DNA, synthesised on the cytoplasmic ribosomes and subsequently targeted into mitochondria (Taylor and Turnbull 2005). There are no introns and almost no intergenic noncoding sequences, but there is a single major, but small (~1kb), non-coding region called the control region or displacement loop (D-loop), which is involved in regulation of transcription and replication (Larsson and Clayton 1995, Tuppen et al. 2010). The D-loop, bounded by genes for tRNA^{Phe} and tRNA^{Pro} is the most polymorphic part of the mtDNA genome, with three small conserved regions (called conserved sequence blocks) (Walberg and Clayton 1981).

MtDNA has a very high mutation rate – it has been estimated that the mtDNA mutation rate is at least 10-fold higher than the observed nuclear DNA mutation rate (Kim et al. 1998, Mambo et al. 2003, Tuppen et al. 2010). This is presumably due to the reduced efficiency of

mtDNA repair systems, coupled with decreased fidelity of mtDNA replication and presence of high levels of reactive oxygen species (ROS) generated by the electron transport chain (Kim et al. 1998, Mambo et al. 2003, Tuppen et al. 2010).

The nuclear and mitochondrial genomes differ in their genetic code and mode of packaging. The mitochondrial genetic code uses codons AUA and AUG to code for methionine, UGA for tryptophan, rather than a termination codon, and AGA with AGG for stop codons, rather than for arginine as in the nuclear genome (although this may be debated) (Stewart and Chinnery 2015). In contrast to nuclear DNA, histones are lacking from mtDNA, perhaps adding to the increased mutation rate (Tuppen et al. 2010). MtDNA is instead packaged, together with proteins, into nucleoids, each of which contains one or two mtDNA molecules (Stewart and Chinnery 2015). In a nucleoid, mtDNA is densely packed by the binding of proteins, such as mitochondrial transcription factor A (TFAM) and mitochondrial single-stranded DNA-binding protein (mtSSB) (Wang and Bogenhagen 2006, Lee and Han 2017).

In 1982, the first evidence for uptake of mitochondrial genetic material by cells was reported (Clark and Shay 1982). The question of whether mitochondrial horizontal transfer occurs in mammalian cells was again revisited only during the last decade, after publication of a seminal paper documenting *in vitro* organelle transfer between cells through tunnelling nanotubes (TNTs) (Rustom et al. 2004). Subsequently, a number of studies have demonstrated occurrence of mitochondrial horizontal transfer *in vitro* (Spees et al. 2006, Berridge et al. 2015, Berridge et al. 2016), suggesting that this phenomenon might also be widely occurring in whole organisms. In the last few years, mitochondrial horizontal transfer has also been shown to occur *in vivo* (Islam et al. 2012, Ahmad et al. 2014, Tan et al. 2015, Dong et al. 2017). The mechanisms by which mitochondria are transferred between cells are still debated, but tunnelling nanotubes (TNTs), micropinocytosis (Patel et al. 2017), extracellular vesicles or cell fusion have been proposed as possible mechanisms (Torralba et al. 2016).

The current belief is that mtDNA does not recombine and this question has been discussed in the literature for a few decades (Lightowers et al. 1997, Taylor and Turnbull 2005). In animals where mtDNA is inherited maternally, presence of a single haplotype presents an enormous challenge for detecting recombination (Chen 2013). A study which analysed a maternal mouse line that has been heteroplasmic for a defined set of mtDNA mutations for more than 50 generations has concluded that mtDNA recombination does not take place (Hagstrom et al. 2014). On the other hand, examples of mtDNA recombination have been

reported between maternal and paternal haplotypes in rare cases of human biparental mtDNA inheritance (Kraytsberg et al. 2004, Zsurka et al. 2005), mtDNA recombination activity was detected in human cell extracts (Thyagarajan et al. 1996) and has been observed in the germline of various eukaryotes (Lunt and Hyman 1997, Ladoukakis and Zouros 2001, Hoarau et al. 2002, Bergthorsson et al. 2003, Gantenbein et al. 2005, Ujvari et al. 2007). Furthermore, the process of mtDNA recombination has been proposed as a mechanism for mtDNA repair (Thyagarajan et al. 1996).

Given the presence of two mitochondrial molecules in the same cell coupled with availability of the recombination machinery in the mitochondria, there seems to be no natural barrier to mtDNA recombination in mammals (Chen 2013). Widespread occurrence and frequency of mtDNA recombination in mammalian cells, however, still remains a controversial topic.

1.4.2 Mitochondrial DNA homoplasmy and heteroplasmy

As discussed above, each mitochondrion contains a number of mtDNA molecules and each cell contains a number of mitochondria. When all mtDNA molecules within a cell are identical, the situation is known as homoplasmy. The high mutational rate of mtDNA may, however, lead to evolution of distinct mtDNA haplotypes. Presence of two or more distinct mtDNA subpopulations within an individual, within a single cell or in discrete organelles results in heteroplasmy (Lightowlers et al. 1997, Taylor and Turnbull 2005, Tuppen et al. 2010) (see Figure 1-10).

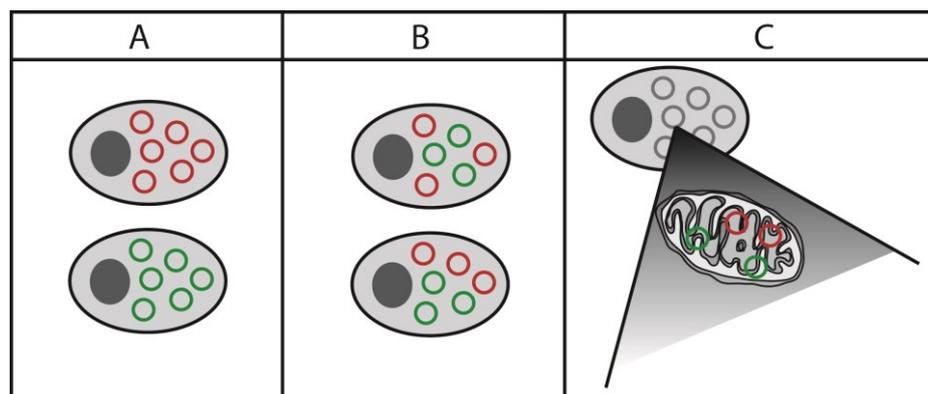


Figure 1-10 Illustration of mtDNA heteroplasmy within (A) different cells of an individual, (B) within a single cell and (C) in discrete mitochondrial organelles. Red and green circles represent different mtDNA haplotypes.

Mutations in mtDNA have been reported to lead to clonal expansion of particular mtDNA haplotypes both in normal and neoplastic tissues (Polyak et al. 1998, Fliss et al. 2000, Habano et al. 2000). In contrast to the nuclear genome, distinct mtDNA haplotypes are placed into direct competition in these cases. The processes by which a mutant mtDNA haplotype becomes dominant include the following: (1) haplotypes may be positively selected at the level of the cell for a functional advantage, e.g. more efficient respiratory function; (2) haplotypes may be selected due to a 'selfish' advantage resulting from more efficient replication or transmission to daughter cells (Ma and O'Farrell 2016); or (3) haplotypes may become homoplasmic by chance through neutral genetic drift (Coller et al. 2001) (see Figure 1-11).

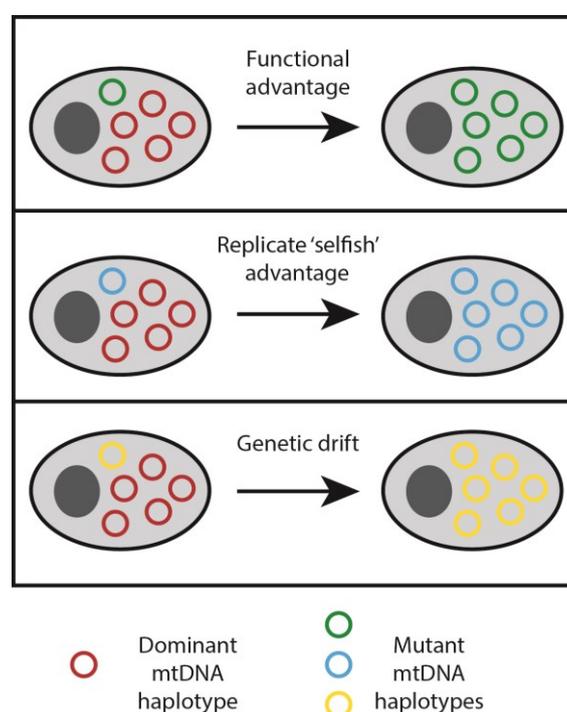


Figure 1-11 Summary of processes by which a heteroplasmic mtDNA variant may become homoplasmic.

1.4.3 Mitochondrial DNA mutations in cancer

Somatic mtDNA mutations have for a long time been proposed to play a role in cancer progression (Brandon et al. 2006, Ohta 2006, Zong et al. 2016, Yuan et al. 2017). Much interest in mitochondrial function in cancer has revolved around the fact that many cancers have altered energy metabolism (Warburg 1956b, Warburg 1956a, Hanahan and Weinberg 2011). Although there are numerous studies reporting somatic mtDNA mutations in cancer,

clear evidence for mtDNA mutations acting as causative 'driver' mutations in cancer, rather than as neutral 'passenger' mutations, are still debated (Schon et al. 2012, Zong et al. 2016).

Recent analysis of mitochondrial genomes in cancer indicated an endogenous mitochondrial mutational process dominated by C>T and T>C mutations with a strong strand bias (Ju et al. 2014, Stewart et al. 2015, Yuan et al. 2017). Importantly, these studies reported presence of negative selection for deleterious, pathogenic mutations, indicating importance of functional mitochondria in some cancers (Ju et al. 2014, Yuan et al. 2017).

1.4.4 Canine mitochondrial DNA

The complete nucleotide sequence of the dog mitochondrial genome (mtDNA) was reported in 1998 by Kim et al (Kim et al. 1998). Even though the full dog reference genome, that of a female Boxer dog, was sequenced by the Broad Institute in 2005 (Lindblad-Toh et al. 2005) and updated in 2011, the dog mitochondrial reference genome remained the previously published Korean ghost dog (Kim et al. 1998).

The dog mtDNA genome is 16,727bp long, with the same organisation of genes encoding 13 polypeptides, 22 tRNAs and 2 rRNAs, as human mtDNA (Kim et al. 1998). The non-coding control region (D-loop) is located between positions 15,458-16,727 (Kim et al. 1998, Pereira et al. 2004) and is the most polymorphic part of the mtDNA. The canine mtDNA control region, in contrast to the human mtDNA, has a 10bp repeat unit (GTACACGT(A/G)C, where A and G are variable between repeat units) that begins at base 16,130 and varies in number and sequence both within and among individuals (Gundry et al. 2007). Additionally, in dog mtDNA, there is variation in presence and length of polyC homopolymer tracts around positions 15512–15535 (Fregel et al. 2015), which have not been reported in humans.

Given the high degree of genetic variation in mtDNA (present especially in the mtDNA control region, which is the most polymorphic part of the mtDNA genome), together with the high copy number of mtDNA within each cell, mtDNA has proven to be useful for genetic investigations. Canine mtDNA has been shown to exhibit inter and intra breed variation (Okumura et al. 1996), and therefore may not be used to assign individuals to breed. The analysis of mtDNA control region variation has, however, been extensively used for canine forensic and phylogenetic studies (Wayne 1993, Savolainen et al. 1997, Savolainen et al. 2000).

Previously published mtDNA analyses date the most recent split between the wolf and the dog at around 15,000 – 30,000 years ago (Vila et al. 1997, Savolainen et al. 2002, Pang et al. 2009, Thalmann et al. 2013). Based on mtDNA sequence data, dogs fall within one of six clades on a maximum likelihood tree when compared with wolves, named clades A-F (Vila et al. 1997, Savolainen et al. 2002, Pang et al. 2009, Thalmann et al. 2013) (see Figure 1-12). Most dogs have mitochondrial haplotypes belonging to clade A, which is represented on all continents. Dog belonging to haplotype groups in clades B and C, on the other hand, have not been observed in the Americas. Haplotypes in clades D, E and F have only been reported regionally in Scandinavia, Turkey and Spain; Japan and Korea; and Japan and Siberia, respectively (Savolainen et al. 2002).

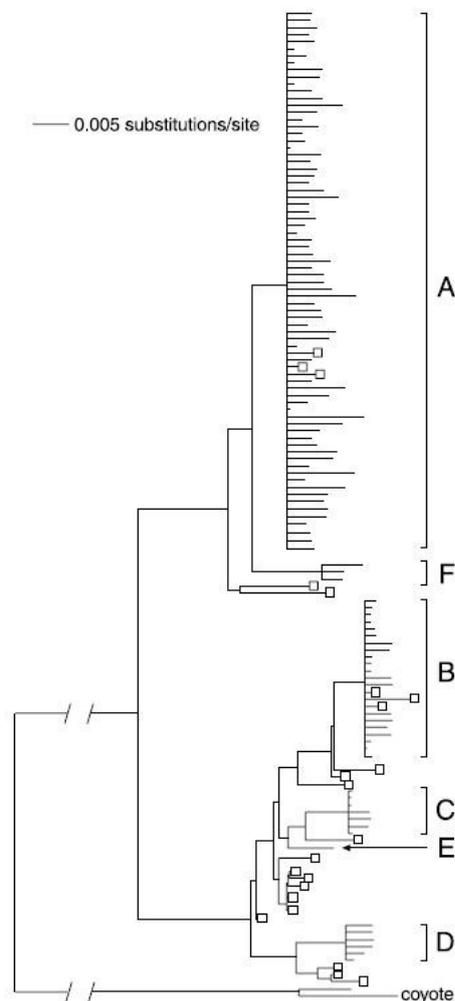


Figure 1-12 Phylogenetic tree of dog (unlabelled) and wolf (open squares) haplotypes, obtained from (Savolainen et al. 2002). Six mitochondrial clades (A to F) of dog haplotypes are indicated. Scale bar indicates base substitutions per site. The branch leading to the outgroup (coyote) was reduced by 50%.

1.4.5 Mitochondrial DNA in CTVT

As discussed in section 1.2.4, canine transmissible venereal tumour (CTVT), arose in an individual dog and has been propagated through the canine population as a single clonal lineage; it has thus been shown to share a monophyletic origin when clustering on a tree of canine nuclear genomes (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014). In contrast, CTVT mtDNA was reported to form two distinct phylogenetic groups on a phylogenetic tree constructed using canine mitochondrial genomes (Murgia et al. 2006, Rebbeck et al. 2011). This finding indicated that CTVT mtDNA is not clonal, but rather has been acquired by periodical mtDNA horizontal transfer events from transient canine hosts (Rebbeck et al. 2011). Further this suggests that replacement of CTVT mtDNA, which presumably was carrying large numbers of possibly deleterious mutations, may have provided a selective advantage to the lineage (although we cannot exclude the possibility that CTVT acquired host mitochondrial DNA via purely neutral processes) (Rebbeck et al. 2011).

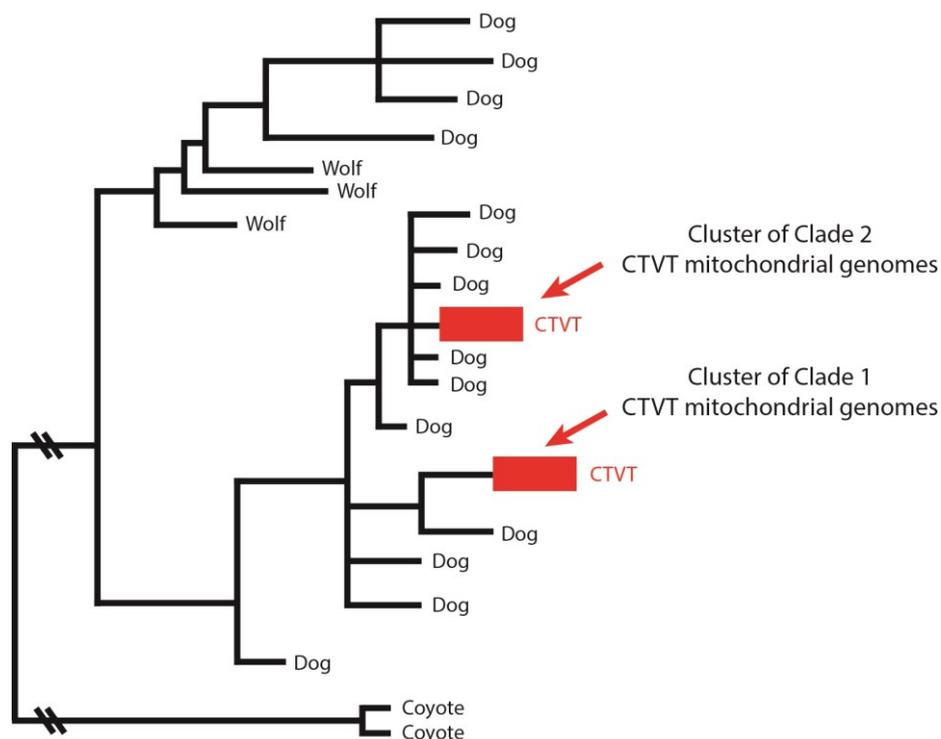


Figure 1-13 Representation of maximum parsimony phylogeny of CTVT mitochondria adapted from (Rebbeck et al. 2011). The red boxes represent mitochondrial genomes from CTVT samples collected from Mexico, Israel, South Africa, Thailand, Kenya, Greece and Malaysia. Note that not all dog mitochondrial clades from Figure 1-12 are represented in this figure. All CTVT samples belong to dog clade A (in contrast, all dog mitochondrial clades are shown in Figure 1-12). The black lines represent normal canine mitochondrial genomes. Note that the CTVTs do not cluster in one monophyletic group; rather they form two distinct clusters, called CTVT clades.

1.5 Goals of this Thesis

At the time when I started my PhD work, the complete genomes of two CTVT tumours had recently been sequenced using next-generation sequencing approaches (Murchison et al. 2014), which confirmed the clonal origin of the cells, but at the same time alluded to the genetic variation present between tumours from different parts of the world.

In this PhD project, my main interest was to understand this genetic diversity and evolution of CTVT tumours worldwide, which has further been alluded to by other genetic studies (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014), but has not been extensively understood. Particularly, as CTVT has spread around the world, distinct lineages may have acquired different adaptations that may be uncovered through genetic and phenotypic analyses. In order to answer these questions, I set about collecting hundreds of CTVT tumour samples from all six inhabited continents. My genetic analysis mostly focused on complete mitochondrial genomes (mtDNA) in CTVTs – as it had previously been suggested that CTVT has acquired mtDNAs through horizontal transfer events, and therefore my interest was to survey this mtDNA genetic diversity in a larger population of CTVT tumours.

Overall, this Thesis addressed questions surrounding the worldwide diversity, spread and evolution of CTVT, with the five main goals of this work discussed below:

(1) To gain further understanding of CTVT prevalence around the world and to decipher historical patterns of global CTVT spread

CTVT arose as a cancer in a single dog that lived around 11,000 years ago (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014). However, published data suggest that CTVT was likely confined to a single place for most of its history, until it rapidly spread around the world approximately 500 years ago (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014). The aim of this Thesis was to survey the current distribution of CTVT around the world, and to use genetics to determine historical patterns of disease spread.

I approached this goal by performing a survey of the historical and current worldwide distribution and prevalence of CTVT (Chapter 2) and by collecting/validating over 1400 CTVT tumour samples from over 50 countries (Chapter 3). Moreover, I utilised CTVT mitochondrial genome sequencing data to understand the phylogenetic relationships between CTVT tumours and to trace the spread of the disease around the globe in the last 2,000 years (Chapter 4).

(2) To determine the frequency of mitochondrial horizontal transfer and search for evidence of mitochondrial genome recombination

CTVT has been previously suggested to periodically acquire mitochondrial DNA (mtDNA) from its hosts, and this process was previously found to have happened at least twice in the history of the lineage (Rebbeck et al. 2011) (see section 1.4.5). The aim of this Thesis was to provide further evidence for this unusual biological phenomenon and to determine the frequency of mtDNA capture in CTVT. Additionally, mtDNA recombination is a process that, to my knowledge, had not previously been detected in cancer at the time that this PhD work was started, and I aimed to use CTVT mtDNA sequencing data to search for evidence of this phenomenon.

I approached these questions by analysing germline mitochondrial variants in diverse CTVT tumours for evidence of mtDNA horizontal transfer and mtDNA recombination. The analysis presented in Chapter 4 showed that CTVT has captured mtDNA through horizontal transfer events at least five times in the history of the lineage, delineating five CTVT tumour clades. Evidence presented in Chapter 4 shows that mtDNA recombination has caused occasional mtDNA re-assortment in CTVT.

(3) To understand the effects of negative selection acting on mitochondrial genomes in CTVT

Natural selection is a fundamental process affecting all evolving populations. CTVT gives us a unique opportunity to study evolution of a cancer cell lineage that has been surviving for thousands of years, in comparison to cancers in humans, which normally survive only for decades. Interestingly, at the time when I started this work, there was little evidence in human cancers for negative selection operating to safeguard essential cellular processes. The aim of this Thesis was therefore to probe the question whether accumulation of mutations may have become a burden rather than an advantage in this long-lived lineage.

I analysed mtDNA somatic mutations to understand the patterns of selection in CTVT mtDNA. Analysis of synonymous and non-synonymous changes in the somatic set of variants indicated that negative selection has operated to prevent accumulation of deleterious mutations in captured CTVT mtDNA (Chapter 4).

(4) To map the phenotypic diversity of CTVT tumours around the world, and explain phenotypic traits by underlying genetic changes

Despite the fact that CTVT is one clonal cell lineage that arose from a single dog living a few thousand years ago (see section 1.2.8), perhaps one of the most interesting features of the disease is the diversity of CTVT tumours around the world. The aim of this Thesis was to map the histopathological phenotypes distinct to major CTVT mitochondrial clades and explain these differences by underlying genetic changes.

To answer the question above, I performed a histopathology based screen of CTVT tumours from mitochondrial clades 1 and 2 (Chapter 3), to decipher any phenotypic changes resulting from mtDNA capture or phylogenetic divergence between clades. This screen did not show any significant differences between groups (Chapter 4).

(5) To understand phylogenetic relationships between individual CTVT mtDNA clades

Mitochondrial DNA sequencing data only permits analysis of genetic events occurring after each mtDNA horizontal transfer event. The aim of this Thesis was to understand the phylogenetic structure of the CTVT lineage, in order to determine how the mtDNA horizontal transfer events fit into the CTVT phylogenetic tree.

In order to determine how CTVT mtDNA clades relate to each other, I analysed exome data from the same set of CTVT tumours analysed in chapter 4. This analysis revealed that a single canine mtDNA haplogroup has recurrently and recently undergone multiple horizontal transfer events (Chapter 5).

The following chapters in this Thesis present work which I undertook to answer the goals listed above. A short summary of each Chapter is provided below:

Chapter 2 describes the design and analysis of a questionnaire-based survey which was used to gain a snap-shot of current CTVT distribution patterns around the world. It also describes how I used a review of published literature to infer recent changes in CTVT distribution patterns in western countries.

Chapter 3 reports the optimisation of a diagnostic validation assay, and its use to confirm CTVT diagnosis of more than 1000 tumour samples.

Chapter 4 presents analysis of mitochondrial DNA (mtDNA) sequencing data from 449 CTVT tumours originating in more than 50 countries. The analysis of CTVT mitochondrial diversity revealed that CTVT has captured mtDNA through horizontal transfer events at least five times during the history of the lineage, delineating five tumour clades.

Chapter 5 explains how I used exome sequencing data to determine the phylogenetic relationships between the five CTVT mitochondrial clades presented in Chapter 4.

Chapter 6 is a discussion of all the work presented in this Thesis and draws conclusions on wider implications of this project.

2

Worldwide distribution and prevalence of canine transmissible venereal tumour

Summary

Canine transmissible venereal tumour (CTVT) arose several thousand years ago and has been reported in dogs worldwide; however, its precise distribution patterns and prevalence remain unclear. In this Chapter, I analysed the historical literature and examined CTVT prevalence information obtained through a crowd-sourcing approach from 645 veterinarians and animal health workers in 109 countries around the world. This analysis confirmed that CTVT is endemic in at least 90 countries worldwide across all inhabited continents. The disease is estimated to be present at a prevalence of one percent or more in dogs in at least 13 countries in South and Central America, as well as in at least 11 countries in Africa and 8 countries in Asia. In the United States and Australia, CTVT was reported by the questionnaire respondents to be endemic only in remote indigenous communities. Comparison of current and historical reports of CTVT indicated that its prevalence has declined in Northern Europe possibly due to changes in dog control laws during the nineteenth and twentieth centuries. Through analysis of factors influencing CTVT prevalence I was able to show that presence of free-roaming dogs is associated with increased CTVT prevalence, while dog spaying and neutering is associated with reduced CTVT prevalence. Overall, my work included in this Chapter describes the first systematic survey of current and historical CTVT worldwide distribution and highlights the factors that continue to modify CTVT's prevalence around the world, implicating free-roaming dogs as a reservoir for the disease.

2.1 Introduction

2.1.1 Previous knowledge of CTVT worldwide distribution

CTVT has been reported in many countries around the world (reviewed in (Stookey 1969, Das and Das 2000, Eze et al. 2007, Purohit 2009, Stimmelmayer 2010, Ganguly et al. 2013) and is the oldest and most prolific mammalian somatic lineage known in nature (see Chapter 1, sections 1.2.6 and 1.2.8). Even though CTVT arose from somatic cells of a dog living several thousand years ago (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014), global populations of CTVT diverged only within the last few hundred years (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014), suggesting that CTVT spread around the globe relatively recently. Despite numerous historical and contemporary reports of the disease (reviewed in (Stookey 1969, Vermooten 1987, Das and Das 2000, Eze et al. 2007, Purohit 2009, Stimmelmayer 2010, Ganguly et al. 2013) no systematic study of CTVT's worldwide distribution and prevalence had been performed at the time when I commenced my PhD (see Chapter 1, section 1.2.6).

One of the reasons why such a systematic study of CTVT prevalence was not previously performed may be because estimating the worldwide distribution and prevalence of a common animal pathogen such as CTVT is a challenging task (Alvar et al. 2012, Kretzschmar et al. 2012, Chapman et al. 2013, Gradoni 2013, Milinovich et al. 2013, Angell et al. 2014). In most countries, the disease is not considered notifiable, and in many areas animals do not have access to veterinary care. Veterinary records are often scant or inconsistent, and variation in CTVT prevalence may exist within countries due to seasonal, demographic or geographic factors. These challenges have been considered when performing the analysis presented in this Chapter.

2.1.2 Goals of my work

The goals of this Chapter were two-fold:

- (1)** To gain a snap-shot of the presence or absence of CTVT in various locations around the world, and to estimate the prevalence of CTVT in these locations.
- (2)** To analyse and study the risk factors affecting CTVT prevalence.

This study enabled me to conclude that CTVT is endemic in at least 90 countries worldwide, across all inhabited continents. Additionally, the analysis showed that the disease's

prevalence is linked to national development status and that dog management policies leading to declines in the population of free-roaming dogs may have caused its recent eradication from the United Kingdom.

2.2 Materials and Methods

2.2.1 Questionnaire design and distribution

This study was approved by the Department of Veterinary Medicine, University of Cambridge, Ethics and Welfare Committee (reference number CR105). The questionnaire used in this study is available in supplementary materials (see Additional file 2-1). Potential questionnaire participants were selected with an internet-based search. The questionnaire was sent by email to more than one thousand individuals and distributed at several veterinary conferences. The questionnaire was also circulated in several veterinary societies' mailing lists and published in veterinary periodicals, newsletters and on social media sites. Of the 645 completed questionnaires received, nine respondents provided information about more than one country. Most of the respondents were private veterinarians (415, 64.3%) or veterinarians working at veterinary schools (143, 22.2%), but they also included other individuals working at charitable organisations (50, 7.8%), government (16, 2.5%), pathology laboratories (13, 2.0%) and research agencies (8, 1.2%). All of the respondents completed the questionnaire within days or weeks of receiving it. The questionnaire was made available in English, Chinese, French, Portuguese, Spanish and Russian. Initial questionnaire distribution was carried out during my Part II undergraduate project at the University of Cambridge (Strakova 2013), but additional questionnaires were distributed during the first year of my PhD.

2.2.2 Questionnaire data analysis

Previous studies indicated that CTVT prevalence rarely rises above 20 percent (see Additional file 2-2); therefore the prevalence categories available in the questionnaire were classified as "none", "less than 0.5%", "0.5-1%", "1-3%", "3-5%", "5-10%", "10-20%" and "more than 20%". The estimated average CTVT prevalence for each country was determined using the average of the mid-point values for each reported prevalence category received from each country. Only countries with three or more responses were included in the analyses in Figure 2-2 and Figure 2-8, but all responses are shown in Additional file 2-3, Figure 2-3 and Figure 2-4.

The World Atlas (World Atlas 2012) was used to generate the list of countries in Additional file 2-3, and Taiwan and Reunion were added as separate countries. Data on the distance from the equator and GDP values were obtained from News Track India (News Track India 2011) and the International Monetary Fund (International Monetary Fund 2013)

respectively. Distance from the equator for each country refers to the distance of the capital city from the equator. Countries were classified as low income, lower-middle income, upper-middle income and high income economies based on the classification scheme defined by the World Bank (World Bank 2013).

2.3 Results

2.3.1 Historical worldwide CTVT distribution

I initially analysed the available published literature and searched for reports of CTVT around the world in order to understand the historical distribution of CTVT. I found 317 reports of primary naturally occurring CTVT cases in the published literature (see Additional file 2-4), including case reports, experimental reports and retrospective studies. These records provide evidence of CTVT on all six inhabited continents and range in date from 1810 until 2014 (see Figure 2-1 and Additional file 2-4). The earliest known record of a disease whose description is consistent with CTVT is from London in 1810 (Blaine 1810). In this report, the disease was noted as one of only two cancers known to affect dogs – indicating that CTVT may have been common in London at that time (Blaine 1810). We found evidence that CTVT was present prior to 1910 in the United States (Beebe and Ewing 1906, Beebe 1907), France (Borrel 1907), Germany (von Bergmann 1895, Sticker 1902, Sticker 1904, Sticker 1906a, Sticker 1906b, Bergell and Sticker 1907, Sticker 1907), Italy (Duplay 1894), the United Kingdom (Blaine 1810, Smith and Washbourn 1897, Smith and Washbourn 1898b, Hobday 1900, Powell White 1902, Hobday 1905, Hobday 1906), Japan (Matsui 1909) and Papua New Guinea (Seligmann 1906). Supporting the idea that CTVT is indeed an ancient disease, the 1906 report from Papua New Guinea stated that CTVT was “endemic before the advent of the white man” (Seligmann 1906).

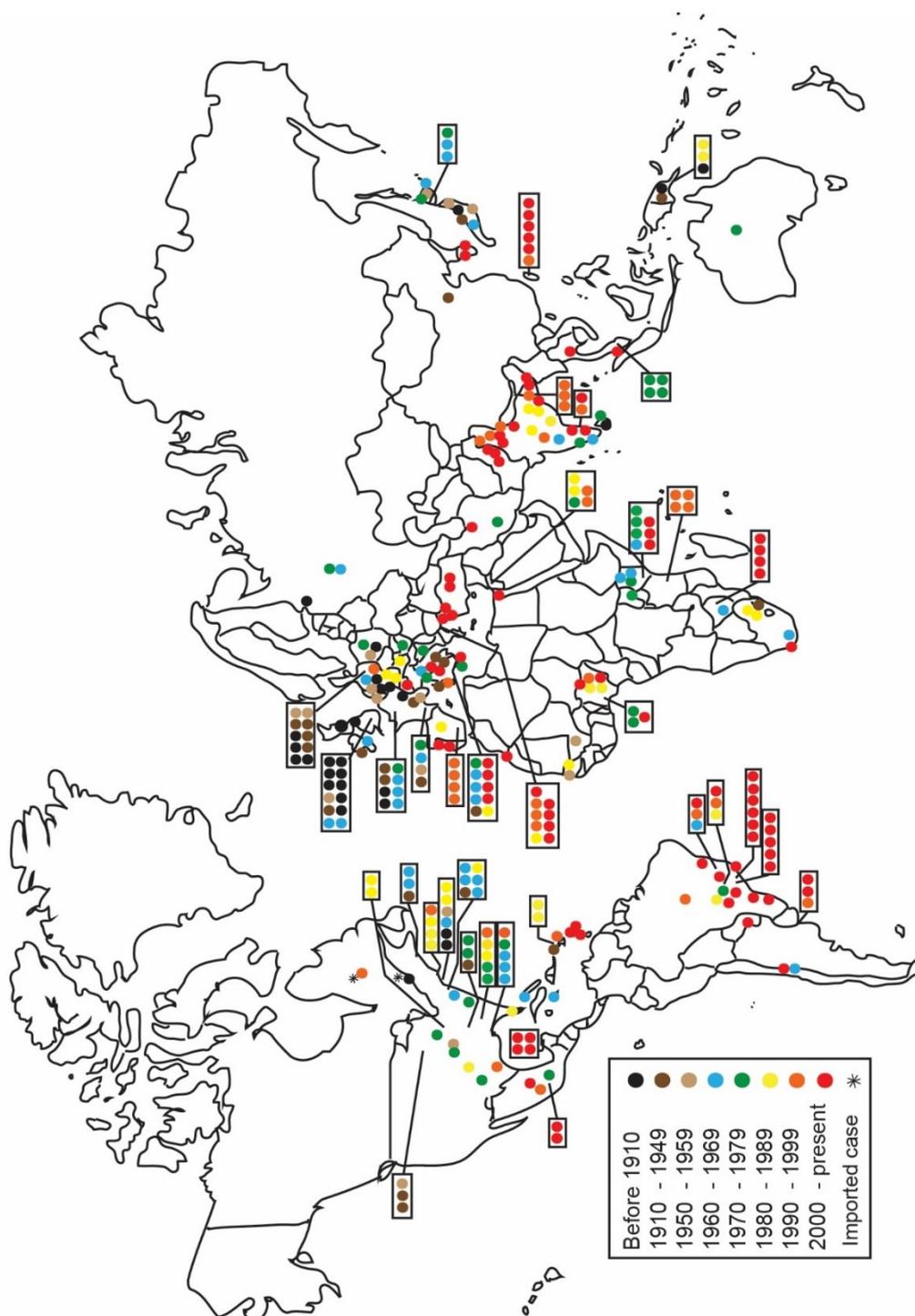


Figure 2-1 Global distribution of published reports of CTVT. Locations in which naturally occurring CTVT cases are reported in the published literature are indicated on the map, classified by date of report. Bibliographical information for each case is found in Additional file 2-4. The two reports in Canada (French 1906, Mikaelian et al. 1998) were specified as imported cases from abroad and are marked with *. Map was created using Adobe Illustrator.

In the second half of the 20th century, the literature provided evidence of declining CTVT prevalence in some locations - two reports published in the 1950s stated that CTVT occurs “less commonly in London dogs” (1954) (Cotchin 1954) and that there has been a “reduction in incidence” of CTVT in New York City (1951) (Bloom et al. 1951). Indeed, my literature review suggested a steady decline in CTVT in the United Kingdom during the twentieth century, as shown by the comments in articles published between 1810 and 1969 (see Table 2-1).

Eighteen of the published reports provided a numerical value for the prevalence of naturally occurring CTVT in their study (see Additional file 2-2). The reported prevalence of CTVT in affected populations ranged from 1% or less (Jamaica, 1968 (Thorburn et al. 1968); Kenya, 1972 (Rottcher 1972); Bangladesh, 2010 (Tarafder and Samad 2010)) to almost 20% (Papua New Guinea, 1985 (Hamir 1985) and 1986 (Hamir 1986); Mexico 2007 (Ortega-Pacheco et al. 2007) and 2010 (Cruz 2010)).

Year	Town	Author	Comment
1810	UK (London)	D. P. Blaine	two parts only [of the anatomy] are subject to a cancerous affection [in dogs] [mammary tumour and canine transmissible venereal tumour]
1897	UK (London)	G. B. Smith	Dog A served 12 bitches, eleven of which became infected.
1898	UK (London)	G. B. Smith	Since the beginning of the year 1896 we have had under observation a series of contagious tumours on the genital organs of dogs
1900	UK (London)	F. Hobday	In stud dogs their [CTVT tumours'] presence is to be regarded with great suspicion
1902	UK (Leeds)	C. Powell White	I have had the opportunity of examining two similar cases, which present several points of interest. The animals affected were valuable pure-bred bulldogs.
1905	UK (London)	F. Hobday	Inside the prepuce and on the mucous surface of the penis one meets with another variety, the infective venereal tumour... It appears to be most commonly met with in the bulldog variety, and the author has also seen it in St. Bernards, terriers, poodles and pugs, although there is no reason why it should be confined to these breeds.
1906	UK (London)	F. Hobday	In the vagina of the bitch one frequently meets with a species of ulcerating contagious venereal tumour which is communicable to the male when the animals are used for stud purposes
1907	UK (Edinburgh, London)	H. Wade	The disease known as infective sarcoma is one which occurs in dogs. It is not uncommon... In this country, on the continent of Europe, and in the United States of America it is frequently met with.
1954	UK (London)	E. Cotchin	[tumours occur] less commonly in London dogs
1969	UK (Liverpool)	J. C. Howell	Since 1959 we have diagnosed 13 cases of canine T.V.T.. Eleven of the dogs were owned... Of the other two, one came from Stattford. In only one case was there a history of the dog having lived abroad and this was an animal that had been in Ghana some four to five years before symptoms associated with T.V.T. developed.
1969	UK (London)	R. C. White	I have not seen this condition in this country, but I have examined a number of cases while I was abroad.

Table 2-1 Historical reports confirming presence of CTVT in the United Kingdom in the 19th and 20th centuries. The "Comment" field reports relevant phrases from the cited reports.

2.3.2 Current worldwide CTVT distribution

The distributed internet-based questionnaire consisted of 18 predominantly multiple-choice questions regarding CTVT prevalence, features of CTVT-infected dogs, and conditions of the local dog population (see section 2.2.1 and Additional file 2-1). Respondents were asked to estimate the prevalence of CTVT in their local area. I received a total of 645 completed questionnaires from 109 countries.

The average estimated CTVT prevalence reported for each country from which a minimum of three responses were received is displayed in Figure 2-2 (see also Additional file 2-3). In addition, a separate map displays each individual respondent's estimate of CTVT prevalence (see Figure 2-3 for map of the world, see Figure 2-4 for higher magnification map of Europe only). The data indicate that CTVT is estimated to occur at between one and ten percent prevalence in dogs in most countries in South and Central America as well as in parts of Africa and Asia, consistent with estimates indicated by published literature (see section 2.3.1). The average reported prevalence by continent is shown in Figure 2-5. The highest estimated CTVT prevalence that we recorded was in Belize, where the average CTVT prevalence (calculated from 6 responses) was estimated to be between 10 and 20 percent. Several countries (Canada, the Czech Republic, Finland, the Netherlands, New Zealand, Sweden, Switzerland and the United Kingdom) were consistently reported by all respondents from that country to be free of endemic CTVT; in these countries, the only CTVT cases were specifically reported to be imported from abroad (see also Figure 2-2, Figure 2-3 and Figure 2-4). CTVT was reported as absent from many regions of the United States and Australia, however it was reported to be present in remote indigenous communities, including Indian reservations in Arizona and North Dakota, as well as in Australian Aboriginal communities in the Northern Territory and Western Australia (see Figure 2-3). There was also geographical variation in estimated CTVT prevalence in Europe; the disease was reported to be absent except for occasional imported cases in many countries of Northern and Western Europe, but was estimated to be present at less than ten percent prevalence in countries in Southern and Eastern Europe (see Figure 2-4).

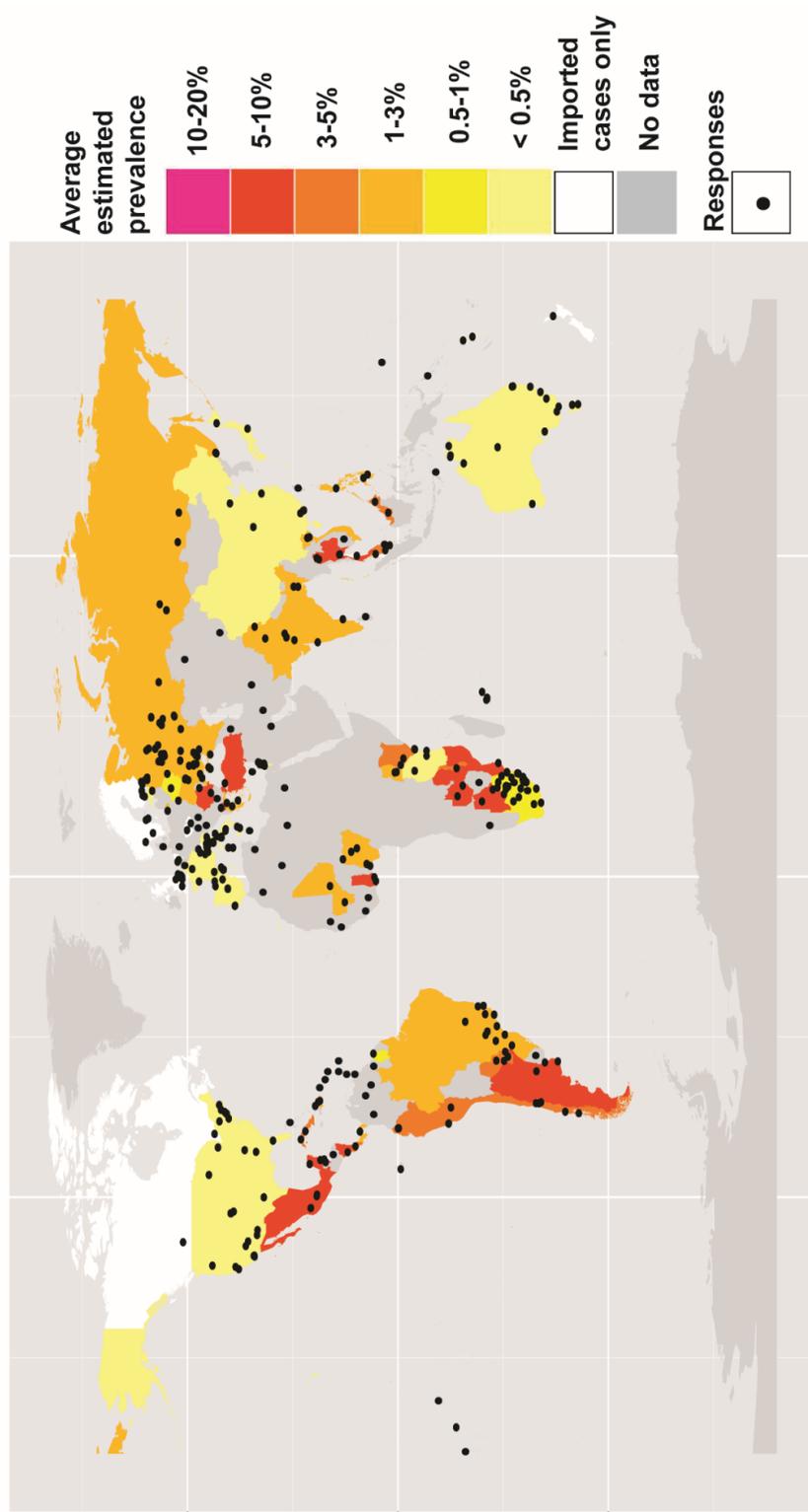


Figure 2-2 Worldwide CTVT distribution and prevalence. CTVT worldwide distribution by country. The colour of each country represents the average of all CTVT prevalence estimates obtained from that country. The location of each response is indicated with a black dot. Countries from which only one or two responses were received are coloured in grey, but location of a response is nevertheless indicated with a black dot. Complete datasets are included in Additional file 5 and Figure 2.3. Map was created using the maptools and ggmap packages, implemented in R, with assistance from Dr Simon Frost (Department of Veterinary Medicine, University of Cambridge) (R Core Team 2013).

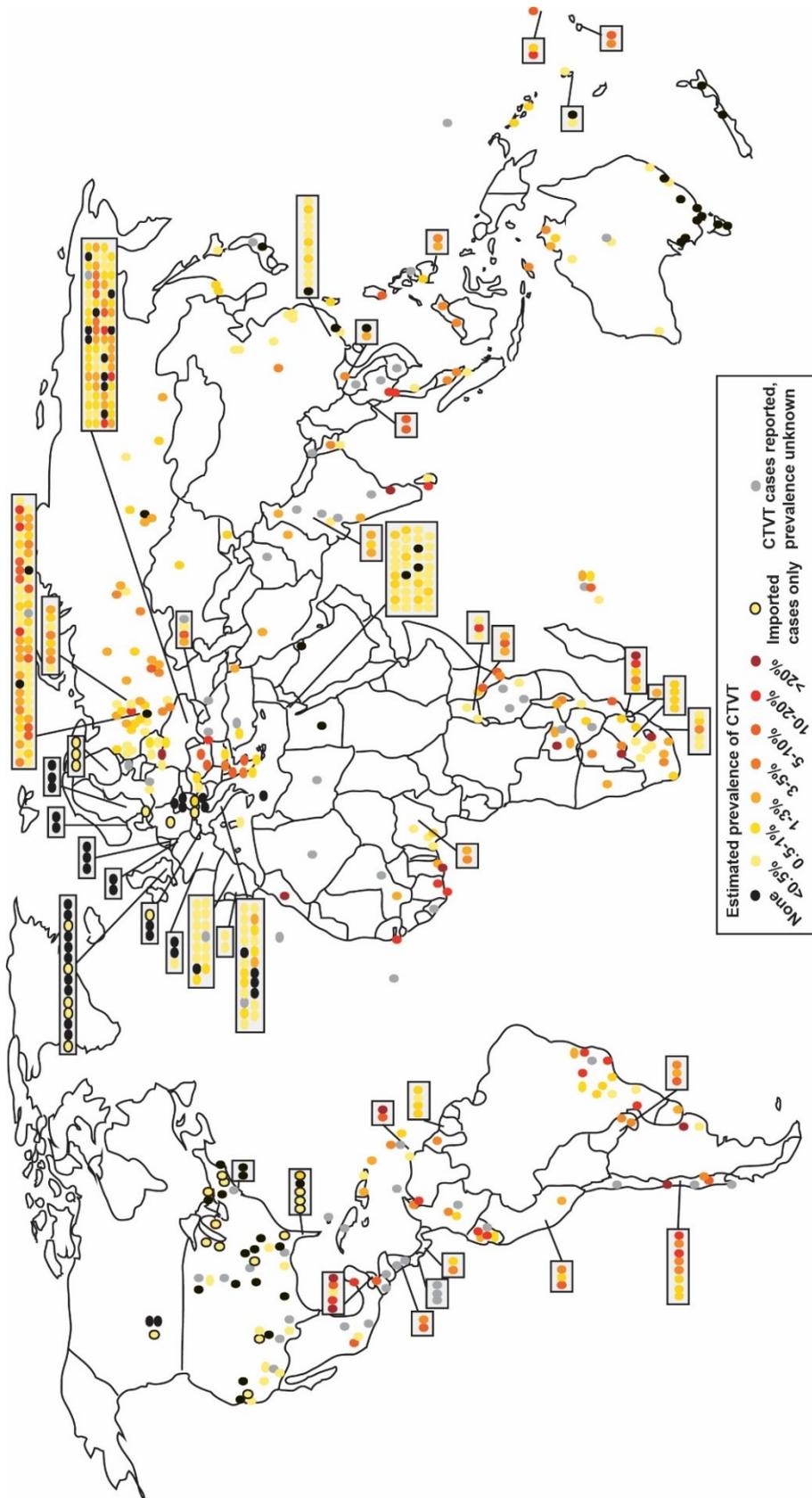


Figure 2-3 All responses to the questionnaire on worldwide CTVT distribution and prevalence. Map indicating CTVT prevalence estimated by each respondent. Each response is represented by a single coloured dot. Map was created using Adobe Illustrator.

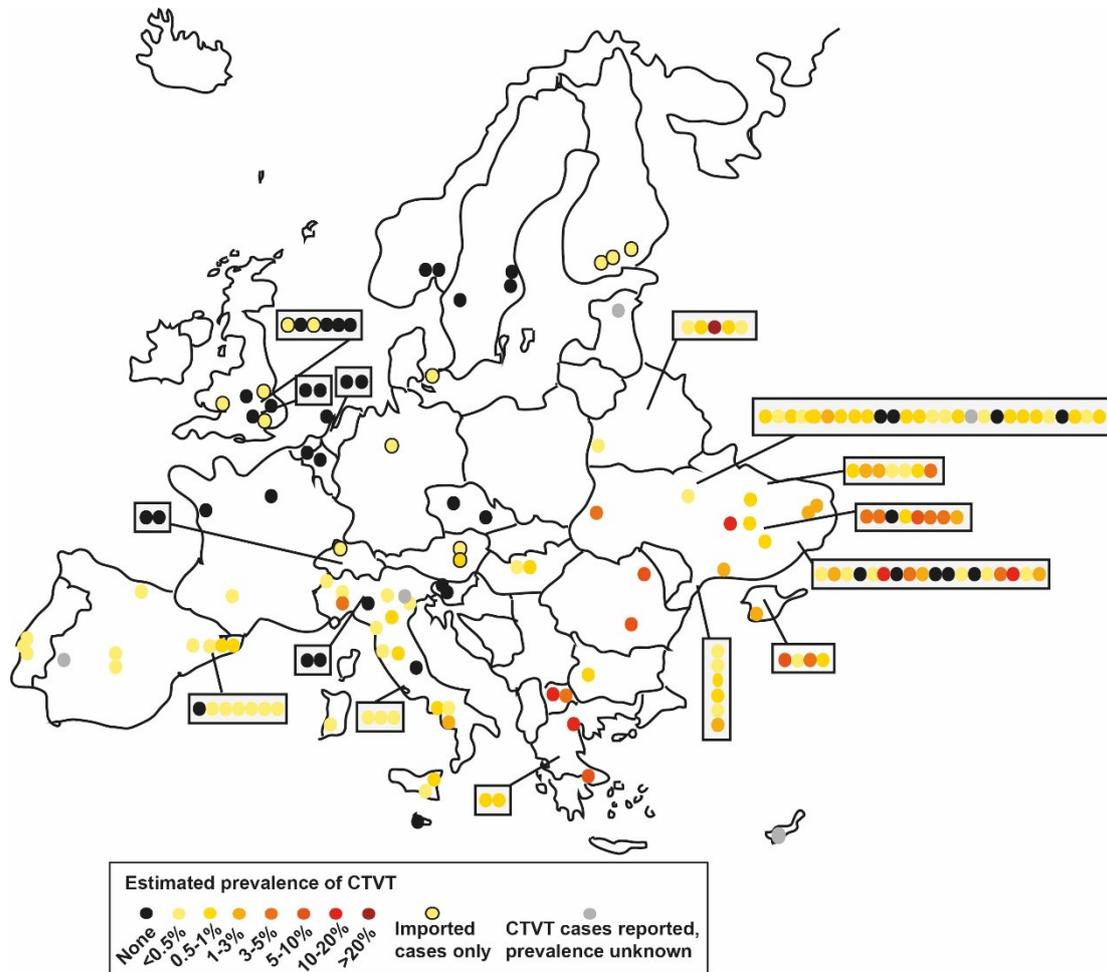


Figure 2-4 All responses to the questionnaire on worldwide CTVT distribution and prevalence. Higher magnification map showing detailed distribution of CTVT prevalence estimates in Europe. Each response is represented by a single coloured dot. Map was created using Adobe Illustrator.

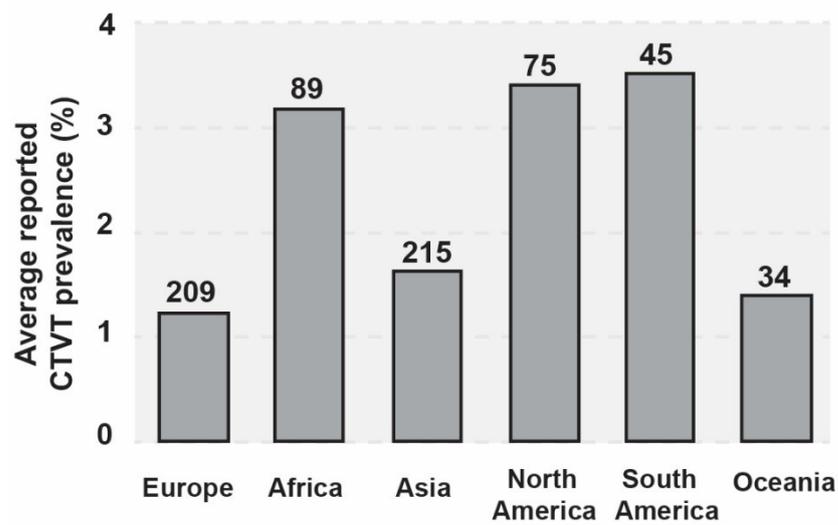


Figure 2-5 Average reported CTVT prevalence by continent. The number of responses estimating prevalence in each continent is shown above each bar.

2.3.3 Status of dogs infected with CTVT

The questionnaire respondents were asked to report the most likely gender and health status of animals with CTVT. Despite previous suggestions that CTVT is more common in males (Osipov and Golubeva 1976, Brown et al. 1980) or females (Ajello 1939, Ajello 1980, Gandotra 1993, Singh et al. 1996, Sobral 1998), responses to the questionnaire indicated that there is no universal detectable gender bias for CTVT infection, with 144 respondents reporting that CTVT was more common in males, 146 reporting that it was more common in females and 168 claiming that it was equally common in both sexes (chi-squared test for no gender difference, $p=0.907$) (Figure 2-6A). The majority of respondents (459 out of 637) claimed that dogs with CTVT were otherwise healthy rather than infected with parasites or otherwise diseased, thin or emaciated, or carrying injuries or bite marks (Figure 2-6B).

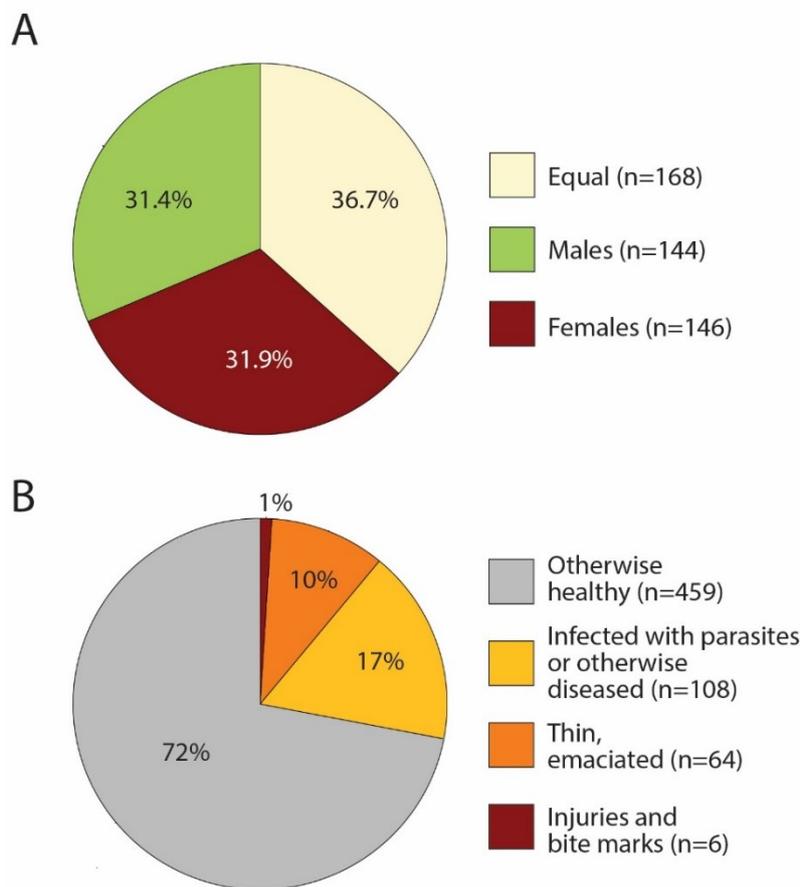


Figure 2-6 Status of dogs with CTVT. (A) Gender of dogs infected with CTVT. Respondents were asked if they observed CTVT “more commonly in males”, “more commonly in females” or “equally in males and females”. The numbers refer to the number of respondents choosing each option. (B) Health condition of dogs infected with CTVT. Respondents were asked to report on the condition of the majority of CTVT-infected dogs by selecting one of the four categories shown. The numbers refer to the number of respondents choosing each option.

I next investigated the association of spaying/neutering with CTVT presence by asking respondents whether the majority of dogs in their country are spayed/neutered or entire. The responses showed that higher estimated CTVT prevalence coincides with higher percentage of respondents claiming that the majority of dogs are entire rather than spayed or neutered (Figure 2-7).

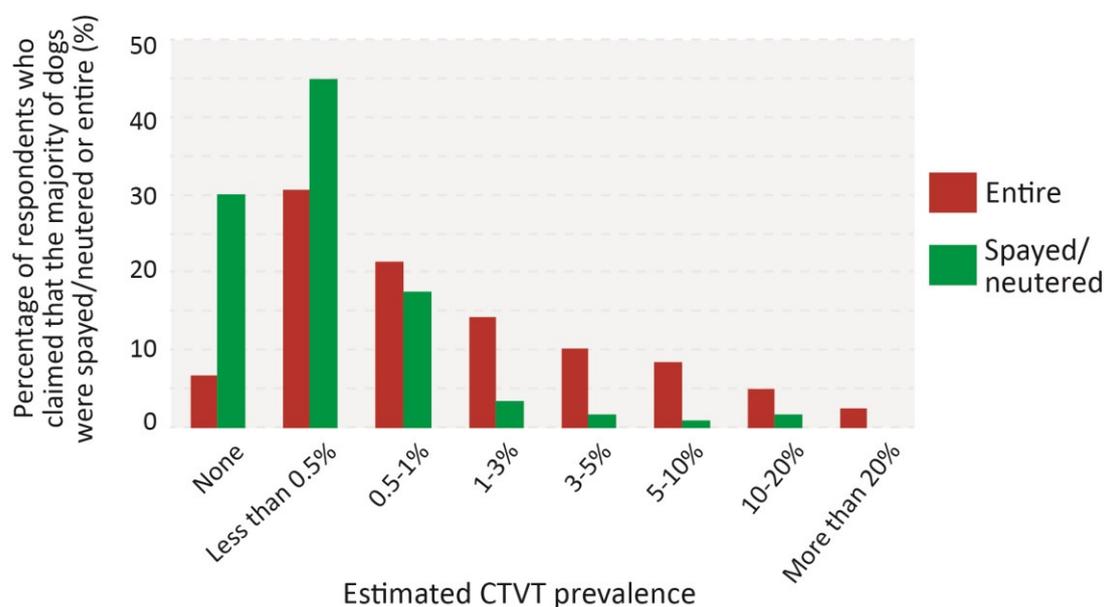


Figure 2-7 Relationship between spaying/neutering and CTVT prevalence. CTVT prevalence estimates from respondents who claimed that the majority of dogs in their area were spayed/neutered (total of 120 respondents) or entire (total of 340 respondents) are coloured in red and green respectively. Percentage of respondents refers to the proportion of respondents choosing each option.

2.3.4 Natural transmission of CTVT to wild canids

Despite previous reports of successful experimental transmission of CTVT cells into wild canids, including wolves, coyotes and red foxes (Sticker 1906a, Wade 1908, Dungern 1912, Cockrill and Beasley 1979, Cohen 1985), no naturally occurring CTVT case has been previously reported in a wild canid. I further investigated this by asking questionnaire respondents if they had observed CTVT in a wild canid (see Additional file 2-1). The responses did not reveal any confirmed reports of CTVT in wild canid populations.

2.3.5 Factors influencing CTVT prevalence

In order to analyse whether socio-economic status may influence the prevalence of CTVT, we stratified countries based on income economy (World Bank 2013). We found lower estimated prevalence of CTVT in countries with high income economies, compared to those with low, lower-middle and upper-middle income economies, which had similar distributions of estimated CTVT prevalence (Figure 2-8A). The proportion of countries with reported presence of free-roaming dogs was also associated with income economy category (Figure 2-8B), presenting the possibility that differences in presence of free-roaming dogs may explain the link between CTVT prevalence and national development status.

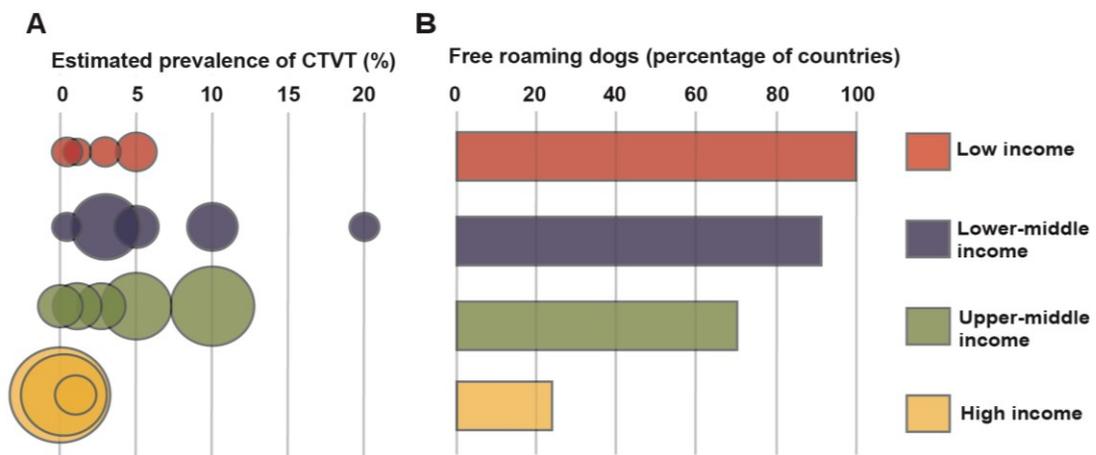


Figure 2-8 Factors associated with CTVT prevalence. (A) Relationship between CTVT prevalence and socio-economic status defined as low/lower-middle/upper-middle/high income economies (World Bank, Country and Lending Groups, 2013 (World Bank 2013)). The size of the dots represents the number of countries with estimated average CTVT prevalence within each interval. CTVT prevalence values represent the upper limit of each categorical interval. (B) Relationship between socio-economic status defined by division into low/lower-middle/upper-middle/high income economies and presence of free-roaming dogs, as reported by respondents to the questionnaire. "Percentage of countries" refers to the proportion of countries within each income category in which the presence of free-roaming dogs was reported by the majority of respondents.

Additionally, I observed a weak negative correlation between average reported CTVT prevalence and socio-economic status, determined by Gross Domestic Product (GDP) per capita value for each country (see Figure 2-9A, Pearson's correlation two tailed significance test between CTVT prevalence and GDP values, $r=-0.504$, $p=4.52 \times 10^{-5}$). Moreover, I detected a weak negative correlation between average reported CTVT prevalence and distance from the equator, as measured by distance from the equator to the capital city of each country (see Figure 2-9B, Pearson's correlation two tailed significance test between

CTVT prevalence and distance from the equator, $r=-0.416$, $p=8.39 \times 10^{-4}$). It is possible, however, that these correlations, as well as a previously identified correlation between CTVT prevalence and distance from the equator within the United States (Hayes et al. 1983), may be explained by the presence of free-roaming dogs.

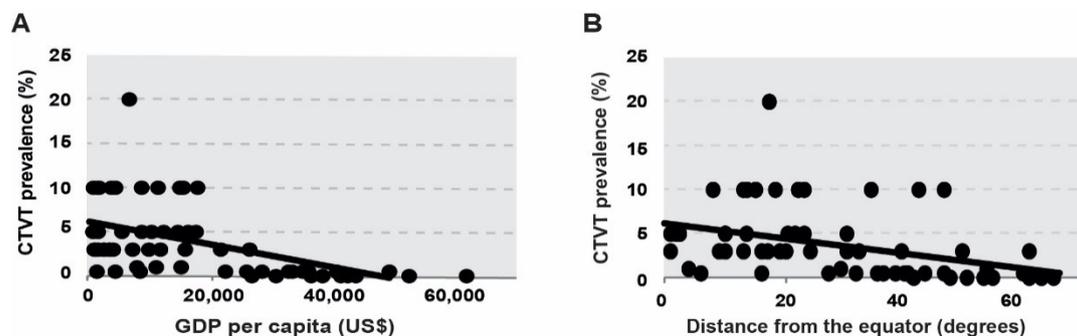


Figure 2-9 Socio-economic and climatic factors associated with CTVT prevalence. (A) Association between GDP per capita and CTVT prevalence. Pearson's correlation two tailed test, $r=-0.504$, $p=4.52 \times 10^{-5}$ (B) Association between distance from the equator and CTVT prevalence. Pearson's correlation two tailed test, $r=-0.416$, $p=8.39 \times 10^{-4}$.

2.3.6 Metastasis and treatment of CTVT

CTVT metastasis has been previously noted in individual case reports (Feldman 1929, Cella 1939, Rust 1949, Belkin 1959, Barron et al. 1963, Prier and Johnson 1964, Sastry et al. 1965, Higgins 1966, Adams and Slaughter 1970, Manning and Martin 1970, Rottcher 1972, Kimeto and Mugeru 1974, Osipov and Golubeva 1976, Idowu 1977, Ajello 1980, Yang 1987, Miller 1990, Moulton 1990, Gandotra 1993, Ferreira et al. 2000, Pereira et al. 2000, Gurel 2002, Abuom 2006, Levy et al. 2006, Park et al. 2006, Batista et al. 2007, Bastan 2008, Mylonakis et al. 2008, Varughese et al. 2012, Chikweto 2013), but no systematic study has summarised frequency of CTVT metastasis around the world. I asked questionnaire respondents to estimate the proportion of CTVT cases in which they observe metastasis. Of the six categorical intervals available to respondents ("none", "0-5%", "5-10%", "10-15%", "15-20%", "more than 20%"), the majority of respondents estimated that metastasis occurs in 0-5 percent of CTVT cases (Figure 2-10A), consistent with previously published estimates (Karlson and Mann 1952, Brown et al. 1980). Additionally, we recorded the most commonly observed sites of metastasis reported by questionnaire respondents (Figure 2-10B).

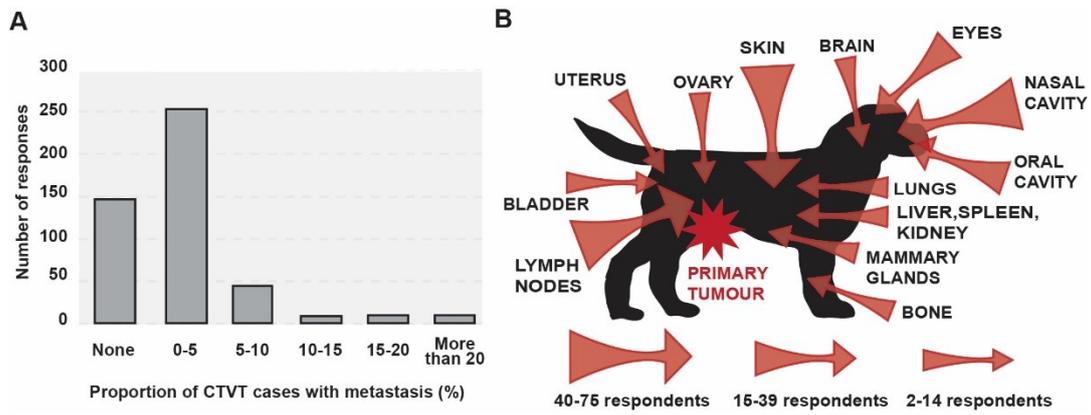


Figure 2-10 CTVT metastasis. (A) Estimated proportion of CTVT cases with metastasis. Number of responses refers to the number of respondents choosing each option. (B) Sites of CTVT metastasis reported by respondents. The size of the arrow is proportional to the number of respondents who reported observations of CTVT metastasis at this site.

Vincristine chemotherapy is known to be a very effective treatment for CTVT (Brown et al. 1980, Calvert et al. 1982, Boscos 1988, Singh et al. 1996, Nak et al. 2005, Said et al. 2009). We asked respondents to report the type and effectiveness of treatments they typically administer for CTVT (Figure 2-11A). The majority of respondents (373 out of 454 respondents, 82.2%) reported that they use either vincristine alone or vincristine in combination with surgery, doxorubicine or radiotherapy. A proportion of respondents (10.9%, 50) reported that they used surgery alone or other non-vincristine treatments (4%, 18). Thirteen (2.9%) respondents stated that the only option is euthanasia. Those who used vincristine treatment for CTVT reported that 80 to 100 percent of tumours usually went into complete remission after treatment (Figure 2-11B), however the number of vincristine doses claimed to be required for complete remission varied between respondents. This contrasted with the poor reported effectiveness of surgery alone or other non-vincristine treatments (Figure 2-11B).

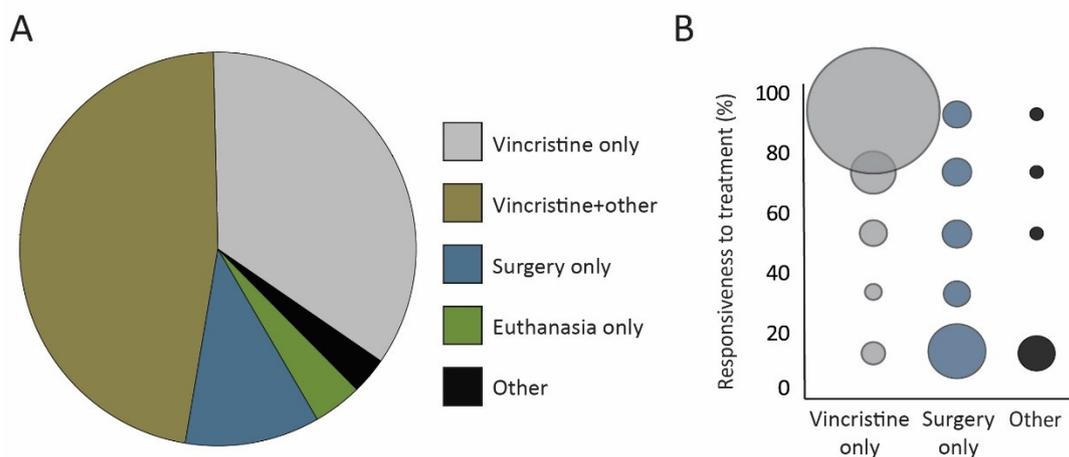


Figure 2-11 CTVT treatment. (A) CTVT treatments reported by respondents. Total number of responses for this question was 454. (B) CTVT response to treatment. "Response to treatment" refers to the proportion of CTVT cases that respondents estimated went into complete remission after the selected treatment. "Other" treatments included aloe vera treatment and electrocautery, but did not include chemotherapy. Circle size represents the number of respondents who estimated effectiveness of treatment within each interval.

2.4 Discussion

As part of the work described in this Chapter, I contacted over 600 veterinarians and animal workers in order to source worldwide CTVT prevalence estimates. This study has enabled me to conclude that CTVT is endemic in at least 90 countries worldwide across all inhabited continents. Additionally, our work has highlighted factors that continue to modify the prevalence of CTVT around the world – free roaming dogs were indicated as the reservoir for the disease and spaying/neutering was found to be associated with lower CTVT prevalence.

2.4.1 Disappearance of CTVT from the United Kingdom

During this study, we have documented the decline and disappearance of CTVT from the United Kingdom during the twentieth century (see summary in Figure 2-12 and Table 2-1). The eradication of CTVT from the United Kingdom may be due to the introduction of a series of dog management laws throughout the nineteenth and twentieth centuries (see summary in Figure 2-12). The Dogs Act, 1871 (National Archives 2017a) imposed civil responsibility on dog owners and stated that dogs must be “under proper control” and “stray dogs may be detained and sold or destroyed”. This was followed by the Dogs Act 1906 (National Archives 2017b), amended in 1928 and 1938, which introduced a requirement to report stray dogs to the police. It is striking that the eradication of CTVT, once a common canine pathogen in the United Kingdom (see Table 2-1), appears to have occurred as an unintentional result of human intervention.

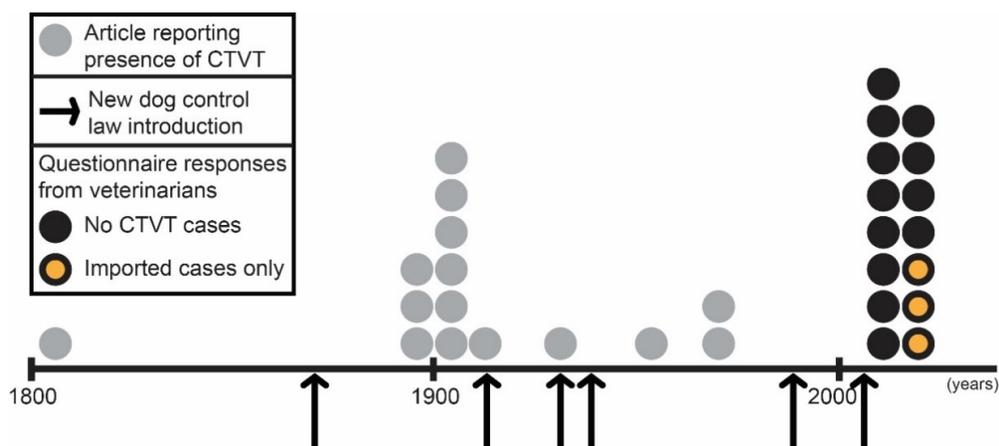


Figure 2-12 Disappearance of CTVT from the United Kingdom during the twentieth century. Timeline showing the declining number of historical reports of CTVT in the United Kingdom, coinciding with the introduction of dog control laws. Data from the questionnaire (indicated with black and yellow dots) were used to confirm current absence (except for occasional imported cases) of CTVT from the United Kingdom. See also Table 2-1.

2.4.2 Implications of the study

Apart from mapping the worldwide distribution and prevalence of CTVT, my study indicated broader implications for effective control of CTVT and additionally highlighted further important questions relevant to understanding the biology of CTVT.

The importance of dog management and spaying/neutering in CTVT control was highlighted by a respondent from Koh PhaNgan Island, Thailand, where breeding control and sterilisation campaigns have almost eradicated CTVT from the island since the commencement of the sterilisation project in 2001. Several respondents, however, when asked to comment on any unusual cases of CTVT that they had seen, commented that they had observed CTVT in dogs years after spay or neuter surgery. This suggests that either the latent period for CTVT development, which previous anecdotal reports have suggested may last for weeks or months (Smith and Washbourn 1898b, Ajello 1939, Locke et al. 1975), can sometimes last years, or, alternatively, that spaying and neutering does not always protect dogs from CTVT. Furthermore, non-coital modes of CTVT transmission, including biting, licking or sniffing, may also contribute to CTVT infection of spayed/neutered dogs (Ndiritu et al. 1977, Weir et al. 1978, Perez et al. 1994, Guedes et al. 1996, Albanese et al. 2002, Chikweto 2013).

Despite the widespread presence of CTVT in dog populations worldwide, the results of our survey suggest that, at least in populations for which we have obtained data, its prevalence rarely rises above 10% (Figure 2-2). This contrasts with epidemiological patterns observed for the only other known naturally occurring transmissible cancer, the Tasmanian devil facial tumour disease (DFTD). Prevalence of DFTD usually rises above 50% in affected Tasmanian devil populations and the disease usually triggers a rapid population decline (reviewed in (Murchison 2008)). Given that the mixed mating system of dogs might be likely to promote widespread exposure to the disease (Pal 2011), this pattern raises the possibility that only a proportion of the dogs may be susceptible to CTVT and, possibly, as one early report of CTVT proposed, that “some animals are naturally refractory” to infection (French 1906). Future studies of CTVT exposure and susceptibility in free-roaming dog populations may reveal further insights into the biological basis of this interesting observation.

Genetic studies indicate that the global spread of CTVT has occurred relatively recently in the history of the lineage, probably within the last 500 years (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014). Although we do not know the location in which CTVT first emerged, our study has highlighted the remarkable efficiency with which CTVT has colonised its global host population. We obtained evidence for CTVT’s presence in some of the world’s

most isolated communities and islands, including the Solomon Islands, Samoa, American Samoa, Fiji, Reunion, Mauritius and several islands in Micronesia. In contrast, New Zealand is free of CTVT, likely due to its rigorous import quarantine rules (New Zealand Government 2013). Together, these findings highlight the mobility of dogs, which, together with humans, have travelled rapidly and extensively around the globe. Further analysis and discussion of how CTVT has spread around the world will follow further on in this Thesis, particularly in Chapter 4.

2.4.3 Challenges and limitations of the study

The global CTVT prevalence data reported in this study are individual estimates of local CTVT prevalence and are thus subject to errors introduced by variation in methodologies used by respondents to estimate CTVT prevalence. The CTVT prevalence figures reported in this Chapter were estimated by questionnaire respondents based on one of the following criteria:

- (1)** An estimate of the number of CTVT patients as a proportion of total canine patients in general clinical practice
- (2)** An estimate of the total number of CTVT diagnostic samples as a proportion of the total number of canine diagnostic samples passing through a pathology laboratory
- (3)** A personal estimate based on previous veterinary experience and discussions with veterinary colleagues
- (4)** The total number of CTVT cases in populations of dogs participating in spay/neuter campaigns.

Difficulties in prevalence estimation may have been further confounded by a combination of absent, inconsistent or incomplete record keeping, personal biases as well as systematic biases introduced where the sampling population was not representative of the population as a whole. Furthermore, variation in CTVT prevalence due to seasonal, demographic or local geographical factors may not have been captured by my crowdsourcing approach. In order to minimise the effects of estimate biases on my analysis, I only included data from countries from which I received at least three responses in Figure 2-2 and Figure 2-8. Despite the limitations associated with estimating global CTVT prevalence, my large sample size supports the conclusion that a strikingly large proportion of the global dog population harbours CTVT infection at a prevalence of between 0.5 and 10 percent. Future studies will however, be important to further validate global variation in CTVT prevalence.

2.5 Conclusion

The study presented in this Chapter has provided information on historical and current CTVT global distribution and prevalence and has illuminated a number of risk factors that may influence CTVT prevalence, including presence of free-roaming dogs, dog spay/neuter practices and enforcement of dog control laws. In addition to providing insight into the global spread of a unique type of pathogen, this study may assist policy-makers and veterinarians in the development of measures to more effectively control and reduce CTVT prevalence and prevent further spread of the disease. Importantly, through performing this work, I have been able to build a supportive network of collaborators (see Table of collaborators, page v) who contributed CTVT samples crucial to my future studies (see Chapter 3).

3

Canine transmissible venereal tumour sample collection and validation

Summary

Although CTVT arose around 11,000 years ago, it appears to have spread around the world only in the last few hundred years (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014), during which it has colonised the global canine population with remarkable efficiency. The main goal of this Thesis was to analyse the genetic and phenotypic diversity that arose as CTVT lineages diverged. In order to investigate this diversity, a crucial step was to collect CTVT samples from different locations around the world. In this Chapter, I describe the sample collection process together with the validation pipeline that was designed to confirm CTVT diagnosis, as well as to estimate tumour purity in confirmed CTVT biopsies. This work resulted in collection of 1816 tumour samples for genetic analysis that were collected from 1498 dogs from 53 countries. Additionally, for 929 dogs, a tumour sample for histology analysis was collected. A subset of these tumours was confirmed to amplify a CTVT-specific *LINE-MYC* rearrangement, with 793 dogs diagnostically confirmed to carry CTVT tumours. The set of samples was biased towards female dogs, possibly reflecting bias towards females in the spay/neuter programmes through which many samples were collected. As expected, most dogs in the sample set were entire adults. More than 50% of sampled dogs were of mixed breed, reflecting the finding that free-roaming dogs are a reservoir for the disease. Overall, the work presented in this Chapter enabled me to collate a robust set of CTVT samples that were subsequently used for DNA sequencing and histopathology analysis.

3.1 Introduction

CTVT is a disease found in dog populations around the world (see Chapter 2). Although the cancer arose around 11,000 years ago, genetic studies indicate that its global spread happened relatively recently, probably within the last few hundred years (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014). My interest in this Thesis was to understand the genetic and phenotypic diversity of CTVT around the world. In order to analyse the genetic and phenotypic diversity of this extraordinary cell lineage, I sought to collect and validate CTVT samples from all six inhabited continents where the disease was reported to be endemic (see Chapter 2).

3.1.1 CTVT sample collection

The overall aim of this Chapter was to collect samples of globally distributed CTVT tumours for two types of downstream analyses (see Chapter 4 and Chapter 5 for sample analysis): (1) genetic analysis and (2) phenotypic analysis. Genetic analysis involved extraction of genomic DNA that was subsequently investigated using DNA sequencing methods discussed in Chapter 4 and Chapter 5. Phenotypic analysis included comparison of histopathological and morphological features of the same set of CTVT tumours. Overall, my aim was to combine the information obtained from both genetic and phenotypic analysis with the goal of explaining any phenotypic differences between tumours from different parts of the world by the underlying genetic changes that took place since these tumour lineages diverged.

Considering the downstream analyses to be performed, the following samples and metadata were to be collected for each case:

(1) Genetic analysis

- Tumour biopsy
- Matched host biopsy

(2) Phenotypic analysis

- Tumour biopsy in 10% neutral buffered formalin
- Photograph of the tumour

(3) Metadata

- Gender of the dog
- Age of the dog
- Breed of the dog
- Ownership status
- Entire/desexed status

In order to analyse the global genetic and phenotypic diversity of CTVT, I aimed to collect samples from at least 500 CTVT cases from a diverse set of countries located on all six inhabited continents.

CTVT is a disease affecting dogs around the world, and is particularly common in low-income countries (see Chapter 2). It was therefore important to approach a diverse set of veterinary workers, in order to reach local communities and places where CTVT is most common. Apart from contacting local private veterinarians, I sought to approach veterinarians in shelters/animal welfare organisations and veterinary organisations performing spay/neuter campaigns. Many of these organisations were working in very low resource settings, which may have affected the quality of tumour samples collected, thus requiring a rigorous validation process to take place before samples were used for DNA analysis.

3.1.2 CTVT validation

CTVT is a transmissible cancer with a characteristic appearance and is commonly located around the genital regions (see Figure 1-3). Nevertheless, clinical diagnosis of CTVT in field conditions may be challenging for cases with atypical presentation or extra-genital location. This poses risk that some samples collected by veterinarians, even though grossly diagnosed as CTVT, may be misdiagnosed.

As discussed in Chapter 1, section 1.2.3.1, diagnosis of CTVT can be performed using the following three methods: (1) cytology, (2) histopathology or (3) presence of *LINE-MYC* rearrangement (Duncan and Prasse 1979, Cohen 1985, Katzir et al. 1985, Katzir et al. 1987, Murgia et al. 2006, Rebbeck et al. 2009) (see Figure 1-5, Figure 1-6 and Figure 3-1A,). Both cytology and histopathology rely on expert interpretation, but presence of *LINE-MYC* rearrangement can be confirmed using molecular biology methods.

CTVT cells grow with the support of a microenvironment (see Chapter 1, section 1.2.2.2), which is formed by stromal and immune cells from the host. As was previously reported, the

immune cell infiltrate and the amount of host supporting stroma varies between CTVT tumours depending on their stage of growth (Cohen 1985). Amount of host cells can therefore vary between tumours as well as between distinct biopsies from an individual tumour. Host contamination can have a major effect on downstream analyses. Importantly, contaminating DNA from the host can affect interpretation of DNA sequencing data, as further discussed in Chapter 4 and Chapter 5.

To overcome the complications in genetic analysis that may arise from host contamination, we used the following two strategies: (1) collect a matched host biopsy that will also be used for DNA sequencing and genetic analysis (see section 3.1.1) and (2) estimate the proportion of host contamination in each CTVT tumour sample.

Estimation of proportion of host contamination was based on presence of *CDKN2A* (cyclin-dependent kinase inhibitor 2A), a gene which is homozygously deleted (i.e. loss of both copies) in CTVT (Murchison et al. 2014) (see Figure 3-1B), but present in two copies in host tissue. Any amplification of *CDKN2A* should therefore be exclusively indicating presence of host tissue.

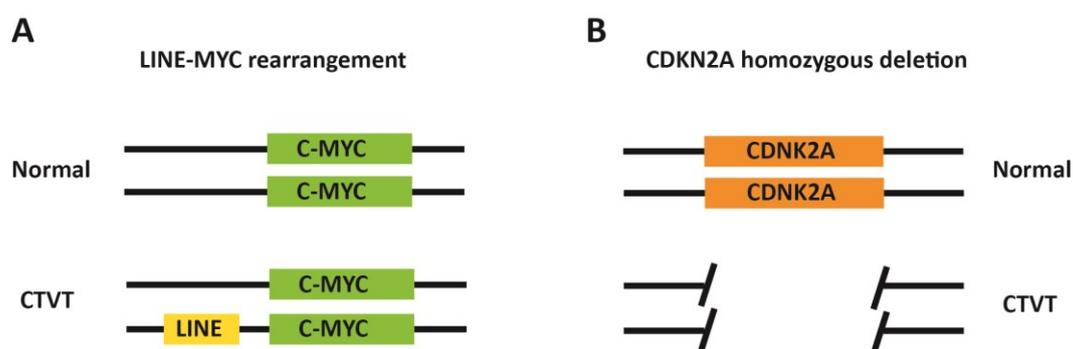


Figure 3-1 Schematic diagram of (A) LINE-MYC rearrangement and (B) CDKN2A homozygous deletion in CTVT. CTVT copy number at the LINE-MYC locus is presented as diploid; however, true copy number at this locus in CTVT is variable.

3.1.3 Goals of this Chapter

The overall aims of this Chapter are summarised below and in Figure 3-2:

- (1) To collect at least 500 CTVT samples from countries on all six inhabited continents
- (2) To optimise DNA extraction and quantification protocols and subsequently extract and quantify DNA from all samples collected
- (3) To design a diagnostic strategy with the goals to:
 - Confirm that samples collected are CTVT tumours
 - Estimate the proportion of host cells in each sample

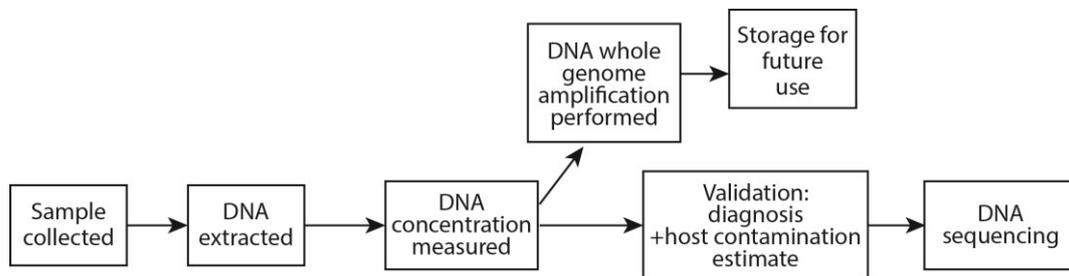


Figure 3-2 Summary of goals and workflow for Chapter 3.

3.2 Materials and Methods

3.2.1 Sample collection

The project and sample collection procedure were approved by the Department of Veterinary Medicine, University of Cambridge, Ethics and Welfare Committee (reference numbers CR174 and CR197). A consent form was completed by the owner or responsible veterinarian for each case where possible (see Additional file 3-1).

3.2.1.1 *Collaborators performing sample collection*

Collaborators involved in sample collection included veterinarians in private clinical practices, veterinarians in a shelter/animal welfare organisations, veterinarians performing short-term spay/neuter campaigns, veterinarians based at veterinary hospitals, veterinary pathologists and veterinary nurses/technicians. The contacts were built up through the CTVT worldwide distribution questionnaire, as discussed in Chapter 2, as well as through internet searches, contacts and colleagues. In addition, I was able to take part in sample collection on field trips to Ecuador and Belize, and I also directly assisted in sample collection from a dog with CTVT in the UK (imported from Romania).

3.2.1.2 *Sample collection kit*

Each collaborator who agreed to collect CTVT samples for my project was sent the following paperwork, which I partly designed myself and partly edited based on a template from Dr Elizabeth Murchison:

- (1) CTVT sample collection information sheet (Additional file 3-2)
- (2) CTVT case metadata collection sheet (Additional file 3-3) - asked to fill in for each case

A sample collection kit was sent by a courier (FedEx) to each veterinarian or animal health worker participating in the project. One sample collection kit included 50ml of RNAlater (Invitrogen), 40 cryotubes (2ml, VWR), plastic pipette (Fisher Scientific) and a marker pen. A hard copy of a sample collection leaflet (see Additional file 3-4) was included in each sample collection kit. A photo of a sample collection kit is shown in Figure 3-3.

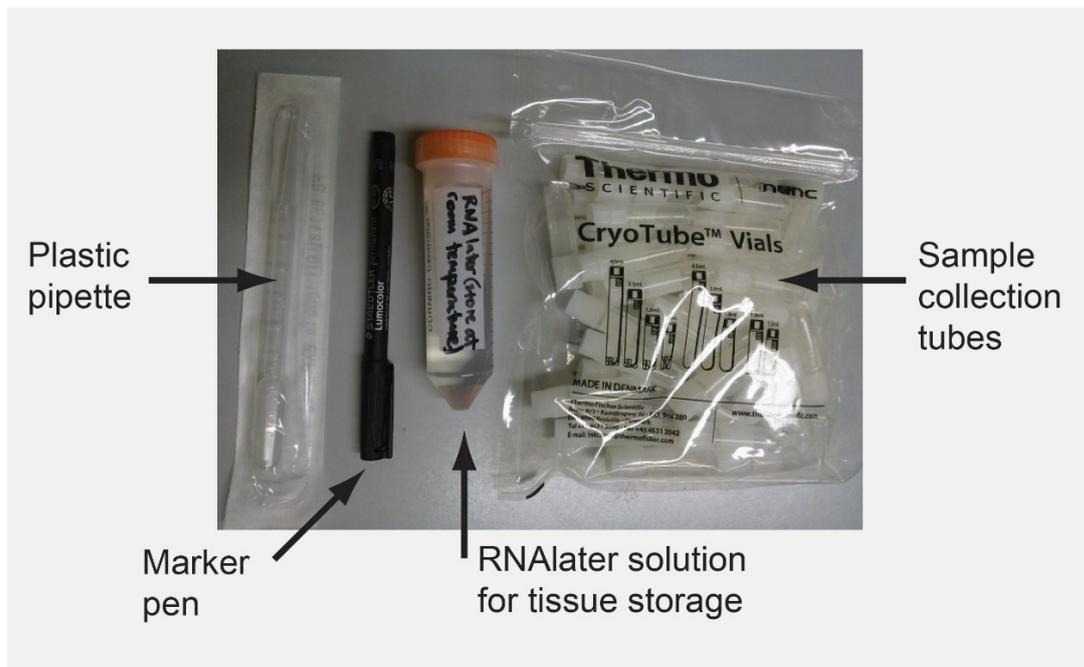


Figure 3-3 Sample collection kit sent to collaborators. Sample collection kit contains RNA later solution, cryotubes for sample collection, plastic pipette and a marker pen. See text for further details.

3.2.1.3 Sample collection procedure

Samples were collected according to the same sample collection protocol (see Additional file 3-2) to minimise the inconsistencies and biases caused by sampling methodology. RNA later, a buffer that maintains stability of nucleic acids in tissue biopsies, was used for collection of genetic samples. The following samples and metadata were collected for each CTVT case:

- (1) Tumour biopsy (size less than 0.5cm³) in RNA later - used for genetic analysis
- (2) Tumour biopsy in 10% neutral buffered formalin (formalin was provided by the collaborators) - used for histopathology analysis
- (3) Host tissue biopsy in RNA later (gonad, skin, blood or other) - used for genetic analysis
- (4) Photograph of the tumour
- (5) Completed case metadata collection sheet - metadata collected included age, gender, ownership and reproductive status of the dog, size and location of the tumour (see Additional file 3-3)

If the dog was noted to have multiple tumours or metastases, collaborators were asked to sample all tumour locations. The preferred host tissue was gonad (or other reproductive

tissues), as most of the dogs involved in the study were undergoing spay/neuter surgery at the time of sample collection. Other host tissues collected included skin, internal organs and blood. If blood was collected, collaborators were instructed to collect blood into EDTA tubes (Venosafe), with addition of RNAlater in the ratio 2:1.

Reported age metadata were categorised into the following three categories: young, adult and old. After a discussion with Dr Mirjam van der Wel (veterinarian at the Animal Anti-Cruelty League Port Elizabeth, South Africa), the age categories were defined as follows, to represent the age demographic structure of dog populations in low-income countries (where majority of sample collection was performed): young ≤ 1 year, adult >1 to <7 years, old ≥ 7 years. The entire/desexed status was reported at the time of CTVT diagnosis.

3.2.1.4 Sample shipment

After collection, collaborators were asked to store samples in the fridge. If samples were stored for more than 4 weeks, due to issues with sample shipment, the RNAlater samples were stored in the freezer, while formalin samples were stored in the fridge or at room temperature. Samples were shipped at room temperature by FedEx, DHL or by regular post. A DEFRA UK import license (TARP/2012/437(a), TARP/2012/406, TARP/2014/020, TARP/2015/293) was included with all shipments.

3.2.1.5 Sample storage and initial processing

Upon arrival, I assigned a unique number to each sample, and entered them into an Excel sample database (stored electronically on my laptop as well as on the departmental shared drive) and transcribed the information from the data collection sheets into the database. Additionally, I assigned a coordinate to each individual sample in the storage boxes (9x9 box, StarLab) and the samples in RNAlater were stored at -80°C (see Figure 3-4). The storage database was also stored electronically on my laptop as well as on departmental shared drive.

The formalin-fixed tissue samples were immediately transferred into a pot with 10% neutral buffered formalin (CellPath, CellStor pot, 20ml).



Figure 3-4 CTVT samples after being assigned a database number and prepared for long-term storage.

3.2.2 DNA extraction and quantification

3.2.2.1 Protocol optimisation

The DNA extraction from tissues stored in RNAlater was optimised by changing three variables: choice of DNA extraction kit, size of tissue input, and effects of RNase addition. The Qiagen DNeasy Blood & Tissue kit, GE Healthcare and Zymo DNA extraction kits were trialled, and the Qiagen DNeasy Blood & Tissue kit was the extraction kit of choice. Optimal size of tissue sample for highest yield was determined as 1.5-2.0mm³. RNase addition proved not to lead to a significant difference in DNA yield. The number of washing steps required to remove RNAlater from blood stored in RNAlater was optimised and determined to be two.

3.2.2.2 Variation in DNA yield between tissue types

DNA yield was dependent on the tissue type, as indicated in Figure 3-5 and Table 3-1. Mean DNA yield from tissue samples was 18.5µg (tumour DNA yield mean was 22.6µg, host DNA yield mean was 13.1 µg), mean DNA yield from blood samples was 2.9µg (see Table 3-1, Figure 3-5). The DNA yield obtained from host tissue (gonad/skin/internal organ) samples was consistently higher than the DNA yield obtained from host blood samples, therefore it was concluded that host gonad/skin/internal tissue samples were preferable to host blood samples.

Sample type	Mean DNA yield (µg)	Median DNA yield (µg)	Standard deviation (µg)
Tumour tissue	22.6	9.4	35.3
Host tissue (gonad, skin)	13.1	6.6	19.3
Host blood	2.9	1.2	4.1

Table 3-1 Summary of DNA yield obtained from different tissue types.

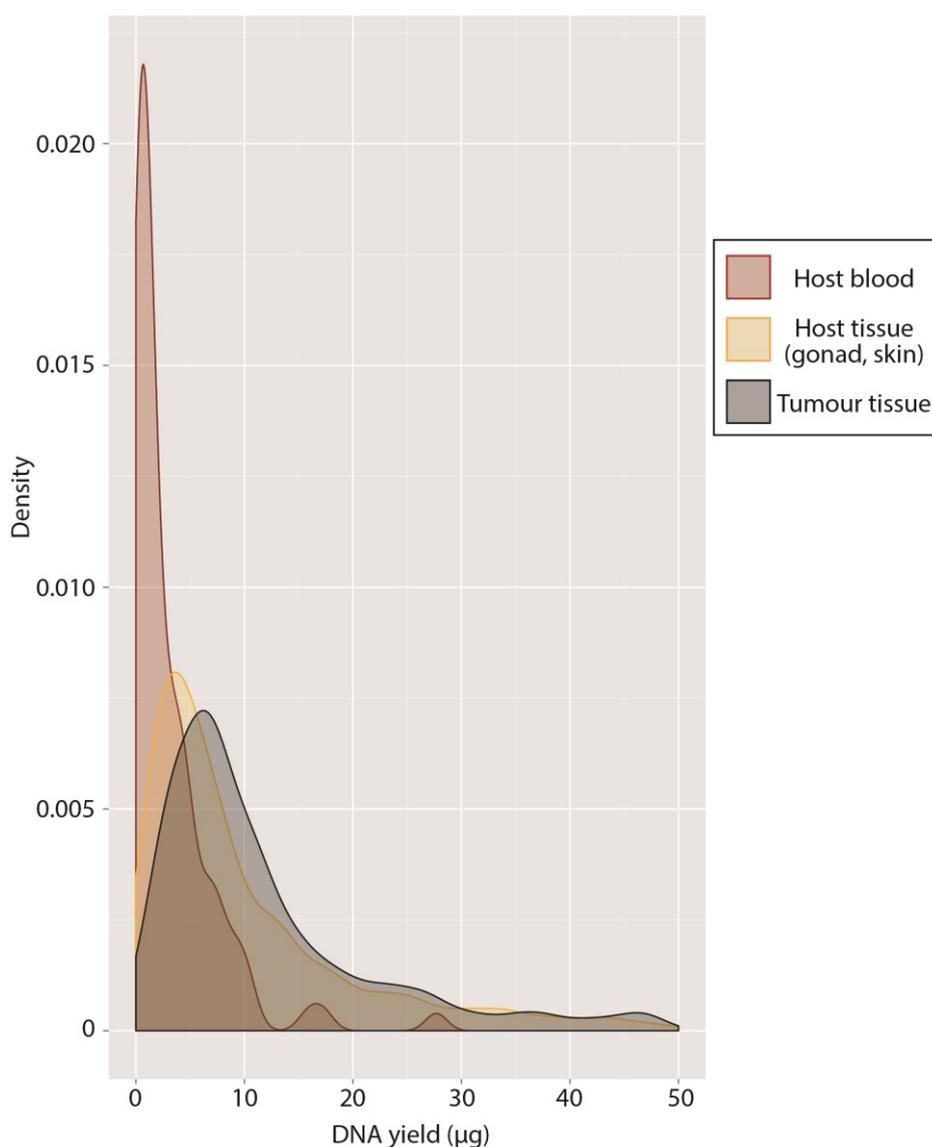


Figure 3-5 Kernel density estimate graph of DNA yield for different tissue types. Area under each curve is equal to 1. Note that x-axis limit is set to 50 µg to clearly display the peaks of each density distribution. Number of samples in this graph is as follows: host blood = 105, host tissue (gonad, skin) = 870, tumour tissue = 1218.

3.2.2.3 DNA extraction and quantification protocols

Based on the protocol optimisation experiments described above (see section 3.2.2.1), the following protocols were used for genomic DNA extraction from tumour/host samples stored in RNAlater and for genomic DNA quantification and DNA amplification.

3.2.2.3.1 DNA extraction from tissues

DNA was extracted from solid tissues using the Qiagen DNeasy Blood & Tissue kit (spin columns or 96-well plates) (see also Figure 3-6). The protocol for DNA extraction from tissues is described below (adapted from the Qiagen DNeasy Blood & Tissue Handbook). The Eppendorf centrifuge 5424 (tubes) or Eppendorf centrifuge 5804 (plates) were used for centrifugation steps.

- (1) Cut a section of tissue (1.5-2.0mm³) into small pieces and place it into an Eppendorf tube

Incubation step

- (2) Add 180µl of Tissue Lysis Buffer (buffer ATL)
- (3) Add 20µl of Proteinase K
- (4) Mix thoroughly using vortex and incubate overnight at 56°C on a rocking platform
- (5) Mix thoroughly using vortex and add 200µl of Lysis Buffer (buffer AL)
- (6) Mix thoroughly using vortex and add 200µl of ethanol (96-100%) and mix again using vortex (do not prolong this step)

Washing step

- (7) Spin the tube briefly and transfer the liquid into the DNeasy Mini Spin column placed in a collection tube (or DNeasy plate if extracting 96 samples)
- (8) Spin at 8,000 rpm for 1 minute
- (9) Discard the flow through and collection tube, place the DNeasy Mini spin column into a new collection tube
- (10) Add 500µl of Wash Buffer 1 (buffer AW1)
- (11) Spin at 8,000 rpm for 1 minute
- (12) Discard the flow through and collection tube, place the DNeasy Mini spin column into a new collection tube
- (13) Add 500µl of Wash Buffer 2 (buffer AW2)
- (14) Spin at 14,000 rpm for 3 minutes

Elution step

- (15) Discard the flow through and collection tube, place the DNeasy Mini spin column into a new Eppendorf tube
- (16) Add 200µl of Elution Buffer (buffer AE) directly onto the membrane
- (17) Spin at 8,000 rpm for 1 minute
- (18) The Eppendorf tube contains the extracted DNA
- (19) Place the DNeasy Mini spin column into a clean Eppendorf tube and repeat steps 16) and 17) to obtain the second elution

3.2.2.3.2 DNA extraction from blood

DNA was extracted from blood using the Qiagen DNeasy Blood & Tissue kit. The protocol for DNA extraction from blood is described below (adapted from the Qiagen DNeasy Blood & Tissue Handbook) (see also Figure 3-6). The Eppendorf centrifuge 5424 was used for centrifugation steps.

- (1) Pipette 1000µl of blood mixed with RNAlater into an Eppendorf tube
- (2) Spin for 5 minutes at 13,000 rpm

Washing step 1 (Phosphate Buffered Saline (PBS) wash)

- (3) Check if the supernatant is red (then the blood sample is in RNAlater)
- (4) Pipette out the supernatant and leave in the pellet together with a small rim of liquid
- (5) Add 500µl of PBS and mix thoroughly using vortex to re-suspend the pellet
- (6) Centrifuge for 5 minutes at 13,000 rpm
- (7) Discard the supernatant
- (8) Add 500µl of PBS and mix thoroughly using vortex to re-suspend the pellet again
- (9) Centrifuge for 5 minutes at 13,000 rpm
- (10) Discard the supernatant

Incubation step

- (11) Add PBS to a total volume of 200µl
- (12) Add 200µl of buffer AL
- (13) Add 20µl of Proteinase K
- (14) Mix thoroughly using vortex to re-suspend the pellet
- (15) Incubate for 10 minutes at 56°C

- (16) Add 200µl of ethanol (96-100%) and mix thoroughly using vortex
- (17) Transfer the liquid (maximum 600µl) into a DNeasy Mini Spin column
- (18) Centrifuge at 8,000 rpm for 1 minute
- (19) Discard the flow through and collection tube

Washing step 2 (Wash buffers)

- (20) Transfer the DNeasy Mini Spin column into a new collection tube and add 500µl of buffer AW1
- (21) Spin for 1 minute at 8,000 rpm
- (22) Discard the flow through and the collection tube
- (23) Transfer the DNeasy Mini Spin column into a new collection tube and add 500µl of buffer AW2
- (24) Spin for 3 minutes at 14,000 rpm
- (25) Discard the flow through and the collection tube

Elution step

- (26) Transfer the DNeasy Mini Spin column into an Eppendorf tube and add 100µl of buffer AE
- (27) Spin for 1 minute at 8,000 rpm
- (28) The Eppendorf tube contains the extracted DNA
- (29) Place the DNeasy Mini spin column into a clean Eppendorf tube and repeat steps 16) and 17) to obtain the second elution

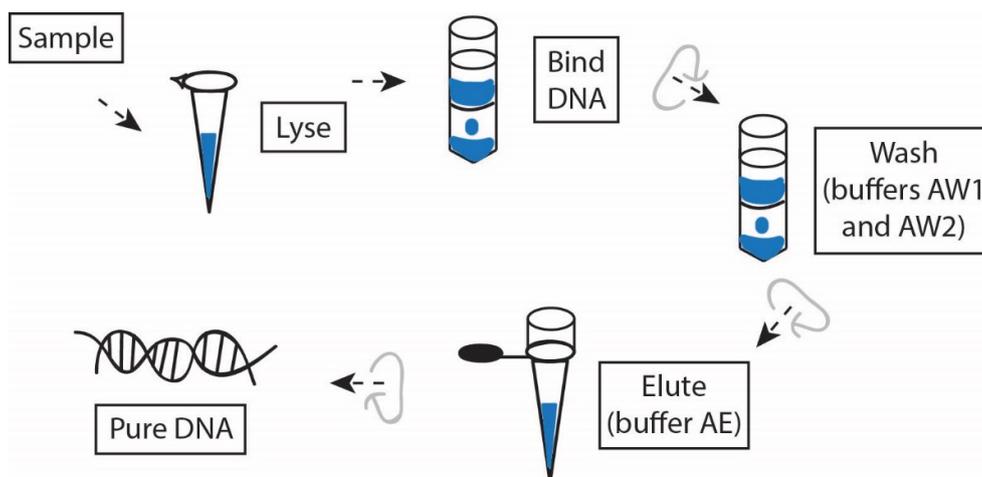


Figure 3-6 The Qiagen DNAeasy DNA extraction protocol (adapted from the DNeasy Blood&Tissue Kit Handbook)

3.2.2.3.3 DNA quantification using pico green

PicoGreen dye (Thermo Fisher Scientific) was used for high-throughput quantification of DNA. The DNA used for the standard curve was the Human Genomic DNA (Promega). The following protocol, adapted from Thermo Fisher Scientific PicoGreen manual (ThermoFisher Scientific 2008) was followed and optimised for use with the available spectrophotometer Envision Reader 2:

- (1) Thaw the PicoGreen reagent at room temperature (the reagent is frozen at 4°C), but keep it in the dark (light sensitive)
- (2) Make up 1X TE buffer (supplied at 20X concentration, dilute 1 in 20)
- (3) Prepare DNA standards - use Human Genomic DNA (Promega) to create a standard curve and create serial dilutions starting with concentration 1µg/µl
- (4) Dilute the test DNA - first 1 in 25 dilution, followed by 1 in 100 dilution
- (5) Pipette 100µl of diluted standard DNA onto a black plate (BD Falcon) (do not touch the bottom of the wells)
- (6) Pipette 100µl of diluted test DNA onto two black plates to create duplicates (do not touch the bottom of the wells)
- (7) Make up 25ml of PicoGreen solution by making 1 in 200 dilution of the PicoGreen reagent
- (8) Pipette 100µl of PicoGreen solution into each well on the test plate run 1, test plate run 2 and standards plate
- (9) Keep the plates in dark once the PicoGreen solution has been added
- (10) Spin the plates briefly (1 minute at 1,500 rpm)
- (11) Read the plates immediately on the Envision plate reader
- (12) Analyse the results using the standard curve method to obtain the final DNA concentration of the sample

3.2.2.3.4 DNA quantification using Qubit

Qubit dsDNA High Sensitivity and Broad Range assay kits (Thermo Fisher Scientific) were used as an alternative method to measure the double-stranded DNA concentration of a small number of samples. The protocol is described below:

- (1) Dilute the Qubit Reagent 1:200 in Qubit Buffer to prepare the working solution
- (2) Prepare two standards by pipetting 190µl of the working solution and 10µl of the supplied DNA standards into a thin-wall, clear 0.5ml PCR tube

- (3) Pipette 199 μ l of working solution into a thin-wall, clear 0.5ml PCR tube
- (4) Add 1 μ l of the sample test DNA
- (5) Mix for 2-3 seconds using vortex
- (6) Leave to stand at room temperature for 2 minutes
- (7) Read the fluorescence on the Qubit Fluorometer (Thermo Fisher Scientific) and note the final DNA concentration of the test sample

3.2.2.3.5 Whole genome amplification (WGA) protocol

The native DNA was amplified using the Whole Genome Amplification (WGA) kit (GE Healthcare Life Sciences, GenomePhi V2 DNA amplification kit) (note: whole genome amplified DNA was not used for subsequent quantitative polymerase chain reaction (qPCR) and DNA sequencing steps; rather, this step was performed to ensure availability of long-term DNA stocks for future screening experiments). The protocol (adapted from the GE Healthcare manual) is described below:

- (1) Pipette 9 μ l of sample buffer into each well on a PCR plate
- (2) Pipette 1 μ l of 20ng/ μ l sample DNA into each well on the PCR plate
- (3) Heat at 95°C for 3 minutes
- (4) Cool to 4°C on ice
- (5) Create a master mix with 9 μ l of reaction buffer and 1 μ l of enzyme for each sample
- (6) Pipette 10 μ l of the master mix into each well on the PCR plate
- (7) Incubate at 30°C for 2 hours
- (8) Heat to 65°C for 10 minutes to denature the enzyme
- (9) Pipette the contents of the PCR plate into a 1ml storage plate with 1ml of distilled water -> assumed concentration is now ~40ng/ μ l

3.2.3 CTVT diagnosis and estimation of host contamination levels

3.2.3.1 CTVT diagnosis

The *LINE-MYC* rearrangement specific to CTVT (Katzir et al. 1985, Katzir et al. 1987, Murgia et al. 2006, Rebbeck et al. 2009) (see Chapter 1, sections 1.2.3.1 and 1.2.7.3.1, Figure 3-1A) was amplified using a quantitative polymerase chain reaction (qPCR) with SYBR Green Master Mix (Thermo Fisher Scientific). The qPCR was performed with two primer sets (see Table 3-2): one primer set was specific for *ACTB*, and was used as a control for normalisation; the second pair of primers spanned the *LINE-MYC* rearrangement junction, and their product was

therefore specific to CTVT. The reaction conditions are displayed in Table 3-3 and Table 3-4. The method was optimised for use with 32 samples to be run on a 384 well plate. Each diagnostic qPCR reaction was performed in triplicate.

Primer	Sequence	
<i>LINE-MYC</i> primers (Rebbeck et al. 2009)	Forward	AGG GTT TCC CAT CCT TTA ACA TT
	Reverse	AGA TAA GAA GCT TTT GCA CAG CAA
<i>ACTB</i> primers (control)	Forward	CTC CAT CAT GAA GTG TGA CGT TG
	Reverse	CGA TGA TCT TGA TCT TCA TTG TGC

Table 3-2 Primers used in the CTVT diagnostic quantitative PCR reaction.

qPCR master mix reagents	Volume per reaction (µl)
SYBR Green Mix	10
Primer forward (5µM)	1.2
Primer reverse (5µM)	1.2
DNA (20ng/µl)	0.5
Distilled water	7.1
Total volume	20

Table 3-3 Input reagent volumes and concentrations for CTVT diagnostic quantitative PCR reaction.

Stage of qPCR amplification	Temperature (°C)	Time (s)
Initial denaturation	95	600
40 cycles	95	15
	60	60
Final dissociation	95	15

Table 3-4 Cycling conditions for CTVT diagnostic quantitative PCR reaction.

3.2.3.2 Estimation of host contamination levels

CDKN2A (cyclin-dependent kinase inhibitor 2A), a gene which is homozygously deleted in CTVT (Murchison et al. 2014) (see Figure 3-1B), but present in two copies in host tissue, was amplified using a qPCR reaction to estimate levels of host contamination. The qPCR was performed with two primer sets (see Table 3-5): one primer set was specific for *ACTB*, and

was used as a control for normalisation; the second pair of primers spanned the *CDKN2A* gene, and their product was therefore specific to host tissue. The qPCR reaction was performed at the same time as the CTVT diagnostic qPCR (see section 3.2.3.1), as a single diagnostic process. The reaction conditions are displayed in Table 3-3 and Table 3-4.

Primer	Sequence	
<i>CDKN2A</i> primers	Forward	CGG CCT TTG GAA GAT TTT GTT TGG CTT GA
	Reverse	CTT CCA CCC CTA GAA ACC TCC TTC T
<i>ACTB</i> primers (control)	Forward	CTC CAT CAT GAA GTG TGA CGT TG
	Reverse	CGA TGA TCT TGA TCT TCA TTG TGC

Table 3-5 Primers used in the quantitative PCR reaction to estimate host contamination levels.

3.2.3.3 Analysis of qPCR results

The analysis of the results from the quantitative PCR reactions to confirm CTVT status and to determine relative host contamination levels was carried out using the standard curve method (based on an example in the Applied Biosystems qPCR manual (Applied Biosystems 2004)). Standard curves were constructed with a two-fold serial dilution of DNA from CTVT tumour 29T1 (see Figure 3-7). Relative DNA input for each qPCR reaction was calculated using standard curves as follows: $(Ct = m(\log_{10}(iA)) + b)$, with parameters defined as follows: Ct = threshold cycle, m = slope of the standard curve, iA = relative DNA input amount, b = y-intercept of the standard curve. Mean relative DNA input for *LINE-MYC* and *CDKN2A* was calculated for each sample from qPCR triplicates, and normalised to mean relative DNA input for *ACTB* as follows: $\text{meanLINE-MYC}/\text{meanACTB}$ or $\text{meanCDKN2A}/\text{meanACTB}$; where meanLINE-MYC = mean relative *LINE-MYC* DNA input amount calculated from qPCR triplicates, meanCDKN2A = mean relative *CDKN2A* DNA input amount calculated from qPCR triplicates and meanACTB = mean relative *ACTB* DNA input amount calculated from qPCR triplicates. *LINE-MYC* and *ACTB* are present in three and two copies respectively in two previously analysed CTVT tumours (Murchison et al. 2014); however, it is possible that copy number at these loci differs between tumours in the current dataset.

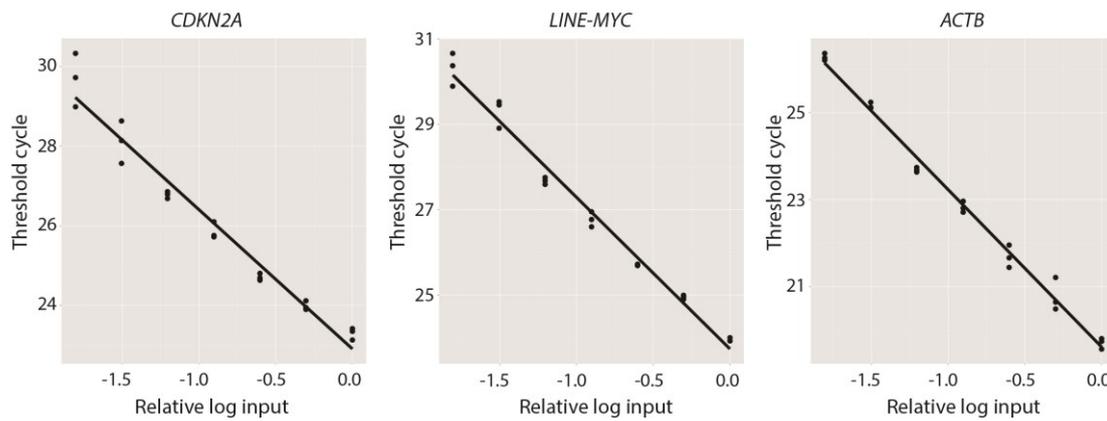


Figure 3-7 Example standard curves for CDKN2A, LINE-MYC and ACTB based on two-fold serial dilutions.

3.2.4 Processing and validation of histology samples

The formalin-fixed tissue samples were processed in the Histopathology laboratory at the Department of Veterinary Medicine, University of Cambridge. Samples were embedded in a paraffin block and haematoxylin-eosin stained slide was produced for each sample.

Visual validation using a microscope was performed to check for quality of cell fixation and for presence of CTVT cells in the tissue biopsy.

3.3 Results

3.3.1 Summary of genetic samples processed

A total of 1816 tumour samples for genetic analysis were collected from 1498 individual dogs (see section 3.3.2). There was a variable number of tumour samples per dog, hence the disparity between the number of samples and the number of dogs. DNA was extracted and diagnostic qPCR was performed on 1036 tumour samples from 918 dogs (see section 3.3.3). Out of these samples, 793 dogs were confirmed to be infected with CTVT (see section 3.3.3 and section 3.3.4), and an estimation of relative host contamination was made for these (see section 3.3.5). A summary of samples processed through the diagnostic pipeline is shown in Figure 3-8.

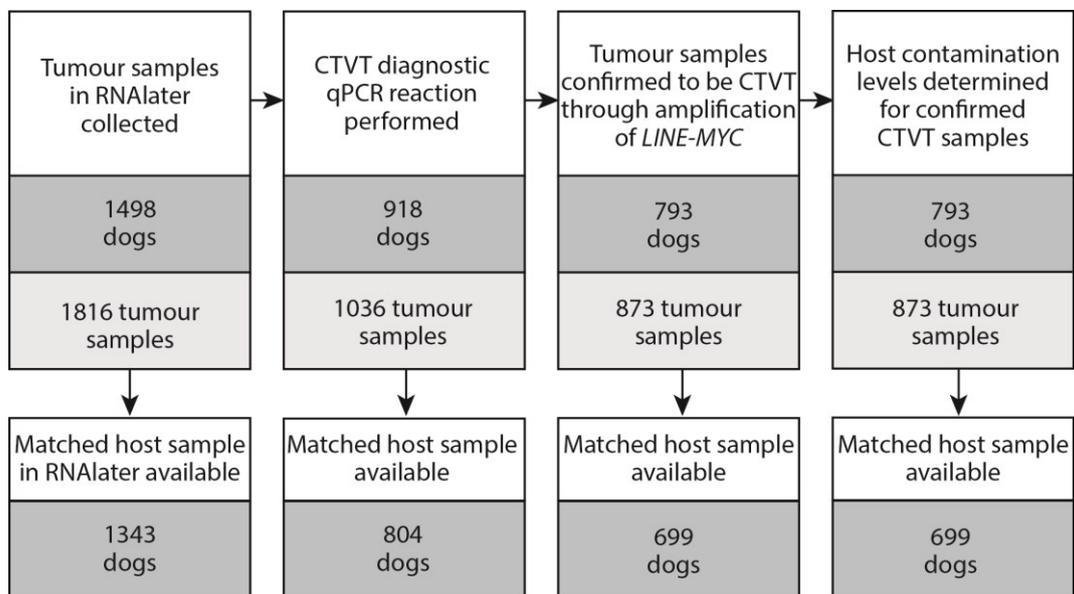


Figure 3-8 Summary of tumour samples collected and processed at each step of the diagnostic pipeline.

3.3.2 Genetic samples collected

The 1498 dogs with suspected CTVT tumours originated from 53 different countries in Europe, Asia, Africa, North America, South America and Australia. The total number of tumour biopsies collected in RNAlater was 1816 (see Figure 3-8); 201 individuals had more than one tumour biopsy collected. Samples from tumours in distinct locations were treated independently in further analysis. If more than one sample from the same tumour was collected, only a single biopsy was used for downstream analyses. A summary of tumour

locations is shown in Table 3-6. Extra-genital locations of tumours included skin (thigh, all over body, chest and neck), eye, mouth, nose, urethra, lymph nodes and rectum.

	Genital CTVT tumours only	Genital + extra-genital CTVT tumours	Extra-genital CTVT tumours only
Number of individuals	1420	27	51

Table 3-6 Summary of tumour locations for individuals with suspected CTVT tumours (n = 1498).

3.3.3 CTVT diagnostic validation

A total of 1036 tumour samples from 918 dogs (see Figure 3-8) were validated through the diagnostic qPCR reaction for *LINE-MYC* rearrangement (see section 3.2.3). The remaining tumour samples collected later were not validated using this method due to time restrictions. Of the 1036 tumours, 873 (84.3%) amplified the *LINE-MYC* rearrangement and therefore were confirmed to be CTVT tumours (Figure 3-9). Positive amplification was defined as *LINE-MYC* to *ACTB* normalised value > 0.01 (see 3.2.3). The mean *LINE-MYC:ACTB* ratio in tumour samples was 0.89. It should be taken into account that the *LINE-MYC:ACTB* ratio may be affected by variation in the *LINE-MYC* and/or *ACTB* copy number between tumours, which could lead to standardization inaccuracies.

Of the remaining 163 samples, 129 tumours (12.4%) did not amplify *LINE-MYC* (defined as normalised *LINE-MYC* value < 0.01) (Figure 3-9) and 34 tumours (3.3%) failed the qPCR, most likely due to degraded input DNA.

The *LINE-MYC* qPCR reaction was also performed on 834 normal host samples in order to test for CTVT contamination and to test the possibility that the rearrangement may naturally occur as a polymorphism in dog populations. Of these, 652 samples did not amplify *LINE-MYC*, as expected (Figure 3-9). On the other hand, 102 host samples showed evidence of amplification of the *LINE-MYC* rearrangement (defined as normalised *LINE-MYC* value > 0.01) (Figure 3-9). The mean *LINE-MYC:ACTB* ratio in host samples was 0.01. Lastly, 80 host samples failed the qPCR. The qPCR analysis allowed me, in addition, to identify samples where the tumour and host tissues were mislabelled.

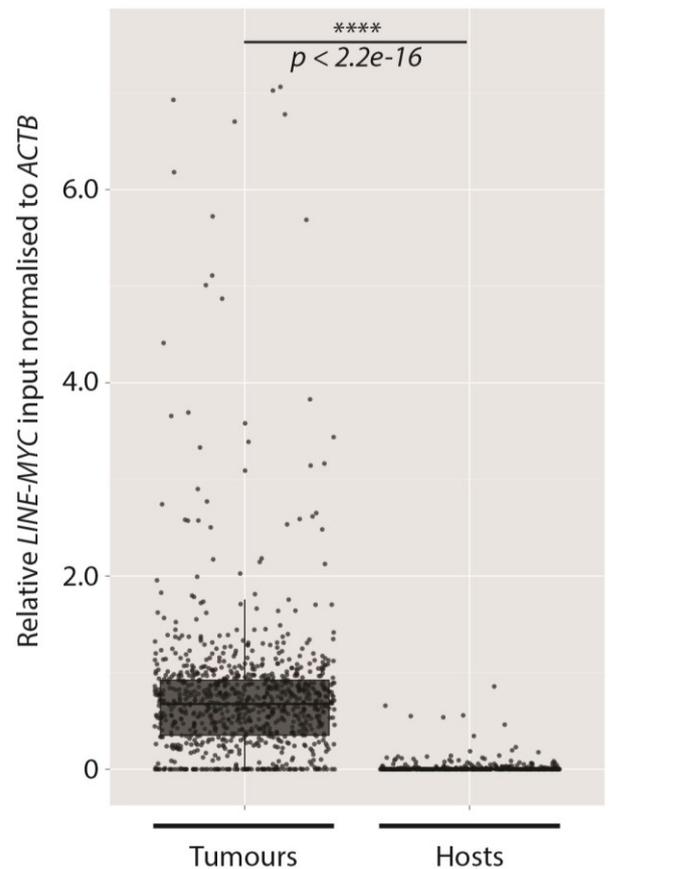


Figure 3-9 Relative DNA input of LINE-MYC for 1036 tumours and 834 hosts normalised to ACTB. Presence of LINE-MYC is a diagnostic marker of CTVT (Katzir et al. 1985, Katzir et al. 1987, Murgia et al. 2006, Rebbeck et al. 2009). LINE-MYC input normalised to ACTB is represented by a dot for each individual sample. Boxes represent the first and third quartiles (inter-quartile range = IQR). Error bars indicate values within 1.5*IQR away from first and third quartiles. P-value was calculated using two-sided student's t-test. Y-axis maximum is set to 7.5, thus 6 outlier points are excluded from this graph.

3.3.4 Set of validated CTVT samples

The set of tumours confirmed to be CTVT through amplification of *LINE-MYC* rearrangement consisted of 873 tumours that were collected from 793 individuals. The samples originated from 45 different countries in Europe, Asia, Africa, North America, South America and Australia (see Figure 3-10).



Figure 3-10 Map showing locations of confirmed CTVT tumours. Red dot represents a location where one or more confirmed CTVT samples were collected. The map in this figure was created using Adobe Illustrator.

Clinical data from 793 confirmed CTVT infected individual dogs are displayed in Table 3-7 and Table 3-8. The ratio of female to male individuals in the set was roughly 2:1. The majority of dogs were between 1 and 7 years of age, and were reported to be owned. Although the majority of dogs with CTVT were entire, a notable proportion (7.7%) were desexed prior to sample collection. The reported time of spay/neuter surgery ranged from a few months to 5 years before sample collection. The majority of dogs with CTVT were dogs of mixed breed (see Table 3-9).

	Genital CTVT tumours only	Genital + extra-genital CTVT tumours	Extra-genital CTVT tumours only
Number of individuals	763	15	15

Table 3-7 Summary of tumour locations for individuals with confirmed CTVT tumours (n = 793).

Clinical data		Number of individuals
Total number of individuals		793
Gender	Females	453 (57.1%)
	Males	246 (31.0%)
	Unknown	94 (11.9%)
Age	Young (≤ 1 year)	27 (3.4%)
	Adult (>1 to <7 years)	467 (58.9%)
	Old (≥ 7 years)	83 (10.5%)
	Unknown	216 (27.2%)
Ownership status	Owned	353 (44.5%)
	Stray/street	162 (20.4%)
	Adopted stray	5 (0.6%)
	Unknown	273 (34.5%)
Entire/desexed status at the time of CTVT diagnosis	Entire	510 (64.3%)
	Desexed	61 (7.7%)
	Unknown	222 (28.0%)

Table 3-8 Summary of clinical data for individuals with confirmed CTVT tumours (n = 793).

Breed of dog	Frequency	Breed of dog	Frequency
Mixed breed	416	Canis Panther	1
Central Asian Shepherd (Alabai)	1	Golden Retriever/Flat-coated Retriever/cross	6
Australian cattle dog	1	Great Dane/cross	2
Basenji	1	Greyhound	1
Beagle	1	Labrador Retriever/cross	14
Boerboel/cross	1	Russo-European Laika	0
Border Collie/cross	4	Maltese/cross	1
Boxer/cross	6	Mastiff/cross	1
Bulldog/cross	1	Pekingese/cross	4
Chihuahua/Chihuahua mix	1	Pitt bull	9
Chow Chow	2	Poodle/cross	9
Cocker spaniel/cross	4	Rottweiler/cross	3
Dachshund	2	Samoyed	1
Dalmatian/cross	0	Shar Pei	1
Doberman/Doberman cross	3	Shetland Sheepdog	0
Dogo Argentino	0	Shih Tzu	0
East-European Shepherd	0	Siberian Husky/cross	3
English springer	1	Staffordshire Bull Terrier	2
German Shepherd/cross	39	Weimaraner	2
Pembroke Welsh Corgi cross	1	Terrier breeds/cross	12
		Unknown	236

Table 3-9 Summary of breed information for individuals with confirmed CTVT tumours (n = 793).

3.3.5 Host contamination levels in confirmed CTVT samples

Estimation of tumour purity was made in the set of 793 dogs with confirmed CTVT tumours (total number of tumour samples collected from these individuals was 873), through amplification of *CDKN2A* - a gene homozygously deleted in CTVT (Murchison et al. 2014). The presence of host cells in the tumour samples is indicated by presence of *CDKN2A* amplification in tumours (Figure 3-11). Relative levels of host contamination were obtained. The mean *CDKN2A:ACTB* ratio in tumour samples was 1.58, which was significantly lower than the mean *CDKN2A:ACTB* ratio in host samples, 3.92 ($p < 2.2e-16$, two-tailed student's t-test).

It should be taken into account that the *CDKN2A:ACTB* ratio may be affected by variation in the *ACTB* copy number between tumours, which could lead to standardization inaccuracies. The assay was therefore only used to determine relative levels of host contamination in CTVT tumours, and will be complemented with more accurate calculations of host contamination levels obtained from the DNA sequencing data (see Chapter 4 and Chapter 5).

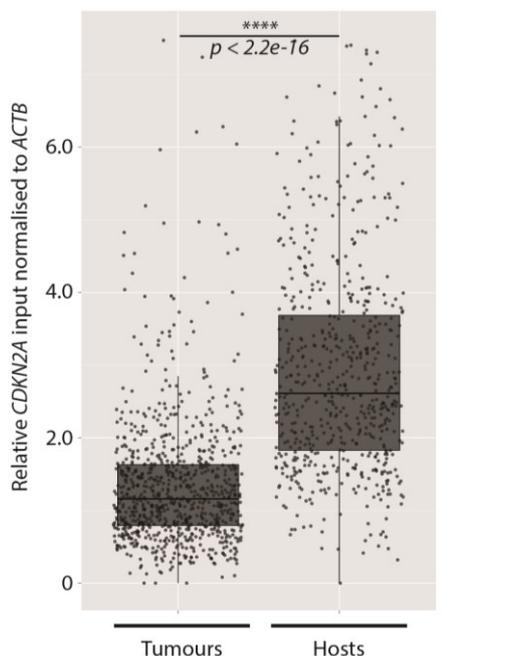


Figure 3-11 Relative *CDKN2A* input for 873 tumours and 652 hosts normalised to *ACTB*. *CDKN2A* input normalised to *ACTB* is represented by a dot for each individual sample. Boxes represent the first and third quartiles (inter-quartile range = IQR). Error bars indicate values within 1.5*IQR away from first and third quartiles. P-value was calculated using two-sided student's t-test. Y-axis maximum is set to 7.5, thus 56 outlier points are excluded from this graph.

3.3.6 Histology samples processed

A summary of histology samples available at each stage of the validation process (see sections 3.3.2, 3.3.3, 3.3.4 and 3.3.5 for details of each stage) is shown in Figure 3-12. Visual validation and inspection rendered a number of haematoxylin-eosin slides unusable for analysis due to problems with cell fixation or absence of CTVT cells from the biopsy (Figure 3-12).

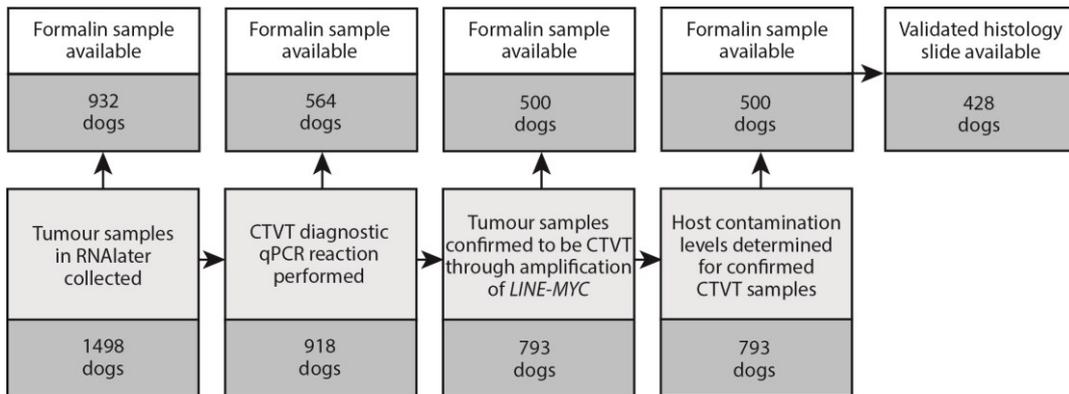


Figure 3-12 Summary of histology samples available at each stage of the validation process.

3.3.7 Dogs with the complete set of samples

The complete set of samples for each dog consisted of the following: diagnostically confirmed CTVT tumour in RNAlater, matched host in RNAlater and validated histology slide. A summary of dogs with the complete set of samples is shown in Figure 3-13.

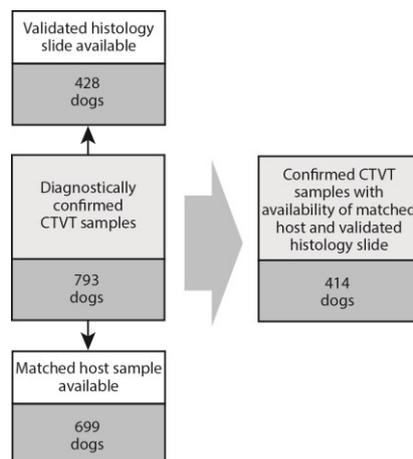


Figure 3-13 Summary of individuals with the complete set of validated samples.

3.4 Discussion

3.4.1 Summary of sample collection process

Overall, in this Chapter I received and analysed samples from 1498 individual dogs from over 50 different countries across inhabited continents. I optimised protocols and subsequently extracted the DNA from tumour and host genetic samples stored in RNAlater. Next, I designed a diagnostic strategy that I used to confirm CTVT diagnosis of 873 CTVT tumours collected from 793 dogs. For these CTVT tumours, I estimated levels of host contamination and I visually validated the histology slides available. The final collection of dogs with diagnostically confirmed CTVT and a complete set of samples for genetic and phenotypic analysis was 414 (see Figure 3-13).

Although I received and analysed a large set of CTVT samples from many countries around the world, some countries have been more difficult to reach than others, in terms of sample collection. These particularly include countries in Africa, central Asia and south-east Asia. It should be noted, therefore, that the sampling coverage across the world is not even and there may be genetic or phenotypic features of CTVT that are not captured in my analysis discussed in following Chapters (Chapter 4 and Chapter 5).

Another general challenge was obtaining an adequate sample quality, when the majority of samples were collected in field conditions with limited access to equipment. These limitations should be taken into account when performing the analysis in subsequent Chapters (Chapter 4 and Chapter 5). This specifically was a limitation for the histology samples, as it was impossible to standardise the fixation time, since some samples took longer periods of time to be shipped.

3.4.2 CTVT diagnosis through *LINE-MYC* amplification

The presence of the *LINE-MYC* rearrangement was used in this Chapter as a diagnostic method for confirming CTVT (Katzir et al. 1985, Katzir et al. 1987, Murgia et al. 2006, Rebbeck et al. 2009). The majority of tumours indeed did amplify this marker (see Figure 3-9), with the threshold defined as *LINE-MYC* to *ACTB* normalised value > 0.01. Even a low amplification of this marker indicates presence of CTVT cells in the biopsy. The variation in observed *LINE-MYC:ACTB* DNA ratios between tumours identified in this study is probably due to variable levels of normal host cells present in tumours; however, it is also possible that there is variation in *LINE-MYC* and/or *ACTB* copy number between tumours.

3.4.2.1 Interpretation of *LINE-MYC-negative* tumours

A proportion of tumours (129 samples, 12.5%) did not show *LINE-MYC* amplification (defined as normalised *LINE-MYC* value < 0.01). There are a number of possible explanations for why these tumours did not amplify the CTVT marker, as discussed below:

(1) These tumours were not CTVT

The diagnosis of CTVT in field conditions may often be challenging, especially in cases with unusual clinical presentation, such as extra-genital tumours. It is possible that these samples were clinically misdiagnosed.

(2) The tumour was CTVT, but CTVT cells were not present at detectable levels in the biopsy

In some cases, it is possible that there were very few CTVT cells in the tumour biopsy, such that *LINE-MYC* amplification was not detected. Alternatively, it is possible that a piece of host tissue was sampled, having been mistaken for tumour tissue. Especially in very advanced tumours, it may be difficult to tell where the boundary between the tumour and the host tissue is, and biopsies at the edge of the tumour may often consist of a majority of host cells.

(3) Not all CTVT tumours carry *LINE-MYC* rearrangement

Although *LINE-MYC* has so far been confirmed to be present in a large set of globally distributed CTVT tumours (Katzir et al. 1985, Katzir et al. 1987, Amariglio et al. 1991, Choi et al. 1999, Choi and Kim 2002, Liao et al. 2003, Murgia et al. 2006, Park et al. 2006, Rebbeck et al. 2009), as discussed in section 3.1.2) we cannot exclude the possibility that a subpopulation of CTVT tumours do not carry this rearrangement, either representing a CTVT lineage that diverged from other lineages before acquisition of *LINE-MYC*, or due to loss of this particular locus in some lineages. Since these tumours tended to have very high levels of host contamination as defined by amplification of *CDKN2A*, which has been found to be deleted in CTVT tumours previously analysed (section 3.1.2), it is more likely that these samples were mostly host tissue.

For cases which did not amplify *LINE-MYC* rearrangement and for tumours which were not analysed through this diagnostic method at all, further diagnosis was based on the clinical history, the appearance of tumours on photographs and histological features.

3.4.2.2 Interpretation of *LINE-MYC* positive hosts

As *LINE-MYC* rearrangement is believed to be specific to CTVT (Katzir et al. 1985, Katzir et al. 1987, Amariglio et al. 1991, Choi et al. 1999, Choi and Kim 2002, Liao et al. 2003, Murgia et al. 2006, Park et al. 2006, Rebbeck et al. 2009), I did not expect any amplification of this marker in host samples. Nevertheless, 102 host samples (12.2%) amplified low levels of *LINE-MYC* rearrangement (defined as normalised *LINE-MYC* value > 0.01) (Figure 3-9). The explanations for this observation are discussed below:

(1) Contamination by CTVT cells

It is possible that host samples were contaminated by CTVT tumour samples. The contamination could either happen at sample collection (if identical scalpel was used for taking biopsy, especially in cases where a piece of gonad was biopsied only after the surgery was completed), or the host sample might be contaminated by metastasised CTVT tumours cells (which may be the case for example in blood samples with free floating CTVT cells). It is also plausible that cross-contamination could have occurred during laboratory handling. DNA sequencing analysis performed at a later stage (see Chapter 4) confirmed presence of low levels of CTVT cells in a number of host samples, thus supporting this explanation.

(2) *LINE-MYC* is a rare germline polymorphism

Although *LINE-MYC* rearrangement has, to date, not been observed in any normal dogs (Katzir et al. 1985, Katzir et al. 1987, Amariglio et al. 1991, Choi et al. 1999, Choi and Kim 2002, Liao et al. 2003, Murgia et al. 2006, Park et al. 2006, Rebbeck et al. 2009), we cannot exclude the possibility that the *LINE-MYC* rearrangement is a rare germline polymorphism, which has not yet been observed in the canine population. However, low levels of amplification of *LINE-MYC* relative to *ACTB* in host tissues (Figure 3-9) suggest that this is unlikely to be the explanation in the set of samples analysed here.

3.4.3 Clinical data summary of CTVT samples collected

My previous work presented in Chapter 2 (Strakova and Murchison 2014) did not provide any evidence for a gender bias for CTVT infection. However, the CTVT sample collection that I obtained is biased towards female dogs (see Table 3-8). It should be noted, however, that a significant proportion of samples in this set was collected as a part of spay/neuter programmes, which focus on sterilising female dogs. This conscious bias may lead to an

increased proportion of female dogs being screened for CTVT tumours and thus to an increased proportion of female dogs in our sample set.

The majority of dogs in the sample set were between 1 and 7 years of age, with a significantly smaller proportion of dogs being 7 years and over and a very small proportion of dogs that were below 1 year of age (see Table 3-8). The age distribution in our sample set reflects the fact that the most common route of CTVT transmission is through coitus. CTVT transmission in young dogs that are not sexually active is likely to occur through other routes, such as sniffing, licking or biting (Ndiritu et al. 1977, Weir et al. 1978, Perez et al. 1994, Guedes et al. 1996, Albanese et al. 2002, Chikweto 2013). Although it has been reported that puppies born to mothers exposed to CTVT are less susceptible to the disease (Yang and Jones 1973), we have come across rare cases of CTVT transmission from mother to offspring during birth, possibly as a result of sparse injuries during the birth process, leading to occurrence of CTVT tumours on skin of infected offspring.

Free-roaming dogs have previously been indicated as a reservoir for the disease (see Chapter 2, (Strakova and Murchison 2014)). Nevertheless, however, significantly more dogs in our sample set were owned rather than 'stray' (see Table 3-8). The term 'stray' is, however, inconsistently defined in the literature and there is no general agreement on what the term means (Morters personal communication 2017). We have reported the term as understood by the submitting veterinary collaborator and we acknowledge the inconsistencies in their use of the term, which are reflected in their answers. It should be considered that in many countries of the world, if a dog is 'owned', it would still spend most of the time outside in the street, thus effectively becoming a free-roaming dog. This inconsistency in the meaning of 'owned' and 'stray' dogs may significantly affect the ratios, especially as most dogs sampled come from lower-income countries where 'owned' dogs are often communal/free-roaming.

The usual route of CTVT transmission is through coitus, and it is therefore to be expected that the majority of dogs in our sample set are entire at the time of CTVT diagnosis (see Table 3-8). Perhaps more surprisingly, there was still a notable number of dogs (7.7%) that had been sterilised prior to diagnosis of CTVT (see Table 3-8), with the time of surgery reported to be up to 5 years before sample collection. As discussed in Chapter 2 - this suggests that either the latent period for CTVT development, which previous anecdotal published reports have suggested may last for weeks or months (Smith and Washbourn 1898b, Ajello 1939, Locke et al. 1975), can sometimes last years, or, alternatively, that spaying and neutering does not always protect dogs from CTVT. Furthermore, as mentioned in Chapter 2, non-coital

modes of CTVT transmission, including sniffing, licking or biting, may also contribute to CTVT infection of spayed/neutered dogs (Ndiritu et al. 1977, Weir et al. 1978, Perez et al. 1994, Guedes et al. 1996, Albanese et al. 2002, Chikweto 2013).

With free-roaming dogs being implicated as a reservoir for the disease (Chapter 2 and (Strakova and Murchison 2014)), it would be anticipated, that the vast majority of dogs would be of local, mixed breed – this indeed was found to be the case, with over 50% of dogs classified as local, mixed breed (see Table 3-8).

3.4.4 CTVT samples in my future work

Sample collection and validation was an important part of this PhD work that formed the basis for tackling the main goal of this Thesis: to analyse the genetic diversity and evolution of CTVT around the world. This set of samples enabled me to perform DNA sequencing of hundreds of CTVT tumours, as further discussed in subsequent Chapters (see Chapter 4 and Chapter 5), thus expanding our knowledge of CTVT genetics.

Moreover, each of the validation steps has generated useful information about the samples collected – this information was valuable in the next stage of my PhD work, to allow me choose which samples should be sequenced using next-generation sequencing technologies (see Chapter 4 and Chapter 5).

Each of the three types of samples collected (see section 3.2.1.1) were used for subsequent parts of my PhD project, as discussed below:

- (1)** Tumour biopsy in RNAlater – used for DNA sequencing and subsequent genetic analysis of CTVT diversity (see Chapter 4 and Chapter 5)
- (2)** Host biopsy in RNAlater – used for DNA sequencing and subsequent filtering of genetic variation arising from host contamination (see Chapter 4 and Chapter 5)
- (3)** Tumour biopsy in 10% neutral buffered formalin – used for histopathology analysis of CTVT diversity (see Chapter 4 and Chapter 5)

4

Mitochondrial genetic diversity, selection and recombination in canine transmissible venereal tumour

Summary

Canine transmissible venereal tumour (CTVT) is a clonally transmissible cancer that originated from a single dog living approximately 11,000 years ago (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014). Despite the clonal origin of the CTVT nuclear genome, CTVT mitochondrial genomes (mtDNAs) have been acquired by periodic capture from transient hosts (Rebbeck et al. 2011). With the goals of tracing the historical spread of CTVT and assessing the global genetic and phenotypic diversity of this disease, we sequenced 449 complete mtDNAs from a worldwide population of CTVTs. The analysis presented in this Chapter shows that mtDNA horizontal transfer has occurred at least five times in the history of CTVT, delineating five tumour clades. The study shows that CTVT appears to have spread rapidly around the world within the last 2,000 years, perhaps transported by dogs travelling along historic maritime trade routes. Further, this work indicated that negative selection has operated to prevent accumulation of deleterious mutations in captured mtDNA, and that recombination has caused occasional mtDNA re-assortment. A histology-based screen of CTVT clades did not show any significant phenotypic differences between groups. These findings implicate functional mtDNA as a driver of CTVT global metastatic spread, further highlighting the important role of mtDNA in cancer evolution.

4.1 Introduction

4.1.1 Mitochondria in CTVT

The canine transmissible venereal tumour (CTVT) has spread around the world as a single clonal lineage; indeed phylogenetic trees constructed using nuclear genetic variation from CTVT and normal dogs indicate that CTVT tumours form a monophyletic group (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014).

Interestingly, previous work has shown that CTVTs do not cluster as a monophyletic group on phylogenetic trees constructed using mtDNA from CTVT and other canids; rather, CTVT mtDNAs cluster on two distinct branches of the tree (see Chapter 1, Figure 1-13) (Murgia et al. 2006, Rebbeck et al. 2011). This finding indicates that mtDNA in CTVT must have been acquired from its hosts through mtDNA horizontal transfer events during the evolution of the lineage (see Figure 4-1). This observation may suggest that replacement of CTVT mtDNA, which presumably was carrying large numbers of possibly deleterious mutations, provided a selective advantage to the lineage (although we cannot exclude the possibility that CTVT acquired host mitochondrial DNA via purely neutral processes) (Rebbeck et al. 2011).

The frequency of mtDNA horizontal transfer in CTVT is less well understood. The work presented by Rebbeck et al. suggests that horizontal transfer of mtDNA has taken place at least twice in the history of CTVT, as displayed in Figure 4-1 (Rebbeck et al. 2011). However, we cannot exclude the possibility that the original CTVT mtDNA has not been completely replaced and therefore only a single mtDNA horizontal transfer has occurred. Taking into account the possibility that mtDNA horizontal transfer is a positively selected process in CTVT, we may expect that it has taken place more than just once or twice during the evolution of CTVT, and that this phenomenon is perhaps even more common than previously appreciated. By analysing mtDNA from a larger set of CTVT tumours around the world, we may be able to identify additional mtDNA horizontal transfer events and thus examine the frequency of this phenomenon.

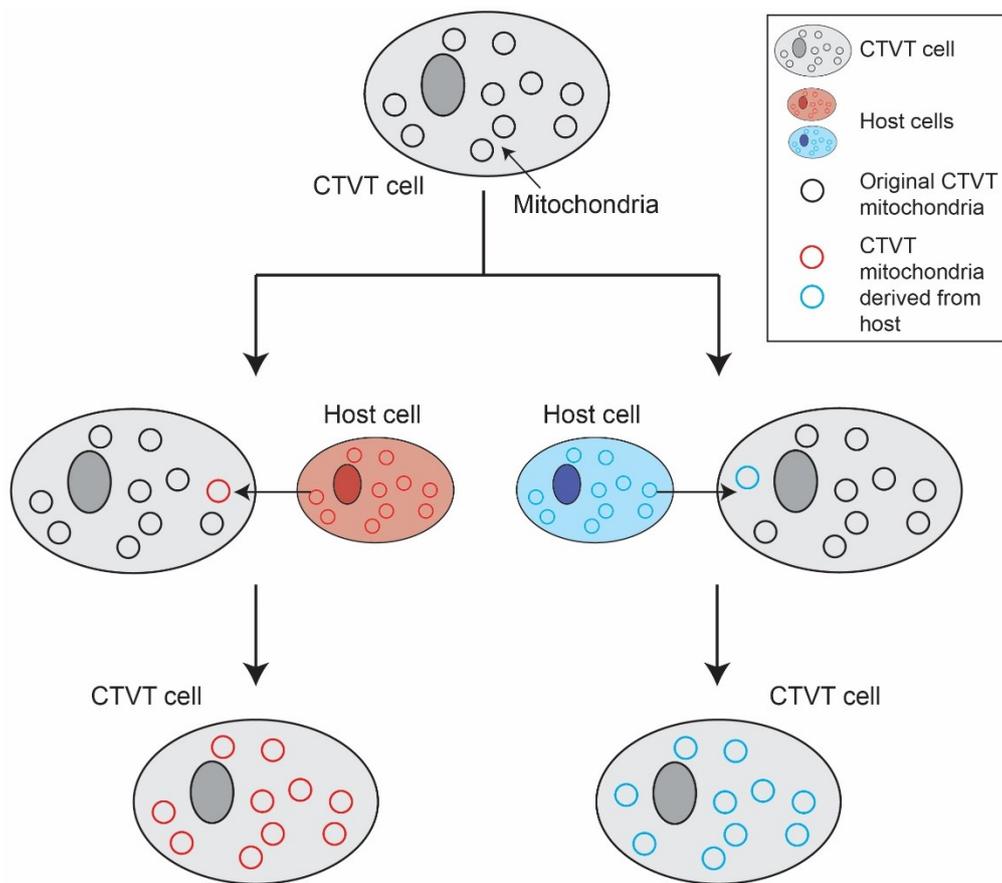


Figure 4-1 Horizontal transfer of mitochondrial genomes in CTVT. CTVT cells (grey) are shown acquiring mitochondrial genomes from host cells (red and blue). Over time, the original CTVT mitochondrial genome (black), which presumably carried a large mutational burden, is replaced by the acquired host mitochondrial genomes (red and blue) via a process of genetic drift or positive selection (Rebbeck et al. 2011). Figure is adapted from (Strakova and Murchison 2015).

4.1.2 Genetic changes in mtDNA

Somatic mutations are genetic changes that are acquired during the evolution of the tumour and are the source of genetic variation within the tumour population. The somatic variants which offer a growth advantage, may be positively selected and act as the ‘driver’ mutations in the formation of a tumour (Ostrow et al. 2014). On the other hand, a mutation, which is deleterious to the cell, may be selected against. Most somatic mutations in cancer are believed to be selectively neutral, captured in the cancerous clone by hitchhiking together with a small number of positively selected driver mutations. Interestingly, there is limited evidence in human cancers for negative selection operating to safeguard essential cellular processes (Campbell et al. 2008, Stratton et al. 2009, McFarland et al. 2013, Ju et al. 2014, Stewart et al. 2015, Martincorena et al. 2017). The extreme longevity of CTVT as a cancer

raises the possibility that the accumulation of mutations has become a burden rather than an advantage in this lineage.

Somatic mutations in mtDNA have for a long time been suspected to play a role in carcinogenesis (Brandon et al. 2006, Ohta 2006, Zong et al. 2016, Yuan et al. 2017), especially as many cancers are considered to have altered energy metabolism (Warburg 1956a, Warburg 1956b, Hanahan and Weinberg 2011). Their specific significance for the development of cancer is still unclear and may depend on cancer type (Ju et al. 2014, Yuan et al. 2017). Previously published work has provided evidence showing that somatic mutations in mtDNA can be positively selected in certain types of cancers (Wallace 2012). Evidence for negative selection acting on somatic mutations in mtDNA was only published after this PhD work was started, indicating that functional mitochondria are essential for some types of cancer cells (Ju et al. 2014, Stewart et al. 2015, Yuan et al. 2017). The exceptionally long life span of CTVT gives us an opportunity to search for evidence of negative selection in mtDNA genomes, which may not be detectable during the life span of other cancers.

4.1.3 MtDNA recombination

It is commonly believed that mtDNA usually does not recombine (see Chapter 1, section 1.4.1). Examples of mtDNA recombination have, however, been observed in various eukaryotes, in rarely-detected cases of human biparental mtDNA inheritance and in human cell extracts (Thyagarajan et al. 1996, Lunt and Hyman 1997, Ladoukakis and Zouros 2001, Hoarau et al. 2002, Bergthorsson et al. 2003, Kraytsberg et al. 2004, Gantenbein et al. 2005, Zsurka et al. 2005, Ujvari et al. 2007). In contrast, mtDNA recombination has not yet been detected in cancer cells. In general, detection of mtDNA recombination is a difficult task, due to lack of genetic markers. In CTVT, however, we have a unique opportunity to search for this phenomenon, as after each mtDNA horizontal transfer, two distinct mtDNA haplotypes co-exist in the same cell.

4.1.4 Genetic diversity in CTVT

Despite the fact that CTVT is a single clonal cell lineage that arose from one dog living a few thousand years ago (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014), one of the most interesting features of the disease is the diversity of CTVT tumours around the world. Genetic analysis of somatic mutations acquired by mtDNA in CTVT tumours from different parts of the world allows us to trace the spread of the disease, through constructing

a phylogenetic tree and using genetic markers to follow its historical travel routes. When analysing the mtDNA in CTVT, it should be considered that we can only trace the disease as far back as the time when this mtDNA haplotype was captured by the cancer lineage.

4.1.5 Linking phenotypic diversity to genetic changes in CTVT

As CTVT has acquired mtDNA and diverged genetically as it has spread around the world, it may have acquired phenotypic changes as a result of underlying genetic differences. Previous studies have indicated variation in lymphocyte infiltration and mitotic index in CTVT tumours (Wade 1908, Bloom et al. 1951, Weir et al. 1978, Chandler and Yang 1981, Gonzalez et al. 2000). These changes have only been associated to different stages of CTVT growth, with regressing tumours often infiltrated by lymphocytes, plasma cells, and macrophages (Wade 1908, Bloom et al. 1951, Weir et al. 1978, Chandler and Yang 1981, Gonzalez et al. 2000). It has not, however, been previously explored whether the phenotypic differences between CTVT tumours could be associated to genetic differences between CTVT lineages. Moreover, it has not previously been investigated how the uptake of mtDNA might affect CTVT tumour behaviour.

4.1.6 Goals of this Chapter

In this Chapter, we sequenced 449 complete CTVT mitochondrial genomes from 39 countries around the world using a whole genome skim sequencing approach and performed a histopathology based screen of these CTVT tumours, with the aim to analyse the worldwide CTVT genetic and phenotypic diversity. The individual goals of this Chapter are outlined below:

- (1)** To call genetic variants in CTVT mitochondrial genomes and construct a phylogenetic tree
- (2)** To determine the number of mitochondrial DNA horizontal transfer events, which will correspond to the number of CTVT mitochondrial clades
- (3)** To analyse the phylogenetic relationships between CTVT tumours with the aim to trace and describe the historical spread of CTVT
- (4)** To search for evidence of negative selection operating on CTVT mtDNA
- (5)** To search for evidence of mitochondrial DNA recombination

- (6)** To perform a screen of histopathological variation in CTVT and to link the results to the underlying genetic changes

The analysis described in this Chapter was published in 2016 (*Strakova et al. eLife, 2016; 5:e14552*; see Additional file 6-1 for full article text). The study was performed in close collaboration with Máire Ní Leathlobhair, another PhD student in the Transmissible Cancer Group. Throughout the Chapter, Máire's contributions have been acknowledged. The datasets generated are publicly available at GenBank (accession no: KU290400 - KU291095) and at European Nucleotide Archive (accession no: ERP014691).

4.2 Materials and Methods

4.2.1 Sample collection, DNA extraction and CTVT diagnosis

Samples were collected and DNA was extracted as described in Chapter 3. Quantitative polymerase chain reaction (qPCR) assays were performed to confirm CTVT diagnosis (Additional file 4-1) by detection of the CTVT-specific *LINE-MYC* genomic rearrangement (Katzir et al. 1985, Katzir et al. 1987, Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014), as described in Chapter 3 (see section 3.2.3).

4.2.2 Library preparation and DNA sequencing

A total of 50µl of native DNA was submitted in 96-well super plates (AB-2800, Thermo Fisher Scientific) to the Wellcome Trust Sanger Institute (WTSI) Sample Management Team. Quality control checks were performed at WTSI (DNA quantification and DNA quality control gel). Whole genome sequencing libraries with insert size of 100 to 400 base pairs (bp) were generated using standard methods according to WTSI Illumina library preparation protocols, and sequenced with 75bp paired end reads on an Illumina HiSeq2000 instrument (Illumina, San Diego, CA). To achieve whole genome depth of ~0.3X, 36-44 samples were pooled per single sequencing lane. The average mtDNA coverage in tumour tissues was ~68X and in host tissues was ~58X (see Figure 4-2). Host samples were found to have a more variable sequencing coverage across mtDNA in comparison to the tumour samples (see Figure 4-2). Some host samples with lower DNA quality (e.g. blood samples) were found to have an extremely low coverage across mtDNA (defined as average coverage <20X, Additional file 4-2A) and were treated with caution in subsequent analyses.

Samples 1380T and 1381T were sequenced separately by Dr Guo-Dong Wang (Kunming Institute of Zoology, Chinese Academy of Sciences), based on the methods described in Pang et al. (Pang et al. 2009). Complete mitochondrial genomes were amplified using the primers listed in Pang et al. and Strakova et al. (Pang et al. 2009, Strakova et al. 2016). The purified PCR products were sequenced on a 3730xl DNA analyser (Applied Biosystems) with a Big Dye Terminator v3.1 Sequencing Kit (Applied Biosystems). The sequenced fragments were assembled by Seqman (DNASTAR, Madison, WI) and the complete mitochondrial genomes were aligned with the CanFam3.1 dog mitochondrial reference genome (Lindblad-Toh et al. 2005).

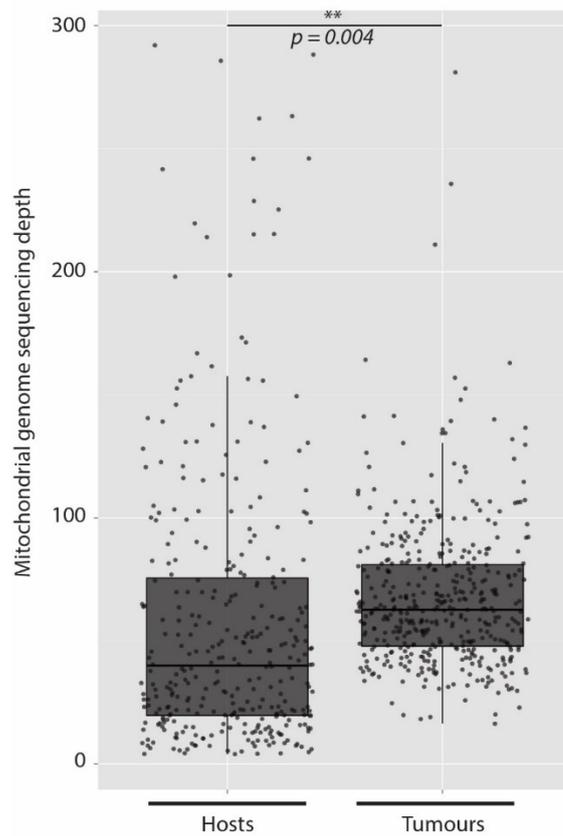


Figure 4-2 Mitochondrial genome sequencing coverage in 338 host samples and 449 CTVT tumour samples. Coverage of each individual sample is represented by a dot. Boxes represent the first and third quartiles (inter-quartile range = IQR). Error bars indicate values within 1.5*IQR away from first and third quartiles. P-value was calculated using two-sided student's t-test.

4.2.3 Sequencing validation and quality assessment

The following validation checks were performed:

- (1) Sample swap check
- (2) Duplicate samples check
- (3) High host contamination check
- (4) Non-CTVT tumour check
- (5) Failed sequencing

These validation checks were performed using copy number plots and variant allele fraction (VAF) plots, as described below and further in section 4.2.5.3.

4.2.3.1 *Copy number plots*

Copy number was plotted by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group) based on scripts provided by Dr David Wedge (Wellcome Trust Sanger Institute). Depth for each 10kb window (= number of sequencing reads mapping to each 10kb window) was plotted against genome position. The CTVT copy number profile is characteristic of this lineage (see Figure 4-3); host samples showed an expected uniform copy number profile of 2 copies across autosomes (see Figure 4-3).

4.2.4 MtDNA copy number calculations

The following equation was used to calculate mtDNA copy number: $(\text{mtCOV}/\text{nuclCOV}) * P$, where mtCOV = average coverage across mtDNA, nuclCOV = average coverage across the nuclear genome and P = ploidy. The ploidy used in our calculations was 2 for both CTVT tumours and CTVT hosts (Murchison et al. 2014). Host and tumour samples with average MT coverage >300X were excluded from the copy number calculations (Additional file 4-2A). Host contamination was not taken into account during the mtDNA copy number calculations.

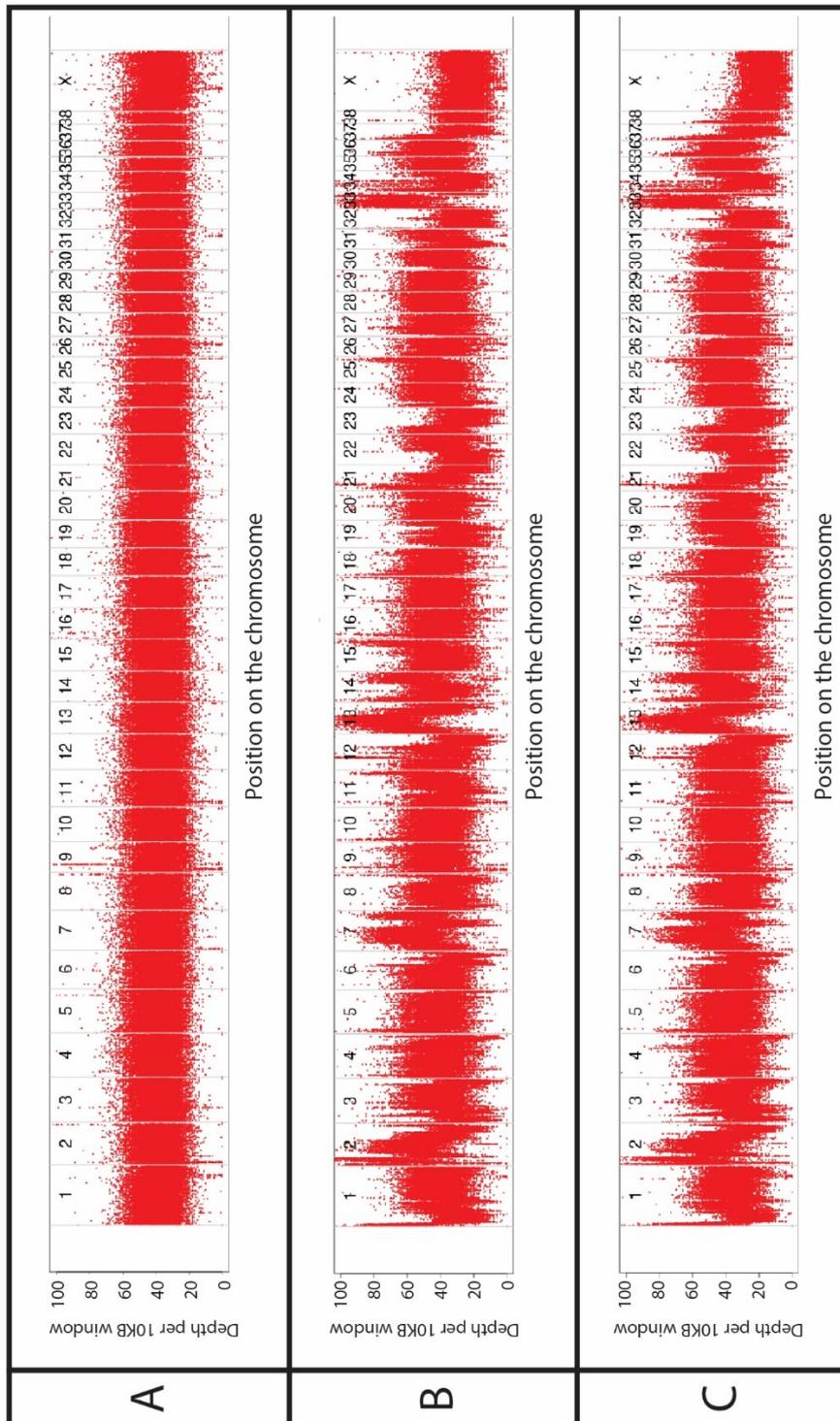


Figure 4-3 Copy number plots showing depth across the whole genome for 10kb windows. (A) Normal female host (B) CTVT tumour from Nicaragua (C) CTVT tumour from India. Chromosomes are labelled at the top of each plot. Plots were made by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group).

4.2.5 Substitution calling

4.2.5.1 *Sequence alignment*

Reads were aligned with the CanFam3.1 dog reference genome (Lindblad-Toh et al. 2005) (http://www.ensembl.org/Canis_familiaris/Info/Index) using the BWA alignment tool (Li and Durbin 2009).

4.2.5.2 *Extraction and filtering*

Substitutions were called by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group) using CaVEMan (Cancer Variants through Expectation Maximisation), an in-house variant calling algorithm, as previously described (Nik-Zainal et al. 2012a) (<http://cancerit.github.io/CaVEMan/>). As CaVEMan is designed for matched tumour-normal data, and CTVT tumours and normals are unmatched (i.e. they are different individuals), we used simulated reads derived from the reference genome as the “normal”, and called all variants relative to this. A variant allele fraction (VAF value, i.e. number of reads supporting the variant as a fraction of the total number of reads covering the variant position) was reported for each substitution detected. The following list of in-built post-processing filters was used to improve the specificity and sensitivity of substitution calls:

- (1) At least one third of mutant alleles must have base quality >25
- (2) Mean mapping quality of reads supporting a variant call must be ≥ 21
- (3) Variant calls supported only by the first or last 15bp of reads were discarded
- (4) Variants were discarded if they occurred 10bp upstream or downstream of an unfiltered indel called in the same sample (as detected by the indel-detecting algorithm *cgpPindel*, see section 4.2.6). The 10bp range was extended by the REP value for samples where an indel had been called with REP>0; REP represents the number of times the inserted/deleted base(s) occurs in the sequence directly 5' or 3' of the putative indel
- (5) Variant calls were discarded if they occurred within region MT:16129-16430 inclusive; this is a simple repeat region as defined by the UCSC (<http://genome.ucsc.edu/>) table browser (Dog, CanFam3.1)
- (6) If the reference allele was supported by at least one read on both strands (forward and reverse), then we required that the mutant allele should be supported by at least one read on both strands

4.2.5.3 *Variant allele fraction (VAF) plots*

Variant allele fraction (VAF) plots for the mitochondrial genomes were plotted by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group) and myself. VAF value represents the fraction of reads at a variant site that support the variant allele. VAF values were plotted against mitochondrial genome position for host samples (see Figure 4-4A) and CTVT tumour samples (see Figure 4-4B). Whereas most mtDNA variants in host samples are homoplasmic, and are present in 100% of reads (Figure 4-4A), CTVT tumours carry variants from both the tumour and from the matched host (Figure 4-4B). Figure 4-4B shows a VAF profile characteristic of CTVT tumours, resulting CTVT tumours being genetically distinct to their hosts (Murgia et al. 2006). Tumour line indicates variants present in the CTVT tumour only, host contamination line indicates variants present in the matched host sample only, tumour + host line indicates variants present in both tumour and host tissue. Heteroplasmy would appear as a tumour variant positioned away from the tumour line (not shown in Figure 4-4B).

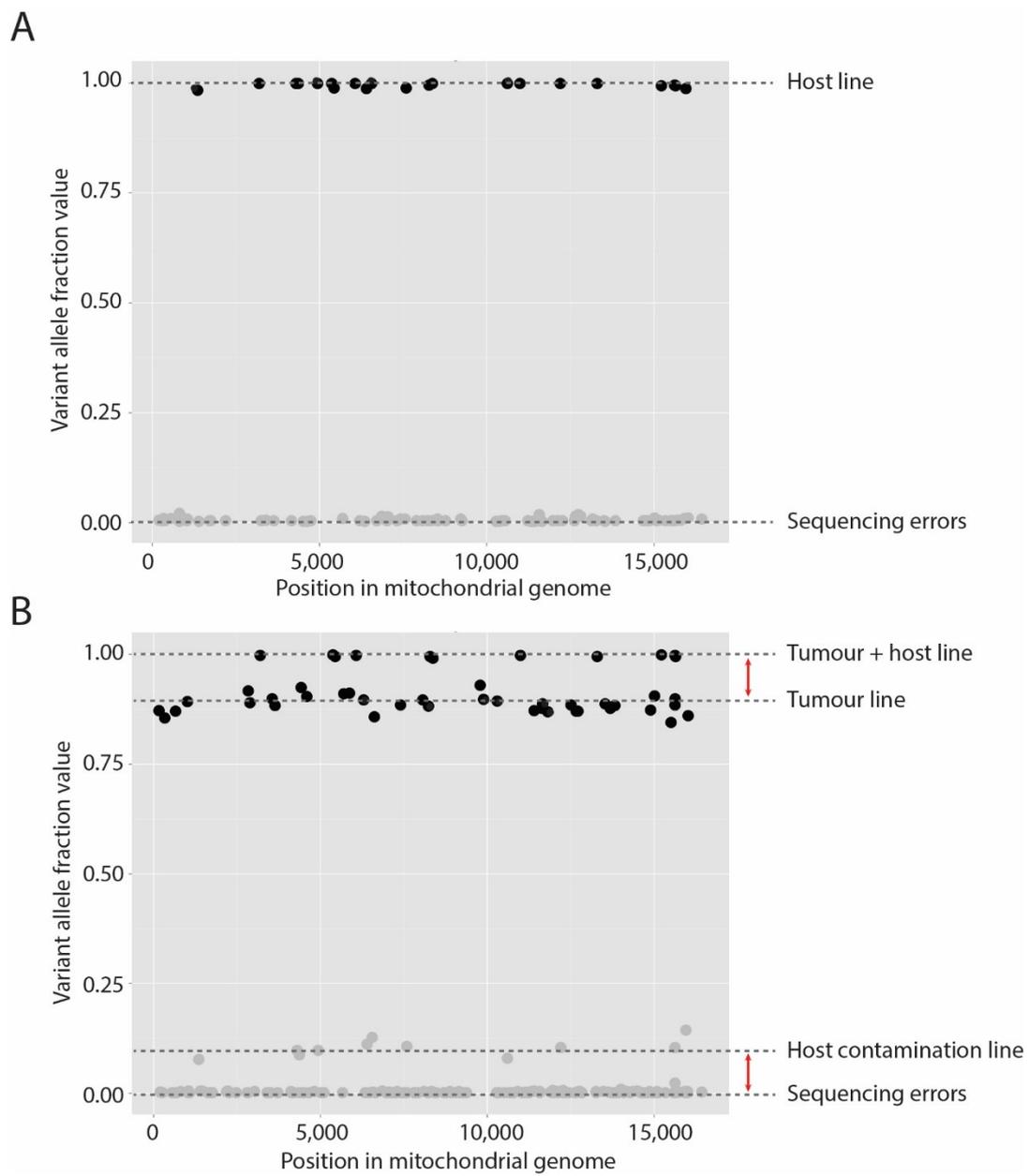


Figure 4-4 Variant allele fraction plot for (A) a host matched to (B) a CTVT tumour. Black dots indicate individual tumour variants. Grey dots indicate sequencing errors in part (A) and individual host variants or sequencing errors (distinguished by labels) in part (B). Dashed lines indicate variants present in tumour only, host only or both (see text for details). Red arrows indicate equal distance between lines. See text for further explanation.

4.2.5.4 *Post-processing*

4.2.5.4.1 *Somatic substitutions in tumours with matched hosts*

To remove tumour variants caused by host contamination, variants that were called in both tumour and matched host, but had VAF<0.9 in the tumour were discarded (see VAF plots in section 4.2.5.3 for visual representation of variants caused by host contamination). Variants with VAF>0.9 in both tumour and matched host were considered to be likely germline variants shared between host and tumour, and were retained (see VAF plots in section 4.2.5.3). Low coverage hosts (defined as average coverage <20X, Additional file 4-2A) were additionally manually scanned for evidence of variants at positions where variants were called in the corresponding tumour, and the variant was discarded in the tumour if at least one read supporting the variant was found in the low coverage host. All tumour variants with VAF<0.5 were discarded if the matched host was of low coverage. Low-level tumour-contaminated hosts were additionally scanned for the presence of substitutions identified in other tumours (see Additional file 4-3) and any variants arising due to contamination were discarded.

4.2.5.4.2 *Somatic substitutions in tumours without matched hosts*

VAF value was used to identify variants likely arising due to host contamination in tumours for which matched hosts were not available (this category included 133 tumours, see Additional file 4-4 for tumours without matched hosts). We used VAF plots, which display VAF value versus genomic position, to identify the level of host contamination in each tumour (for further information about VAF plots, see section 4.2.5.3). We then discarded any variant below a VAF cut-off, specified uniquely for each tumour based on its estimated level of host contamination (for most tumours, the VAF cut-off was 0.5 or 0.6) (see Figure 4-5). If VAF plots did not show clear distinctions between tumour variants and host contamination (i.e. host contamination was > ~40%; this category included 9 tumours), we identified likely tumour variants as those which were present in phylogenetically-related tumours. Remaining variants were removed if they were also found in normal dogs (see “CTVT host and published dog genome germline substitutions” in Additional file 4-3F). Those variants that were not found either in phylogenetically-related tumours or in normal dogs were kept as putative somatic variants (total of 11 variants).

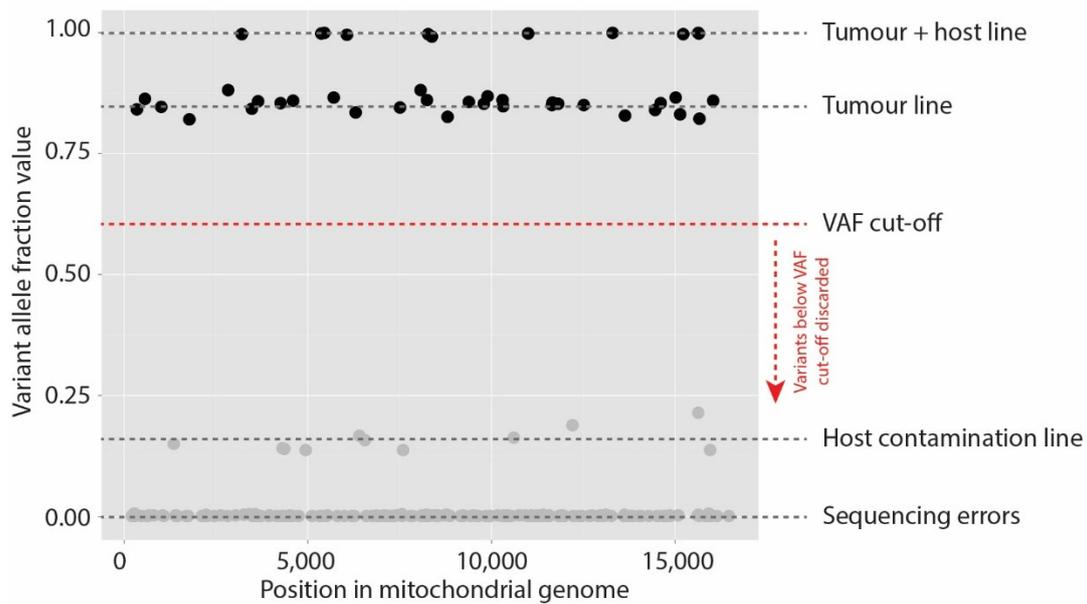


Figure 4-5 Example of VAF plot for tumour without matched host. Black dots represent individual tumour variants, grey dots represent individual host variants or sequencing errors (distinguished by labels). Red line represent VAF cut-off (see text for details). Note the clear distinction between tumour variants and host contamination.

4.2.5.4.3 Germline substitutions in hosts

Substitutions in hosts were filtered as described in section 4.2.5.2. Low coverage hosts (defined as average coverage <20X, Additional file 4-2A) were further screened for evidence of substitutions at positions where a variant was called in the corresponding tumour as described in section 4.2.5.4.1. Low coverage hosts and hosts with regions of low coverage (low coverage defined as <20X, see Additional file 4-2 for average coverage per sample and list of samples with low coverage regions) should be treated with caution, as these may contain false negatives due to low coverage.

4.2.5.5 Additional quality checks and validation

Variants present in low coverage regions, regions with low-mapping quality, and regions containing variable-length polyC homopolymer tracts (Fregel et al. 2015) were subjected to additional quality filtering. Substitutions that were subsequently completely excluded from the analysis as a part of this check are listed in Table 4-1.

Position	Base change	Justification for excluding substitution
15493	G>A	Inconsistently called due to low coverage and decreased mapping quality in this region
15505	T>C	Inconsistently called due to low coverage and decreased mapping quality in this region
15632	C>T	Frequently miscalled due to decreased mapping quality in this region
15639	T>A	Frequently miscalled due to decreased mapping quality in this region
15639	T>G	Frequently miscalled due to decreased mapping quality in this region
15931	A>G	Frequently miscalled because of its presence in the same position as a frequently miscalled indel (middle of a homopolymer tract)
16672	C>T	Frequently miscalled because its proximity to a frequently miscalled indel (associated with a very long homopolymer tract)
16705	C>T	Frequently miscalled due to low coverage in this region

Table 4-1 List of substitutions excluded from the analysis as a result of additional quality checks.

Visual inspection of substitutions discarded due to proximity to an indel was performed using Integrative Genomics Viewer (Robinson et al. 2011, Thorvaldsdottir et al. 2013). A list of substitutions with a substantial support was rescued, and is shown in Table 4-2.

Position	Base change	Position	Base change
381	T>A	8703	G>A
1481	T>C	9825	G>A
1683	T>C	9896	T>C
2682	G>A	13708	C>T
2683	G>A	14977	T>C
3028	A>C	15524	C>T
6629	T>C	15526	C>T
6882	A>G	16660	T>C
7014	T>G	16663	C>T
8281	T>C	16671	T>C
8368	C>T		

Table 4-2 List of substitutions that had substantial support and were rescued after being discarded due to proximity to an indel.

4.2.5.6 *Host contamination*

Host contamination levels were estimated from sample variant allele fraction (VAF) plots (see section 4.2.5.3 for further information on VAF plots). The average VAF of substitutions present in tumours but not in matched host was used to estimate the proportion of tumour mtDNA (Additional file 4-2C). Tumour cell fraction is indicated by the tumour line in Figure 4-4B, and estimates of tumour cell fraction for all CTVT tumours are shown in Additional file 4-2C. To take account of host contamination, the VAF value for all substitutions was normalised based on the estimated proportion of tumour cells within each sample (Additional file 4-3B and C), using the following equation: $VAF/ETCF$, where VAF=VAF value for substitution and ETCF=estimated tumour cell fraction for each tumour. Tumour variants with normalised $VAF < 1$ most likely represent heteroplasmic variants; however, we cannot exclude the possibility that these represent cellular subclones harbouring distinct homoplasmic mtDNA populations.

4.2.5.7 *Recurrent mutations and back-mutations*

Back mutations (somatic mutations back to reference base) and recurrent mutations (independent somatic events) occurring in tumours were identified by inspecting phylogenetic positions of tumours carrying each substitution on phylogenetic trees (see section 4.2.7 for further information on phylogenetic trees). Phylogenetically independent events were identified as back mutations and recurrent mutations.

4.2.5.8 *Predicted functional consequences of substitutions*

Variant effect predictor (VEP) (McLaren et al. 2010) was used to predict the functional effect of single point substitutions, as presented in Additional file 4-5.

4.2.5.9 *Extracting substitution variants from publicly available dog sequences*

In order to enrich our panel of germline substitutions, obtained from 338 host samples, and to improve our ability to detect host contamination, we included additional 252 domestic dog mtDNA sequences publicly available from NCBI (see Additional file 4-6).

Substitution variants were extracted from the publicly available sequences by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group). Sequence fasta files were aligned with the CanFam3.1 dog mitochondrial reference genome using Clustal Omega (Sievers et al. 2011). Manual inspection of alignment errors in the multiple sequence

alignment (usually due to miscalls caused by closely mapped indels) was performed and errors were corrected to minimise gaps. Substitutions were extracted using snp-sites (https://github.com/sanger-pathogens/snp_sites). In samples missing data in regions MT: 15510–15532 and MT: 16040–16550 we substituted the most likely substitution at polymorphic sites based on phylogenetic position of the sample. Our filtering rules were applied to the substitution set where applicable (see section 4.2.5.2 and section 4.2.5.5). Substitutions called before MT position 48 or after MT position 16671 were excluded due to low coverage. Substitutions represented by International Union of Pure and Applied Chemistry (IUPAC) codes R, Y, S, W, K or M where one of the two possible calls was the same as the reference were changed to the base which was different to the reference. In cases where the IUPAC code represented >2 bases (B, D, H, V, or N), the reference base was substituted.

4.2.6 Indel calling

4.2.6.1 *Extraction and filtering*

Insertions and deletions (indels) were extracted from the sequencing data by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group) using cgpPindel (<https://github.com/cancerit/cgpPindel>). The following list of in-built filters was used to improve the specificity and sensitivity of indel calls:

- (1) Indels were required to have at least 3 or more supporting reads on either the forward or reverse strands or at least 2 supporting reads or more on both the forward and reverse strands
- (2) Indel calls with at least 4 supporting pindel-mapped reads were required to have at least 1 supporting BWA-mapped read or, failing that, if REP value = 0 (see section 4.2.5.2), then at least one supporting pindel-mapped read on both strands.
- (3) Indels called within the simple repeat region were excluded (16129bp-16430bp inclusive, based on the simple repeat region as defined by the UCSC (<http://genome.ucsc.edu/>) table browser (Dog, CanFam3.1)

Samples with extremely high coverage of the mitochondrial genome (24T-Dog, 24H-Dog, 1T-Dog, 2T-Dog, 3T-Dog, 4T-Dog, 4H-Dog, 498H-Dog, 432T-Dog, 455T1-Dog, 231T-Dog, 79H-Dog, 79T-Dog, see Additional file 4-2A) were excluded from the analysis, due to high numbers of

false positives in calling indels in these samples. Samples 1380T-Dog and 1381T-Dog were also excluded from the analysis, due to being sequenced separately.

4.2.6.2 Post processing

4.2.6.2.1 Somatic indels

The somatic list represented indels found in tumour samples. Indels that were called in tumours and in at least one host were discarded from the somatic list, due to high possibility of being caused by host contamination. Indels present in unmatched tumour samples were discarded, as we could not rule out the possibility that they were caused by host contamination. All remaining indels were visually validated using Integrative Genomics Viewer (Robinson et al. 2011, Thorvaldsdottir et al. 2013), and as a result indels listed in Table 4-3 were discarded from the analysis as miscalls. In total 27 somatic indels were included in the analysis (see Additional file 4-7A).

Position	Indel	Sample	Justification for discarding indel
9891	TGATTTATCTCATAATTATCA> TATCTCATAATTATCATG	324T2-Dog	Indel miscalled due to close proximity of two other indels in the same sample
9910	C>CAT	401T-Dog	Indel miscalled, due to presence of a substitution at the same position

Table 4-3 Somatic indels discarded from the analysis as miscalls.

4.2.6.2.2 Germline indels

Only homoplasmic host indels were used in the analysis where homoplasmic indels were defined as any indel with VAF \geq 0.9. Only homoplasmic germline indels that were visually validated using Integrative Genomics Viewer (Robinson et al. 2011, Thorvaldsdottir et al. 2013) were included on the host variants list (see Additional file 4-7B).

4.2.6.3 Recurrent indels

Recurrent indels (independent somatic events) occurring in tumours were identified by inspecting positions of tumours carrying each indel on phylogenetic trees. Phylogenetically independent events were identified as recurrent indels.

4.2.6.4 *Variant Allele Fraction (VAF) calculation for indels*

Although primary indel calling was carried out using cgppindel, indel allele fraction for wild type and mutant indels was calculated using vcfCommons (unpublished in-house software developed at the Wellcome Trust Sanger Institute) by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group). The algorithm takes all mapped and unmapped reads in the region of an indel and performs an alignment of the reference and the predicted mutated path using Exonerate (Slater and Birney 2005).

Indel VAF values were normalised to take account of host contamination in Figure 4-18B.

4.2.6.5 *Predicted functional consequences of indels*

Variant effect predictor (VEP) (McLaren et al. 2010) was used to predict the functional effect of indels, as presented in Additional file 4-8.

4.2.7 Phylogenetic analyses

4.2.7.1 *Phylogenetic trees*

Phylogenetic trees were constructed by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group). A maximum likelihood (ML) method was used and implemented in PhyML 3.0 (Guindon et al. 2010) using the General Time Reversible (GTR) + G + I nucleotide substitution model with transition/transversion (ts/tv) ratio, gamma distribution shape parameter and proportion of invariant sites estimated in PhyML for each dataset. Phylogenetic trees constructed using RAxML 8.2.6 with a rapid hill climbing algorithm and identical model specification were consistent with the phylogenetic trees obtained using PhyML. jModelTest2 (Darriba et al., 2012) (<https://code.google.com/p/jmodeltest2>) was used prior to tree construction to determine the best fitting substitution model for the alignment, inferred using the Akaike Information Criteria (AIC). The tree topology was estimated using a combination of Nearest Neighbor Interchange (NNI) and Subtree Pruning and Regrafting (SPR) (Hordijk and Gascuel 2005) algorithms. Dendrocope (Huson and Scornavacca 2012) was used to visualise the trees. Bootstrap support values were obtained from 100 bootstrap replicates. Nodes with $\geq 60\%$ support are labelled in Additional file 4-9 and Additional file 4-10. The long branch-lengths in CTVT clade 4 were further investigated for possibility of errors by visually validating sequence alignments involving these samples.

4.2.7.2 Confirmation of individual horizontal transfer events

4.2.7.2.1 Classification of tumour substitutions

CTVT substitutions in tumours were classified as shown in Table 4-4 (see Figure 4-6 for illustration, see Additional file 4-3A for the full lists of tumour substitutions).

Note that another list of more stringently chosen somatic substitutions was created, called **Tumour somatic substitutions - conservative list** (see Additional file 4-3C for the full list) and was used for analysis in Figure 4-18A, Figure 4-19. The criteria for this list were the same as for Tumour somatic substitutions list (see Table 4-4), but any substitutions which were classed as somatic but had the potential to be germline were excluded from the conservative somatic list, as described below:

- Association with inferred recombination events
- Present uniquely in clade 3; clade 3 tumours carry very few somatic mutations, and so we cannot exclude the possibility that clade 3 tumours arose from several independent mtDNA horizontal transfers
- Ancestral trunk mutations; these include trunks defining CTVT_1A, CTVT_1B1, CTVT_2A haplogroups - due to their ancestral divergence, it cannot be excluded that these haplotypes are derived from independent mtDNA horizontal transfer events
- Present as possible somatic mutations in tumours without hosts (see section 4.2.5.4.2).

	Origin and characteristics	Definition	Additional notes
Tumour germline clade-defining substitutions	Present on the reconstructed donor mtDNA haplotype which founded each of the five horizontal transfer events	Shared by all tumour samples within each clade derived from the original CTVT cell which received the horizontally transferred donated mtDNA	Present within the pool of substitutions in the current domestic dog population (see Additional file 4-3D)
Tumour somatic substitutions	Arose after the clade-defining horizontal transfer event	Variable and phylogenetically informative within the set of tumours in one clade (see Additional file 4-3B)	Any substitutions which were associated with inferred recombination event (see section 4.2.9) and uniquely present in 559T were discarded
Tumour potential somatic substitutions	Substitutions where it cannot be determined whether they arose before the horizontal transfer event (i.e. germline) or very soon after the horizontal transfer event but before the divergence of tumours within that clade (i.e. early somatic)	Common to all tumour samples within one clade, but not observed in the domestic dog population (590 dogs) analysed as part of this study (see Additional file 4-3E)	

Table 4-4 Classification of tumour substitutions into (1) Tumour germline clade-defining, (2) Tumour somatic and (3) Tumour potential somatic. See Figure 4-6 for illustration. See Additional file 4-3 for the full lists of tumour substitutions.

4.2.7.2.2 *Definition of a CTVT clade*

A CTVT clade is defined as a group of tumour samples arising from a single horizontal transfer event, using the following criteria:

- (1)** CTVT tumours from the same clade cluster very closely on a phylogenetic tree constructed using normal canine mtDNA (see Figure 4-13).
- (2)** CTVT tumours from the same clade arose from a single donor mtDNA haplotype and therefore share the same set of germline substitutions which was originally present in the donor mtDNA (see Figure 4-14, section 4.2.7.2.1).
- (3)** The number of somatic mutations (i.e. mutations arising after the horizontal transfer event) is consistent across tumour samples from a single clade, indicating that the time since the horizontal transfer event took place is comparable across the clade (see Figure 4-15).
- (4)** The reconstructed donor mtDNA haplotype for each clade has a phylogenetically closely related/identical haplotype in the current dog population (see Figure 4-14).

4.2.7.3 *Estimated timing of clade divergence*

Average number of somatic mutations (i.e. mutations arising after the horizontal transfer event) within each clade was calculated for each of the five CTVT clades (see Table 4-11 and Table 4-12). For potential somatic mutations (shown in grey in Figure 4-15 and Figure 4-17), we were unable to determine whether they occurred before or after the horizontal transfer event, as they are present in all samples from the same clade, but not in the pool of germline variants (see Additional file 4-3E). As the number of somatic mutations influences the time of divergence, we included estimates both with and without potential somatic mutations in our analysis. Presence of mitochondrial recombination in haplogroup CTVT_1B2b2 (clade 1) was taken into account when calculating the average number of mutations (see section 4.2.9). The timing of clade divergence was estimated independently based on the following three methods, as described below: timing based on nuclear DNA (Alexandrov et al. 2013, Murchison et al. 2014), number of cell divisions per homoplasmic mitochondrial mutation (Cohen and Steel 1972, Ju et al. 2014) and number of mitochondrial mutations per year (Ju et al. 2014). All estimates of CTVT timing assume a constant rate of accumulation of somatic mtDNA mutations both within and between clades, including constant activity of selection (Figure 4-18 and Figure 4-19).

4.2.7.3.1 Timing based on nuclear DNA

Based on the most recent common ancestor of samples 24T (CTVT clade 1) and 79T (CTVT clade 2) which was estimated to have existed 460 years ago (Murchison et al. 2014), we can assume that the maximum time since the mtDNA horizontal transfer event that defined clade 2 (i.e. the more recent clade) is 460 years. Calibrated according to this dating (Alexandrov et al. 2013, Murchison et al. 2014), and assuming a constant rate of accumulation of mutations with time, the maximum number of CTVT somatic mtDNA mutations per year is 0.0205 (calculated using the following equation: average number of mutations in clade 2 samples/age of clade 2 = 9.437/460). The calculated ages of clades 1, 3, 4 and 5, based on the average number of substitutions per clade, are shown in Figure 4-17, Table 4-11 and Table 4-12.

4.2.7.3.2 Timing based on number of cell divisions

A recent study of mtDNA mutations in cancer found that one homoplasmic mutation arises every ~1000 cell generations (Ju et al. 2014). Moreover, an experimental CTVT study estimated CTVT cell division rate to be 4 days in first stage and >20 days in second stage tumours (Cohen and Steel 1972). Using the data presented above and assuming a constant accumulation of somatic mutations in CTVT, we estimated a minimum and maximum mutation rate of ~ 0.0183 and ~ 0.0913 mutations/year respectively. The corresponding age calculations for each clade 1, 2, 3, 4, and 5 are displayed in Table 4-11 and Table 4-12.

4.2.7.3.3 Timing based on number of mutations per year

A previous study indicated that the rate of accumulation of mtDNA mutations in human cancers is about ~0.75 mutations per 30 years (Ju et al. 2014). Extrapolating this mutational rate to CTVT, we estimated the accumulation of 0.025 mutations per year. Assuming a similar rate in CTVT somatic mtDNA mutations and a constant accumulation of mutations in CTVT, the estimated ages of clades 1, 2, 3, 4 and 5 are presented in Table 4-11 and Table 4-12.

4.2.8 Haplotype analysis

4.2.8.1 Haplotype nomenclature

4.2.8.1.1 Host haplotypes

The host haplotype naming system used in this Chapter is adapted from the cladistic canine mitochondrial DNA phylogeny nomenclature proposed by Fregel et al (Pereira et al. 2004,

Fregel et al. 2015). All host haplotype names are presented in Additional file 4-4 and Additional file 4-11. Corresponding CTVT normal dog host samples were assigned into one of the major clades (A, B, C, D, E and F). For subsequent levels, haplogroups were defined by specific diagnostic variants, as defined by Fregel et al (Fregel et al. 2015). A unique number, following an underscore after the haplogroup name, distinguished haplotypes within each haplogroup. Haplogroup defining variants 15632 C>T, 15639 T>A and 15639 T>G were excluded from our analysis and therefore we were unable to distinguish between haplogroups A1, A1c and A1e (these haplogroups are therefore referred to as A1/A1c/A1e). Any haplotypes of hosts with low coverage (defined as average coverage <20X, Additional file 4-2A) and low coverage regions (see Additional file 4-2B) were individually assessed and assigned to haplogroups.

4.2.8.1.2 Tumour haplotypes

The haplotype naming system used is adapted from the cladistic canine mitochondrial DNA phylogeny nomenclature proposed by Fregel et al (Fregel et al. 2015). All CTVT tumour haplotype names are presented in Additional file 4-4 and Additional file 4-11. Each CTVT haplotype has a prefix “CTVT_”, indicating a tumour haplotype. Five CTVT clades which have been assigned names are numbered (CTVT_1, CTVT_2, CTVT_3, CTVT_4 and CTVT_5). For subsequent levels, hierarchical notation is used, where subsequent haplogroups are named by alternating letters and numbers and the maximum number of levels included in the hierarchical notation is five – i.e. 3 numbers and 2 letters (e.g. 1A1a1). The first letter is a Roman capital; subsequently used letters are lower case Roman letters. Any haplogroups beyond the maximum number of levels are considered as a single subgroup, in which individual haplotypes are distinguished using a non-hierarchical numbering system - an underscore followed by a number (e.g. 1A1a1_1, 1A1a1_2, etc.). Underscores are only used to distinguish individual haplotypes after the haplotype has been assigned to all 5 hierarchical levels (e.g. 1A_1 does not exist, as this haplotype would be classified as 1A1 instead).

4.2.8.2 Reconstructed donor haplotypes

A “donor haplotype” was reconstructed for each of the clades, representing the inferred donor mtDNA in each horizontal transfer event, and was used to root the trees for each clade in Additional file 4-9. The donor haplotypes were reconstructed from the clade-defining germline variants and the clade-defining potential somatic variants (see section 4.2.7.2.1 for further details about variant classification) and are shown in Figure 4-14. The clade-defining germline variants represent substitutions present on the reconstructed donor mtDNA

haplotype, which founded each of the five horizontal transfer events; they are shared by all tumour samples within each clade and are present in the pool of substitutions in the current domestic dog population. The clade-defining potential somatic variants are substitutions where it cannot be determined whether they arose before the horizontal transfer event (i.e. germline) or very soon after the horizontal transfer event but before the divergence of tumours within that clade (i.e. somatic); they are common to all tumour samples within one clade, but not observed in the domestic dog population. The phylogenetically closest haplotype present in the current canine population is shown in the same figure.

4.2.9 Mitochondrial recombination analysis

4.2.9.1 *Automated mtDNA recombination analysis*

The automated mtDNA recombination analysis was performed by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group). The RDP4 package (Martin et al. 2015) was used to detect recombination events within the complete sample set of CTVT samples, matched host samples and published sequences (449 CTVT tumours, 338 CTVT hosts and 252 additional dogs) using a Bonferroni corrected p-value cut-off of 0.05. The following programs were implemented within the RDP package (using default parameters): RDP (Martin and Rybicki 2000), MaxChi (Smith 1992), Chimaera (Posada and Crandall 2001), 3Seq (Boni et al. 2007) and SiScan (Gibbs et al. 2000).

4.2.9.2 *Analysis of phylogenetic outliers*

We searched the phylogenetic trees for tumours with unusually long branch length appearing as unexplained outliers on the mtDNA phylogenetic trees. Tumours identified using this method were scanned for abnormally high numbers of apparent back-mutations and variants with unexpected VAF values.

4.2.9.3 *Haplotype cloning and sequencing*

Polymerase chain reaction was performed on an MJ Research Tetrad PTC-225 Peltier Thermal Cycler, spanning a region between MT positions 4450-5024. Primers and conditions are specified in Table 4-5, Table 4-6 and Table 4-7. PCR products were used for a TOPO TA cloning reaction according to the manufacturer's protocol (Thermo Fisher Scientific). The plasmid resulting from the TOPO cloning reaction was transformed into DH5 α competent cells (Thermo Fisher Scientific) and cells were cultured overnight. DNA from individual colonies was purified using the QiaPrep MiniPrep Kit (Qiagen) and sequenced using capillary

sequencing (Source BioScience). Capillary sequencing traces were analysed for presence of variants at MT positions 4591 and 4940 using BioEdit Sequence Alignment Editor (Hall 1999).

Primer		Sequence
MT-recombination-short primers	Forward	AGC ATA CTC CTC CAT TGC CC
	Reverse	TCA GGG GTT AAG TGT GTA GCA

Table 4-5 Primers used for PCR reaction spanning a region between MT positions 4450-5024.

PCR master mix reagents	Volume per reaction (µl)
Platinum™ Taq DNA Polymerase (Invitrogen™)	0.08
10X PCR buffer (-Mg)	2
MgCl ₂ (50mM)	0.6
dNTP mix (2.5mM each)	1.6
Primer forward (10µM)	0.4
Primer reverse (10µM)	0.4
DNA (~40ng/µl)	2
Distilled water	12.92
Total volume	20

Table 4-6 Input reagent volumes and concentrations for PCR reaction spanning a region between MT positions 4450-5024.

Stage of qPCR amplification	Temperature (°C)	Time (s)
Initial denaturation	94	120
34 cycles	94	15
	60	30
	72	30
	72	300
Final dissociation	72	300

Table 4-7 Cycling conditions for PCR reaction spanning a region between MT positions 4450-5024.

4.2.9.4 PacBio long-read sequencing and analysis

A genomic library was created directly using 5µg of genomic DNA from sample 559T, not utilizing shearing or amplification techniques. The library was sequenced using two PacBio SMRT cells with the Pacific Biosciences RS sequencer (Pacific Bio-sciences, Menlo Park, CA).

Each out of two SMRT cells yielded ~ 1Gb of sequence data with mean read length 11,421bp and N50 read length 19,382bp. Average sequence coverage across the mitochondrial genome was 111.3X. PacBio sequence reads aligning to mtDNA were viewed in SMRT view (Pacific Biosciences) as well as in Integrative Genomics Viewer (IGV) (Robinson et al. 2011, Thorvaldsdottir et al. 2013) and used to phase the mitochondrial substitutions previously called in 559T using Illumina sequencing data. The three most common recombinant haplotypes in 559T were completely phased, as shown in Figure 4-23. Additional haplotypes, which we were unable to phase completely and which were present at very low level (less than 5%), were also identified. The reads used to phase the substitutions in the most common haplotypes in tumour 559T are shown in Table 4-8:

559T haplotype 1	559T haplotype 2	559T haplotype 3
m150625_205815_00127_c100 80924255000000182317731008 1544_s1_p0/80046	m150623_234631_00127_c10 078774255000000182317300 8251557_s1_p0/328	m150625_205815_00127_c100 8092425500000018231773100 81544_s1_p0/64645
m150625_205815_00127_c100 80924255000000182317731008 1544_s1_p0/50141	m150623_234631_00127_c10 078774255000000182317300 8251557_s1_p0/89158	m150625_205815_00127_c100 8092425500000018231773100 81544_s1_p0/29160
m150623_234631_00127_c100 78774255000000182317300825 1557_s1_p0/32367	m150625_205815_00127_c10 080924255000000182317731 0081544_s1_p0/45213	Many reads in the region between 9790-16627
m150623_234631_00127_c100 78774255000000182317300825 1557_s1_p0/145458		

Table 4-8 PacBio sequencing reads used for phasing of individual haplotypes in tumour 559T.

4.2.10 Selection analyses

4.2.10.1 Variant allele fraction

4.2.10.1.1 Substitutions

Normalised variant allele fraction (VAF) value for somatic substitutions was calculated as described in section 4.2.5.6. Cumulative distributions of normalised VAF scores (i.e. levels of heteroplasmy) were plotted for nonsense substitutions (n = 10) and missense and synonymous substitutions (n = 610). Statistical significance testing was performed using the two-sample Kolmogorov-Smirnov test implemented in R (R Core Team 2013).

4.2.10.1.2 *Indels*

Normalised variant allele fraction (VAF) value for somatic indels was calculated as described in section 4.2.6.4. Cumulative distributions of normalised VAF scores (i.e. levels of heteroplasmy) were plotted for frameshift ($n = 18$) and non-frameshift ($n = 9$) indels. Statistical significance testing was performed using the two-sample Kolmogorov-Smirnov test implemented in R (R Core Team 2013).

4.2.10.2 *dN/dS*

The dN/dS ratio was calculated according to a method adapted from (Martincorena et al. 2015) by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group). The analysis was restricted to conservative somatic substitutions only (see section 4.2.7.2.1). Synonymous substitutions were used to estimate the background mutation rate.

4.2.11 Nuclear copies of mtDNA (NuMT) analysis

Nuclear copies of mtDNA (NuMTs) are mtDNA fragments that have been incorporated into the nuclear genome. Over 150 NuMTs have previously been identified in the canine genome (Verscheure et al. 2015). Somatically acquired NuMTs have also been described before in human cancer (Ju et al. 2015). Given that our study design did not involve purification of cytoplasmic mtDNA genomes, we assessed the possibility that our mtDNA variant analysis has been influenced by presence of NuMTs as follows.

4.2.11.1 *NuMTs in CanFam3.1*

Assessment of the potential contribution of NuMTs present within the CanFam3.1 assembly to our variant calling was performed by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group). Wgsim (<https://github.com/lh3/wgsim>) was used to simulate reads from CanFam3.1 genome (excluding the MT chromosome) to sequencing coverage of 0.3X (i.e. the average nuclear genome coverage sequenced as part of this study). BWA (Li and Durbin 2009) was then used to map the simulated reads against the CanFam3.1 MT reference and Samtools depth (Li et al. 2009, Li 2011) was utilised to measure MT coverage. Any MT genome coverage detected from this analysis would be expected to arise from NuMTs. The average MT genome coverage from this analysis was 0, indicating that the NuMTs known to be present within CanFam3.1 are too divergent to map to the MT reference genome using the alignment parameters used in this study.

4.2.11.2 *Somatically acquired NuMTs*

Based on the analysis above (see section 4.2.11.1) we cannot exclude that somatically acquired NuMTs that are not in the CanFam3.1 genome assembly could confound our variant analysis. The following observation, however, argues against this possibility - as CTVT is a clonal lineage, somatically acquired NuMT-derived variants would be expected to present as stable low-VAF variants across all tumours within a phylogenetic group. Variants with these features were not observed, suggesting that NuMT-derived variants did not have a significant impact on our tumour variant calling. However, we cannot exclude the possibility that low-VAF variants specific to a single tumour could be derived from recently somatically acquired NuMTs.

4.2.12 Histopathology analysis

A blinded histopathology scoring study was performed to analyse phenotypic differences between CTVT clades 1 and 2. 253 tumours were included in this study, 93 clade 1 tumours and 160 clade 2 tumours. All histology samples available at the time were used for the study with the intention to perform a pilot screen. Tumour slides were randomised and visually validated to confirm presence of CTVT cells. Histology slides were scanned using the Hamamatsu Nanozoomer 2.0-HT slide scanner (C9600) at the Wellcome Trust Sanger Institute, Hinxton and visualised for scoring using the NDP.view2 software (Hamamatsu). A proportion of histopathology slides (111 tumours) were double-scored by Isobelle Bolton (Part II pathology veterinary student in the Transmissible Cancer Group) and myself to test reproducibility of results. The remaining slides (142 tumours) were scored by Isobelle Bolton (Part II pathology veterinary student in the Transmissible Cancer Group).

4.2.12.1 *Histopathology screen design*

Histopathology screen design was based on previous literature (see section 4.1.5), discussions with veterinary pathologists Dr Katherine Hughes (Department of Veterinary Medicine, University of Cambridge) and Dr Hannah Bender (Murdoch University, Western Australia) and pilot screening. The scoring parameters were developed with an emphasis on making biologically relevant deductions from the histopathological features of tumours. The following list of parameters was chosen for the histopathology screen: normal mitotic index, bizarre mitotic index, lymphocyte density in parenchyma, lymphocyte density at interface, haemorrhage and necrosis.

4.2.12.2 Scoring criteria

4.2.12.2.1 Selecting a field

- At least half of the field must consist of CTVT parenchyma (an exception may be made when scoring lymphocyte percentage, but the field must be within the tumour parenchyma)
- Avoid scoring in areas of artefactual distortion or near ulceration and haemorrhage
- Avoid overlap between fields through use the NDP.view2 software navigation map

4.2.12.2.2 Scoring mitotic index

- Use NDP.view2 software to create a mitotic box of area 0.0325cm³, perimeter 722µm
- Select a random field at 40X magnification
- Insert the mitotic box and count the total number of normal mitoses present. Note as a mitotic figure if: no nuclear membrane, fuzzy outline, evident cytoplasm (refer to examples of mitotic figures in Figure 4-7 below). Do not score if in fibrous trabeculae. Do score if on the boundary of the mitotic box.
- Count the total number of bizarre mitoses present. Follow criteria for a mitotic figure as above; count as bizarre if tripolar or otherwise misshaped - refer to examples in Figure 4-7.
- Repeat for 9 other fields, using the NDP.view2 navigation map and move systematically across the tumour in a staircase fashion.

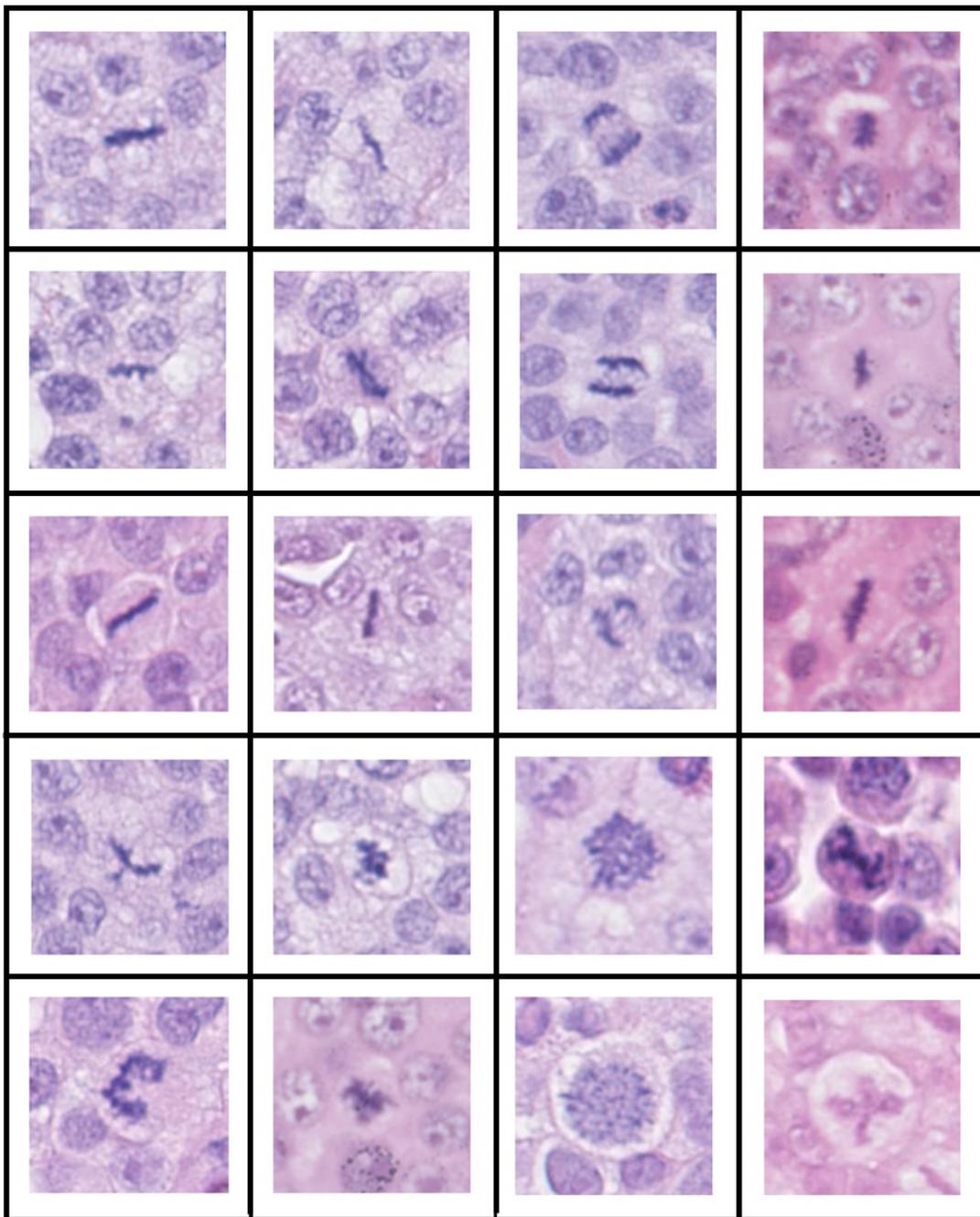


Figure 4-7 Examples of mitotic figures to be counted in the mitotic figure scoring system. First three rows show examples of normal mitotic figures, last two rows show examples of bizarre mitotic figures. Photos were taken by Isobelle Bolton (Part II pathology veterinary student in the Transmissible Cancer Group).

4.2.12.2.3 Scoring percentage of lymphocyte infiltration in parenchyma

- Overlay a 14cm (vertical, labelled 1-14) by 18cm (horizontal, labelled A-R) grid (numbers and letters start in bottom right hand side corner) on the NDP.view2 navigation map
- Select a 20X field in the centre of grid box identified as G8 and estimate the percentage lymphocyte cellularity (higher magnification may be used to confirm presence of lymphocytes), choosing from following categories: 0%, 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 95%.
- Using the coordinates listed below, note the percentage lymphocyte cellularity for further four fields – if the grid box contains no CTVT cells, or is not scorable based on the field criteria above, then move to the next coordinate. If after 20 coordinates (listed below) five suitable fields have not been obtained, select a field directly within the tumour parenchyma.

Order or coordinates for scoring percentage lymphocyte cellularity									
G8	C6	E9	I5	O7	B2	F7	N8	K6	D4
J9	M8	G5	I3	N11	H6	F5	I10	C4	H11

4.2.12.2.4 Scoring percentage of lymphocyte infiltration at interface

- If present, identify the boundary between tumour and surrounding stromal tissue. If boundary not present, do not score.
- Select a 20X field, positioning so the interface between tumour and stromal tissue is approximately central, and estimate the percentage lymphocyte cellularity (higher magnification may be used to confirm presence of lymphocytes), choosing from following categories: 0%, 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 95%.
- Move field at evenly spaced intervals along the tumour-stromal tissue interface, and note the percentage lymphocyte cellularity for further four fields.

4.2.12.2.5 Scoring haemorrhage

- Scan across the whole sample at 2.5X magnification to identify regions of haemorrhage (higher magnification may be used to confirm presence of haemorrhage).

- Score as absent or present. Refer to Figure 4-8 for examples of haemorrhage. Only score haemorrhage within the tumour parenchyma. Do not score sampling induced haemorrhage (at the surface of the tumour).

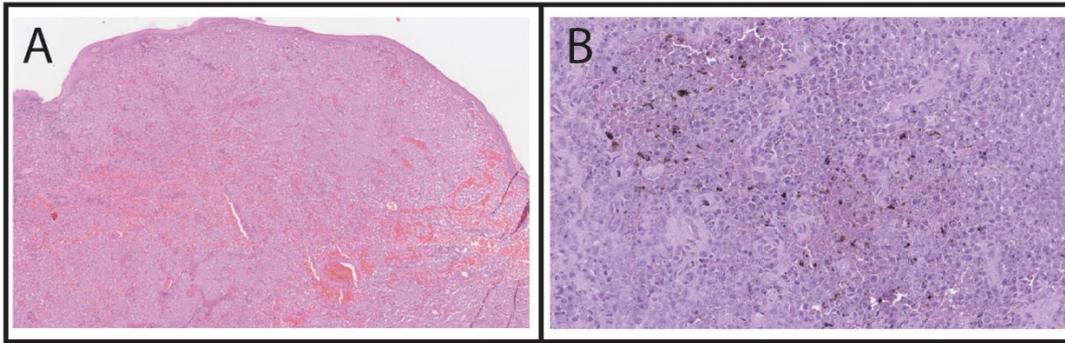


Figure 4-8 Examples of haemorrhage within the tumour parenchyma. (A) Intra-tumoural haemorrhage (magnification 2.5X) (B) Haemosiderin indicating old haemorrhage (magnification 20X). Photos were taken by Isobelle Bolton (Part II pathology veterinary student in the Transmissible Cancer Group).

4.2.12.2.6 Scoring necrosis

- Scan across the whole sample at 2.5X magnification to identify regions of necrosis (higher magnification may be used to confirm presence of necrosis).
- Score as absent or present. Refer to Figure 4-9 for examples of necrosis (characterised as an amorphous, eosinophilic region with fragmented nuclear remains (karyorrhexis) and infiltration of neutrophils and macrophages). Do not score necrosis if present under an ulcerated epithelium.

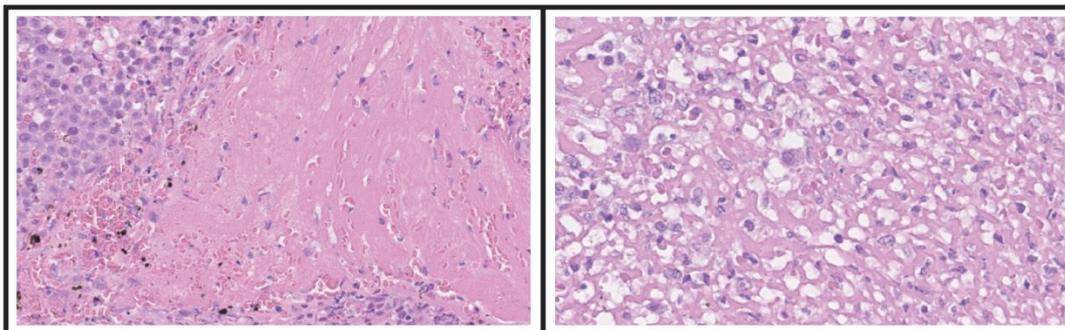


Figure 4-9 Examples of necrosis (magnification 40X). Photos were taken by Isobelle Bolton (Party II pathology veterinary student in the Transmissible Cancer Group).

4.2.12.3 Statistical analysis

A generalised linear mixed effect model (GLMER) utilised in R (R Core Team 2013) was used to perform statistical analysis of the histopathological data and enabled inclusion of a number of confounding factors. The following parameters were analysed using the model: total mitotic index; normal mitotic index; bizarre mitotic index; bizarre mitoses as a proportion of total mitoses; parenchymal lymphocyte density. Clade was imputed as the fixed effect, and the following confounding factors were imputed as random effects: host age, host gender, tumour ulceration and sample (to avoid sample quality bias). A Poisson distribution was used in the GLMER model, based on statistical advice from Dr David Wedge, Wellcome Trust Sanger Institute. The R command used to utilise the model is shown in Figure 4-10.

```
#### GLMER model ####
# Parameters:
  - total mitotic index
  - normal mitotic index
  - bizarre mitotic index
  - bizarre mitoses as a proportion of total mitoses
  - parenchymal lymphocyte density
# Fixed effect: Clade
# Random effects: Sample/Gender/Age/Tumour ulceration
# Family: Poisson

glmer_name <- glmer (Parameter ~ Clade
+(1|Sample)+(1|Gender)+(1|Age)+(1|Tumour_ulceration),
input_data, family = "poisson")
```

Figure 4-10 Command to utilise GLMER model in R (R Core Team 2013).

Lymphocyte density at interface was discarded as a parameter from the statistical analysis due to very low number of samples covering an interface (5 out of 253 samples). The GLMER model reports the effect size as $\ln(\text{ratio_to_baseline})$, with clade 1 defined as the baseline. GLMER output estimate (see Table 4-13) generated is defined as $\ln(\text{parameter_clade2}/\text{parameter_clade1})$; exponentiating the estimate gives the factor by which clade 2 differs from clade 1 (see Table 4-13).

4.2.12.3.1 Random effects

Random effects used in the GLMER model were included to take account of the following confounding factors: host age, host gender, tumour ulceration and scoring slide (to avoid sample quality bias) (see Table 4-9, Table 4-10). Age data were categorised into the following three categories: young, adult and old. After a discussion with Dr Mirjam van der Wel (veterinarian at the Animal Anti-Cruelty League Port Elizabeth, South Africa), the age categories were defined as follows, to represent the age demographic structure of dog populations in the developing world (where most samples were collected): young ≤ 1 year, adult >1 to <7 years, old ≥ 7 years.

Parameter		Number of cases
Total number of cases		253
Categorical Age	Young	8
	Adult	184
	Old	24
	Unknown	37
Gender	Male	103
	Female	142
	Unknown	8
Tumour Ulceration	Yes	124
	No	82
	Unknown	47

Table 4-9 Summary of random effects and number of cases within each category.

Random effect	Reasoning for included as random effect
Host age	Immuno-naivety and immuno-senescence of young and old dogs respectively may lead to a lower lymphocyte count and consequently may result in more aggressive tumours.
Host gender	Different hormone profiles between males and females could may influence host-tumour interaction between sexes.
Tumour ulceration	Ulceration is a sign of bacterial infection, which may have an impact on the immune infiltrate in tumour. Chronic infection may lead either to higher lymphocyte counts or immune exhaustion, thus impacting the immune infiltrate.

Table 4-10 Summary of random effects and reasoning for why they were included in the screen.

4.3 Results

4.3.1 Samples analysed in this Chapter

The number of samples submitted for DNA sequencing was 821 tumour and matched host samples. Out of these, 34 samples were discarded from the analysis for the following reasons: low DNA sequencing quality and low sequencing coverage (see Materials and Methods, section 4.2.3). The final set of samples used for analysis consisted of 449 confirmed CTVT tumours and 338 matched hosts (Additional file 4-4).

4.3.2 Sequencing of mtDNA

To investigate the global CTVT population structure and estimate the frequency and timing of mtDNA horizontal transfer, we analysed mitochondrial DNA (mtDNA) sequencing data from 449 CTVT tumours and 338 matched hosts from 39 countries around the world (see Figure 4-11, Additional file 4-4). The DNA was sequenced at $\sim 0.3X$ whole genome coverage, leading to $\sim 70X$ coverage across mtDNA and therefore indicating that each CTVT cell carries approximately 470 mtDNA copies (see Figure 4-12). The host tissue samples included skin, gonad and blood samples, thus explaining the great variability in mtDNA copy number in host samples (Figure 4-12).

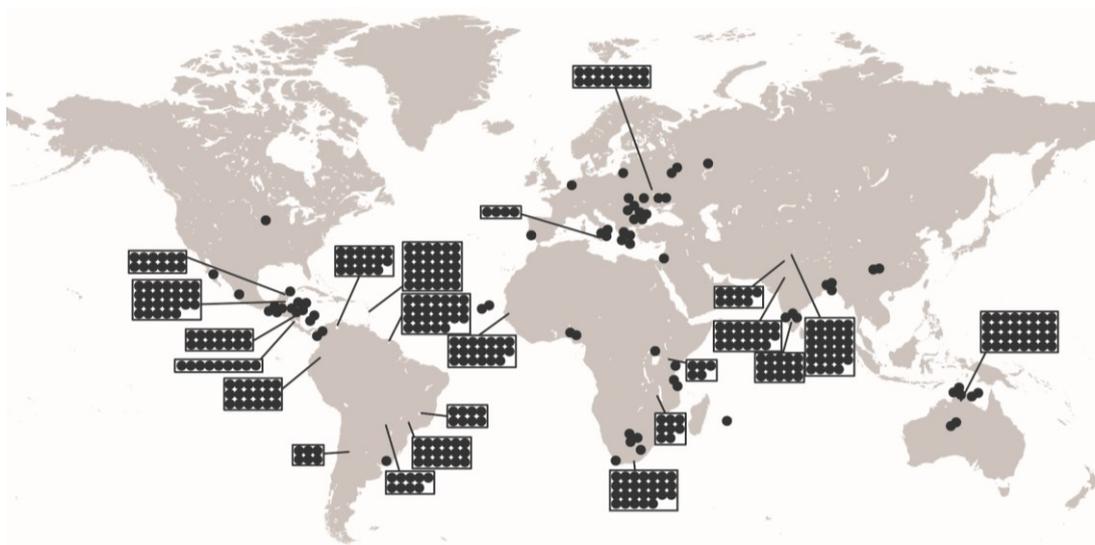


Figure 4-11 Geographical locations of CTVT samples analysed in this Chapter (see Additional file 4-4). Each dot represents the location where one CTVT tumour was collected from.

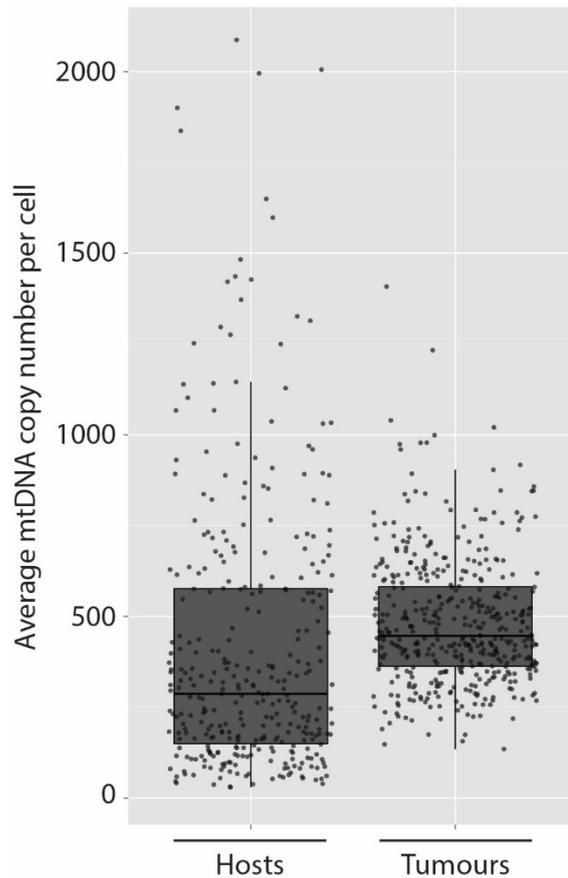


Figure 4-12 mtDNA copy number in CTVT and host samples. MtDNA copy number was estimated by normalising mtDNA sequence coverage to whole genome sequence coverage (Additional file 4-2). Each point represents an individual tumour or host. MtDNA copy number in tumours was not normalised for host contamination. Host and tumour samples with average MT coverage >300X (see Additional file 4-2) were excluded from the analysis and from calculation of average number of mtDNA copies per cell.

4.3.3 Number of mtDNA horizontal transfer events in CTVT

A maximum likelihood phylogenetic tree including mtDNA sequences from CTVT and matched hosts revealed that CTVT mtDNAs cluster in five distinct groups (Figure 4-13) (see Additional file 4-12, Additional file 4-15, Additional file 4-16) (tree was constructed by Máire Ní Leathlobhair, PhD student in the Transmissible Cancer Group). These data suggest that mtDNA in CTVT has at least five independent origins, demarcating five groups that we have named CTVT clades 1 to 5.

Moreover, Figure 4-13 indicates that the donor mtDNA for all clades 1-5 cluster within the previously identified dog mitochondrial clade A (from which follows that the 'donor haplotype' belongs to clade A) (see Figure 4-13 Figure 4-14). 'Donor haplotypes' were

reconstructed for each of clades 1-5 (see Figure 4-14). Each ‘donor haplotype’ represents the inferred donor mtDNA haplotype in each of these 5 horizontal transfer events. ‘Donor haplotypes’ were reconstructed from the clade-defining germline substitutions (i.e. substitutions common to all tumours within a clade and present on the most closely related haplotypes from the normal dog population) and the clade-defining potential somatic substitutions (i.e. substitutions absent from normal dog mtDNA haplotypes but common to all tumours within a clade) (see Materials and Methods, section 4.2.7.2.1). The phylogenetically closest haplotype to each CTVT clade found in the sequenced dog population is also shown in Figure 4-14.

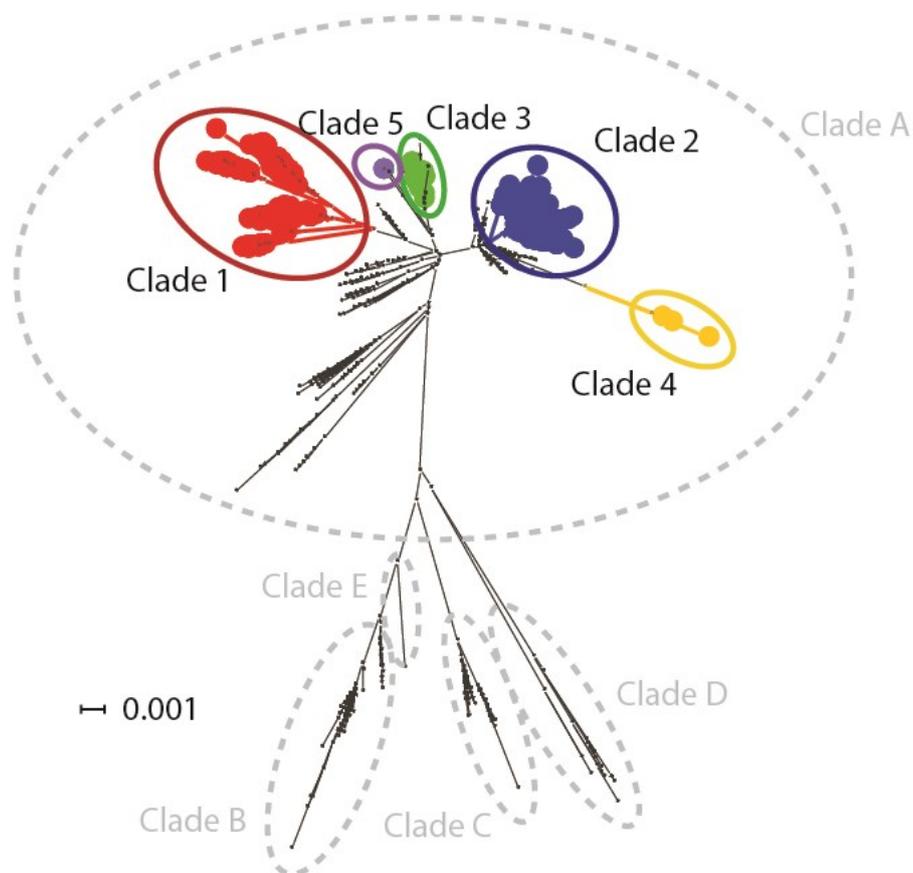


Figure 4-13 CTVT has acquired mtDNA by horizontal transfer at least five times and CTVT mtDNA clades 1-5 all arose from dog mtDNA clade A. Representation of maximum likelihood phylogenetic tree constructed with complete mtDNA sequences from 449 CTVT tumours and 590 dogs. Coloured dots and black lines represent CTVT and dog mtDNA respectively. CTVT mtDNA clades 1-5 are labelled by coloured circles. Dog mtDNA clades A to E are labelled by grey circles (Vila et al. 1997, Savolainen et al. 2002). Scale bar indicates base substitutions per site. Figure is adapted from (Strakova et al. 2016).

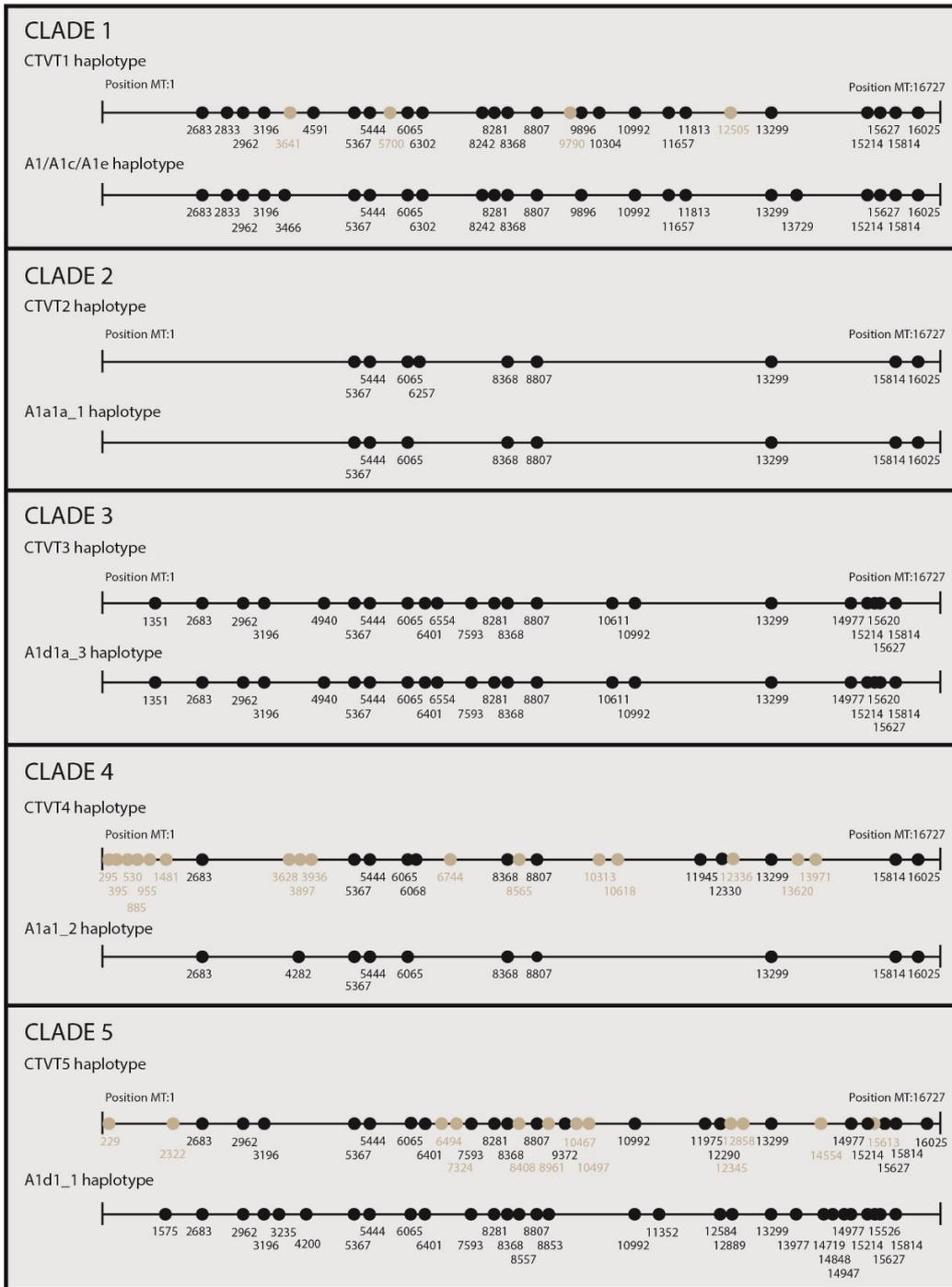
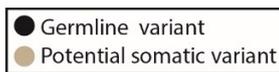


Figure 4-14 Reconstructed donor haplotypes for CTVT mtDNA clades 1 to 5. MtDNA is represented linearly and genome coordinates are labelled. Diagrams representing the likely donor haplotype for each of the CTVT mtDNA clades 1 to 5. The coordinates for each substitution variant position are shown, and substitutions are colour-coded either as “germline” (i.e. they are present in all tumours within a clade and are found in the most closely related dog mtDNA haplotype, which is represented below each of the clade diagrams) or they are found in the most closely related dog mtDNA haplotype only); or “potential somatic” (i.e. they are present in all tumours within a clade but are not found in the most closely related dog mtDNA haplotype).

4.3.4 Relative time since mtDNA horizontal transfer events

Since the horizontal transfer events took place independently, it would be expected that each of them took place at different time points during the history of the lineage. CTVT clades 1-5 were analysed for presence of somatic mutations arising after the mtDNA horizontal transfer events (see Figure 4-15). Variants were called, filtered and classed as somatic as described in Materials and Methods, section 4.2. This analysis revealed that clade 1 mtDNA carry more than double the number of mtDNA somatic mutations (22.5 mutations average) compared with clade 2 mtDNA (9.4 mutations average), which was almost equivalent in number to clade 5 (12.0 mutations average). Clade 3 carried the lowest number of mutations (5.0 mutations average), while clade 4 was identified to have the highest number of mtDNA somatic mutations (34.7 mutations average) (see Figure 4-15).

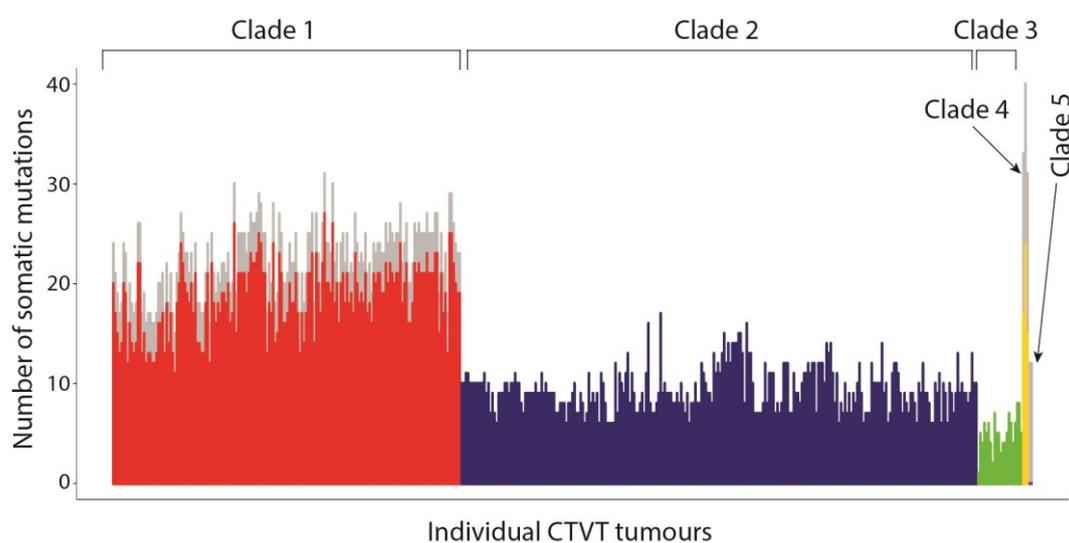


Figure 4-15 Number of somatic substitution mutations per CTVT tumour. Coloured bars indicate somatic mutations acquired by each tumour since mtDNA capture. Grey bars indicate substitutions absent from normal dog mtDNA haplotypes but common to all tumours within a clade; therefore, the early somatic or rare germline status of these variants is not known. Figure is adapted from (Strakova et al. 2016).

Previous work has shown that although CTVT originated about 11,000 years ago, whole genome sequences of two CTVT tumours derived from clades 1 and 2 indicate that tumours belonging to these two clades share a common ancestor approximately 460 years ago (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014). By inferring that the clade 2 mtDNA horizontal transfer event occurred no more than 460 years ago, our subsequent analysis suggests a maximum time since mtDNA uptake of 1097 years for clade 1, 244 years for clade 3, 1690 years for clade 4 and 585 years for clade 5 (all calculations are assuming a

constant mtDNA somatic mutation rate in CTVT) (see Materials and Methods, section 4.2.7.3.1).

Importantly, we used two additional mutation rate estimates, derived using human data (Ju et al. 2014) to estimate the timing of mtDNA horizontal transfer events. Both estimates suggested similar timing for CTVT clade origins (see Table 4-11). The second timing estimate was based on number of cell divisions and third timing estimate was based on number of mutations per year (see Materials and Methods, section 4.2.7.3).

Clade	Average number of somatic mutations	Timing based on nuclear DNA	Timing based on number of cell divisions (first stage tumours)	Timing based on number of cell divisions (second stage tumours)	Timing based on number of mutations per year
		Years before present	Years before present	Years before present	Years before present
1	22.5	1097.1	246.5	1232.4	900.2
2	9.4	460.0	103.3	516.7	377.5
3	5.0	243.7	54.8	273.8	200.0
4	34.7	1689.9	379.6	1898.2	1386.7
5	12.0	585.0	131.4	657.1	480.0

Table 4-11 Timing analysis. Methods used to estimate the time since the origin of CTVT clades 1 to 5. The table lists time estimates (years before present for the origin of each clade) assuming potential somatic substitutions are somatic and arose after mtDNA horizontal transfer (see Materials and Methods, section 4.2.7.3).

Potential somatic substitutions (see Materials and Methods, section 4.2.7.2.1) are those which are common to all tumours within a clade, but that are not found in the normal dog population; therefore we cannot confirm their germline or early somatic status. In Table 4-11, all estimates are calculated using assuming potential somatic substitutions are somatic. The same analysis was repeated assuming potential somatic substitutions are germline (i.e. present on the mtDNA haplotype that founded each clade), thus suggesting minimum times since mtDNA uptake, as shown in Table 4-12.

Clade	Average number of somatic mutations	Timing based on nuclear DNA	Timing based on number of cell divisions (first stage tumours)	Timing based on number of cell divisions (second stage tumours)	Timing based on number of mutations per year
		Years before present	Years before present	Years before present	Years before present
1	18.7	911.3	204.7	1023.6	747.8
2	9.4	460.0	103.3	516.7	377.5
3	5.0	243.7	54.8	273.8	200.0
4	18.7	909.9	204.4	1022.1	746.7
5	0.0	0.0	0.0	0.0	0.0

Table 4-12 Timing analysis. Methods used to estimate the time since the origin of CTVT clades 1 to 5. The table lists time estimates (years before present for the origin of each clade) assuming potential somatic substitutions are germline and were originally present on the mtDNA haplotype that founded the clade (see Materials and Methods, section 4.2.7.3).

Thus, this analysis suggests that the original mtDNA, that was present in the founder dog that first spawned CTVT, is not detectable in CTVT tumours analysed in this Chapter, and provides further evidence showing that CTVT cells have captured mtDNA from transient hosts a number of times within the last two thousand years.

4.3.5 Phylogeography of CTVT

The geographical distribution of all five CTVT clades presented in section 4.3.3 reveals the dynamic worldwide spread of the disease around the world (see Figure 4-16). The distribution of the individual clades indicates that CTVT has spread around the world in at least three individual sweeps (see Figure 4-16), shown by clades 1-3 being present in many countries around the world. Clades 4 and 5 on the other hand, were each detected in only a single country. Compared to a clear geographical pattern shown by the distribution of CTVT clades, the corresponding host dog mitochondrial clades have no detectable geographical pattern, as previously described (Savolainen et al. 2002) (Additional file 4-13).

The phylogenies of CTVT clades 1-5 were constructed (see Materials and Methods, section 4.2.7) (see Figure 4-17). Each tree in Figure 4-17 represents the evolution of a single CTVT clade since it was founded by a horizontal transfer event, with the timeline indicating the maximum number of years since the horizontal transfer event using timing system presented in section 4.3.4 (Table 4-11 and Table 4-12). The two most common clades were clades 1 and

2. Tumours that diverged early in the clade 1 lineage occur in Russia, Ukraine, China and India (see Figure 4-17). Further, clade 1 tumours in Central and South America share a single common ancestor that existed approximately 511 years ago (see Figure 4-17; Materials and Methods, section 4.2.7.3.1). Similarly, our data suggest a single introduction of CTVT to Australia approximately 116 years ago (see Figure 4-17; Materials and Methods, section 4.2.7.3.1). In contrast, CTVT tumours in clade 2 appear to have rapidly disseminated to many parts of the world over a short period of time (see Figure 4-17; Materials and Methods, section 4.2.7.3.1). The more recent clade 3 lineage was found in Central and South America and India, and the less frequent clades 4 and 5 occurred only in India and Nigeria respectively see Figure 4-16 and Figure 4-17.

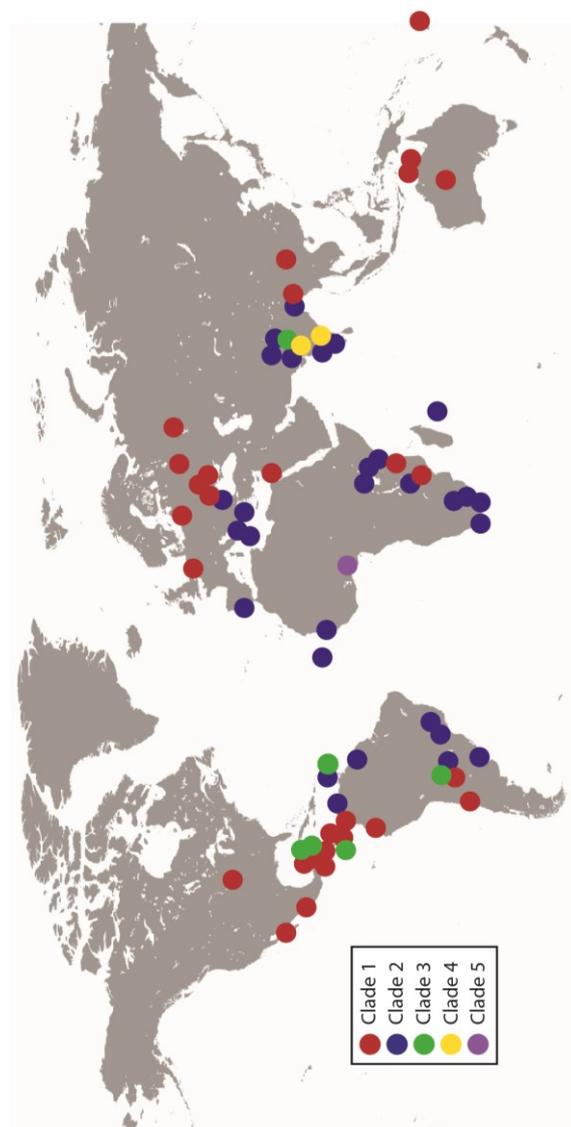


Figure 4-16 Geographical distribution of CTVT clades. Coloured dots represent locations from which one or more CTVT tumours were collected.

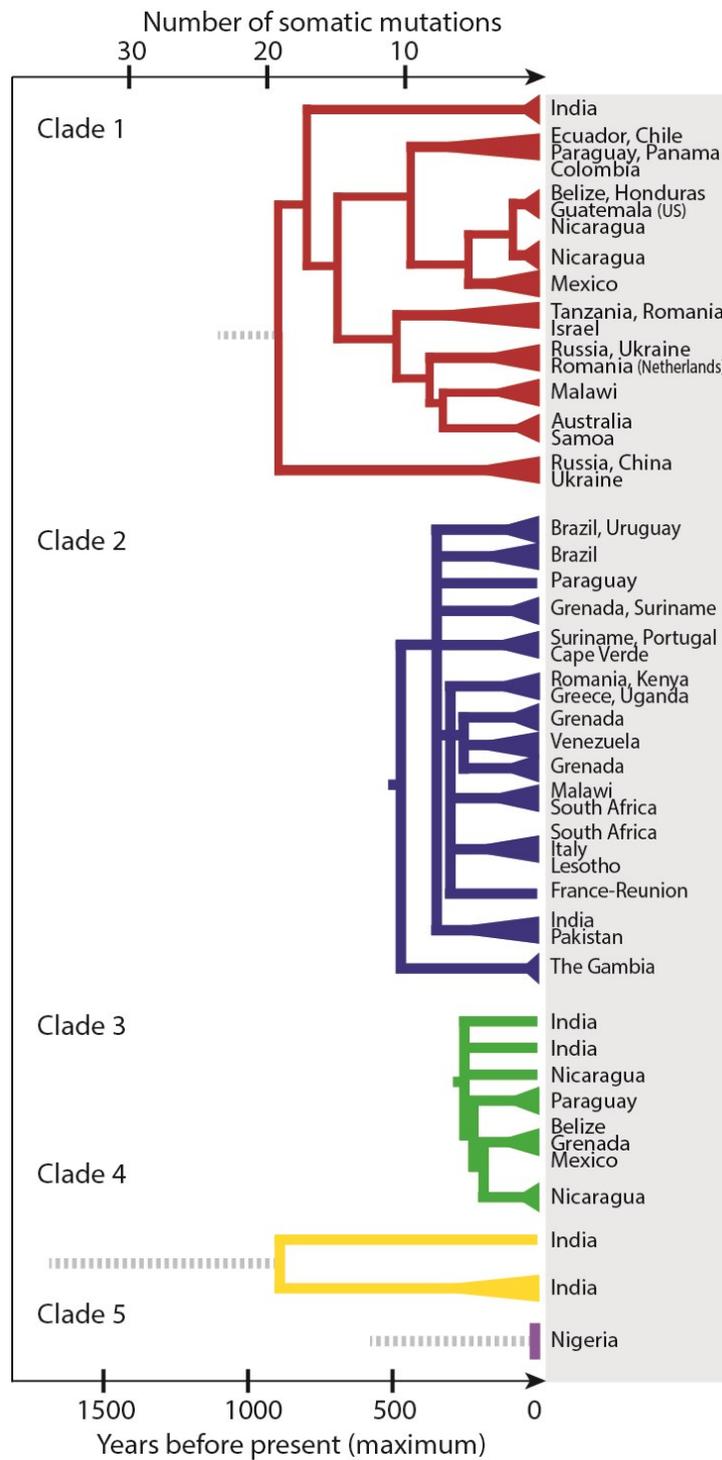


Figure 4-17 Simplified representation of maximum likelihood phylogenetic trees for each clade. Trees illustrate nodes with bootstrap support >60, and shaded triangles represent coalescence of individual branches within each group. Two tumours were collected in the United States and the Netherlands respectively from dogs imported from Guatemala and Romania. Discontinuous grey lines represent contributions of substitutions absent from normal dog mtDNA haplotypes but common to all tumours within a clade. Assuming a constant accumulation of mutations within and between clades, approximate number of somatic mutations and estimated timing is shown. Maximum likelihood trees upon which these representations are based are found in Additional file 4-9. Figure is adapted from (Strakova et al. 2016).

4.3.6 Negative selection in CTVT mtDNA

Most somatic mutations in cancer are believed to be selectively neutral, and there is little evidence in human cancers for negative selection operating to protect essential cellular processes (Stratton et al. 2009). We therefore searched for evidence of mtDNA functionality and importance in CTVT cells by examining CTVT mtDNA for signals of negative selection, which would prevent accumulation of deleterious mutations in mtDNA.

If present, negative selection acting on mtDNA would be expected to prevent homoplasmy of deleterious mutations. Consistent with this prediction, the variant allele fraction (VAF) of nonsense substitutions as well as frameshift indels were significantly lower than VAF for other substitutions and indels (as indicated by difference between red and black cumulative distribution functions in Figure 4-18A and B, $p=0.00019$ and $p=3.03\times 10^{-5}$ respectively, two-sample Kolmogorov-Smirnov test). Interestingly, the only identified homoplasmic nonsense mutation was found at the end of the cytochrome b gene, and therefore effectively had most likely very little effect on the gene function.

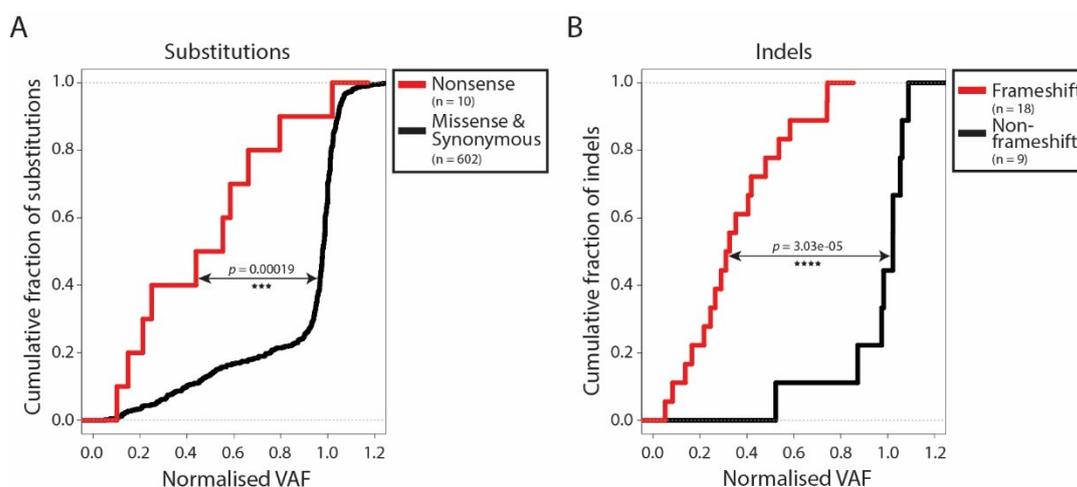


Figure 4-18 Negative selection operates to prevent the accumulation of gene-disrupting mutations in CTVT. Cumulative distribution functions for variant allele fraction (VAF) for tumour (A) substitutions and (B) indels. Figure is adapted from (Strakova et al. 2016).

Furthermore, dN/dS for somatic mtDNA mutations in CTVT showed significant deviation from neutrality both for nonsense (0.187 , $p=1.02\times 10^{-7}$) and missense (0.748 , $p=4.18\times 10^{-3}$) mutations (Figure 4-19) (dN/dS calculations were performed by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group), as described in Materials and Methods, section 4.2.10.2).

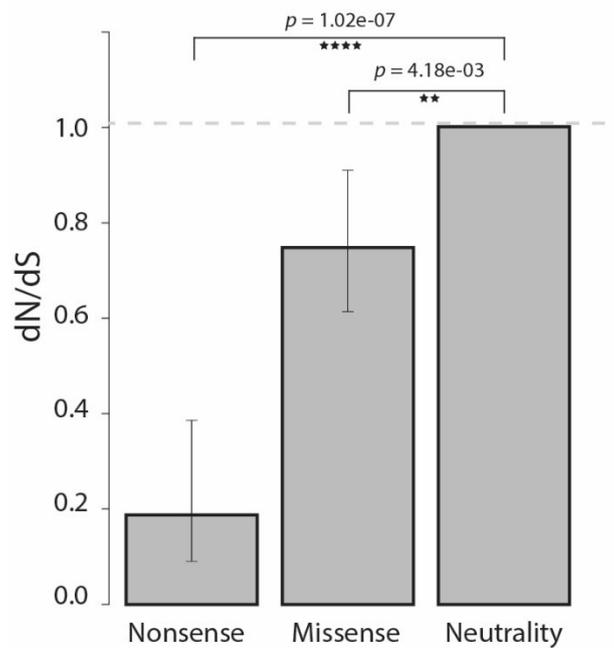


Figure 4-19 Negative selection operates to prevent the accumulation of gene-disrupting mutations in CTVT. dN/dS for somatic nonsense and missense substitutions is shown. P-values were calculated using a likelihood ratio test with parameters estimated using a Poisson model. Calculations were performed by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group). Error bars indicate 95 percent confidence intervals. Figure is adapted from (Strakova et al. 2016).

4.3.7 Ancient mtDNA recombination in CTVT

As discussed in section 4.1.3, mtDNA is usually assumed to be recombinationally inactive and mtDNA recombination has not yet, to our knowledge, been detected in cancer. One of the reasons why mtDNA recombination has not been frequently observed may be the fact that mtDNA is clonally inherited. Thus, in normal somatic cells, only one haplotype is observed, and there are no distinct genetic markers that can be used to test for recombination. Given the possibility for presence of two distinct mtDNA haplotypes in one CTVT cell (a situation that would arise immediately after each mtDNA horizontal transfer event), we searched for evidence of recombination in this cancer lineage. Recombination detection algorithms 3seq and SiScan (Boni et al., 2007; Gibbs et al., 2000) were utilised by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group) (see Materials and Methods, section 4.2.9.1).

We observed significant evidence for mtDNA recombination in CTVT clade 1, detecting recombination breakpoints at around MT:5430 and MT:16176. Maximum likelihood phylogenetic trees constructed using segments MT:1–5429 and MT:5430–16176 derived from clade 1 mtDNA differed in their topology (see Figure 4-20A). Further inspection of clade

1 mtDNA haplotypes revealed that recombination replaced a segment of mtDNA in region MT:1–5429 in clade 1 CTVT tumours that diverged from Central American clade 1 CTVT haplotype (see Figure 4-20B). The recombined mtDNA haplotype has subsequently colonised areas of South and Central America (Chile, Colombia, Ecuador, Panama, Paraguay). In summary, this analysis provides evidence for mtDNA recombination in an ancestral CTVT lineage.

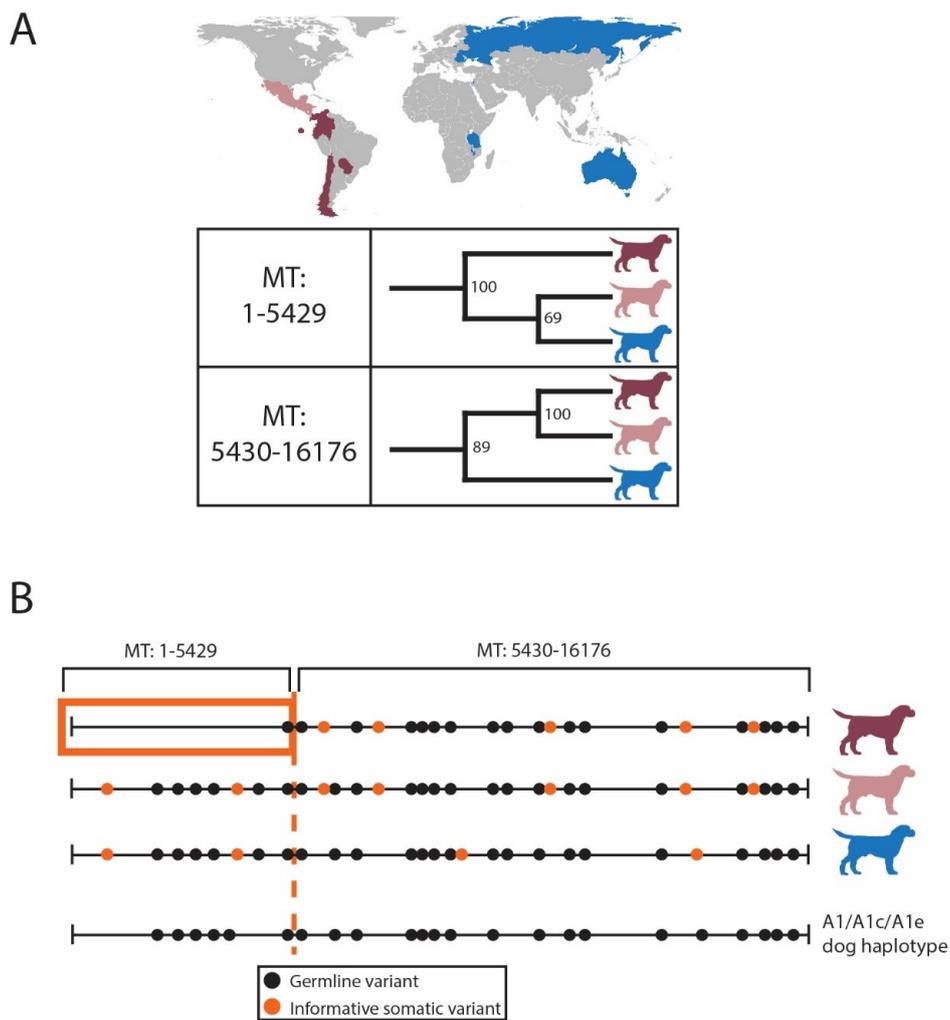


Figure 4-20 Ancient mtDNA recombination in CTVT. (A) Maximum likelihood phylogenetic trees constructed using segments MT:1–5429 and MT:5430–16176 from clade 1 CTVT mtDNAs by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group). Three clade 1 mtDNA haplotype groups are represented by coloured dog silhouettes, and their geographical distributions are colour-coded on the map. Bootstrap values were calculated from 100 iterations. Maximum likelihood trees upon which these representations are based can be found in Additional file 4-10 (B) Simplified haplotype diagrams for clade 1 CTVT mtDNAs derived from groups shown in (A). Germline variants that were present in the donor mtDNA that founded clade 1 are represented by the A1/A1c/A1e dog haplotype (see Figure 4-14). Region putatively replaced by recombination is outlined with orange box. Figure is adapted from (Strakova et al. 2016).

4.3.8 Recent mtDNA recombination in CTVT

After detecting ancient mtDNA recombination, we considered the possibility that a more recent mtDNA recombination event has taken place. Through searching for outliers on CTVT mtDNA phylogenetic trees, thus highlighting samples with unusual numbers and combinations of mutations, we searched for evidence of a more recent occurrence of this phenomenon (see Materials and Methods, section 4.2.9). This analysis identified 559T, a CTVT tumour derived from a male dog from Nicaragua (see Figure 4-21 and Additional file 4-12). In the variant allele frequency (VAF) plot for this sample, rather than seeing the usual three populations of variants separated by their VAF values (see Materials and Methods, section 4.2.5.3), we observed what appeared to be four clusters of variants, suggesting that there may be more than two haplotypes present in this sample (see Figure 4-22).



Figure 4-21 Section of the phylogenetic tree showing sample 559T with its abnormal branch length. Note that only samples from Nicaragua, Belize, Guatemala and Honduras are shown in this tree for purposes of clarity. Tree was constructed by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group).

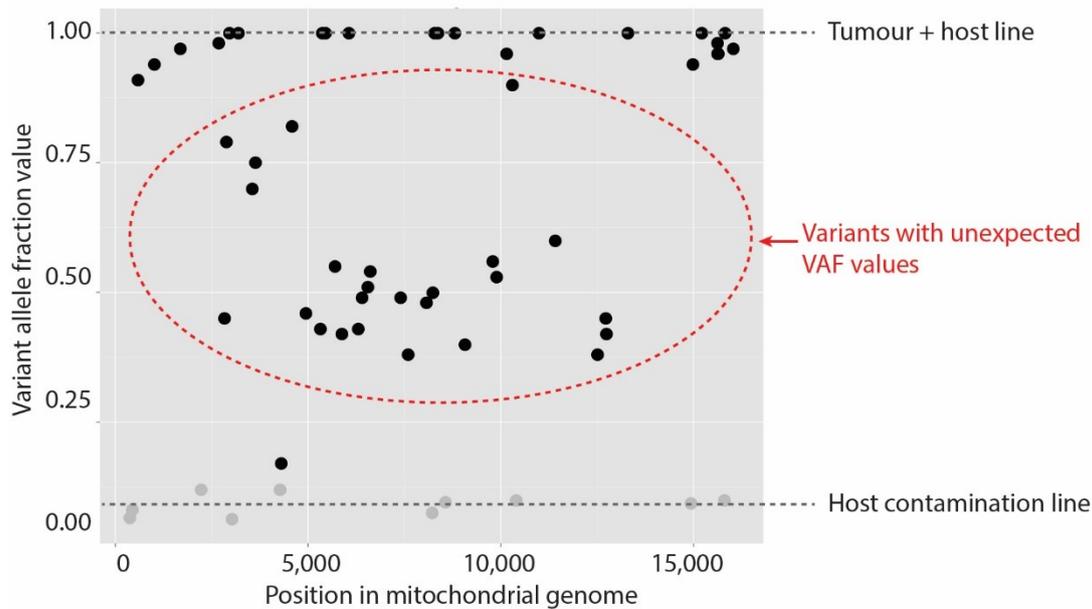


Figure 4-22 Variant allele fraction plot representing sample 559T, identified as a phylogenetic outlier on the mtDNA tree (see Figure 4-21), suspected to be undergoing mtDNA recombination. Note the presence of variants with unexpected VAF values.

Cloning experiments were performed to check for presence of clade 1 mtDNA haplotype-specific substitution (4591 G>A) and A1d1a_1 host mtDNA haplotype specific substitution (4940 T>C) on the same mtDNA molecule, which would be suggestive of occurrence of mtDNA recombination (however at this stage, we cannot exclude the possibility that this may appear on a single mtDNA molecule as a somatic mutation). Indeed, we detected both variants to be present on the same mtDNA haplotype in 3 out of 15 clones.

Further investigation of mtDNA in 559T using PacBio long-read sequencing revealed what appeared to be a CTVT clade 1 mtDNA haplotype (CTVT_1B2b1_29) superimposed upon a normal dog mtDNA haplotype (A1d1a_1), neither of which resembled the mtDNA haplotype found in normal tissues from this host dog, 559H (B1_1 haplotype). Initial analysis of cloned PCR products confirmed presence of A1d1a_1 specific variant (MT: 4940) and CTVT_1B2b1_29 specific variant (MT: 4591) on the same mtDNA molecule. Subsequent phasing of mtDNA variants in 559T using long read sequencing indicated the presence of at least three distinct mtDNA haplotypes in this tumour, each representing a recombination product apparently derived from mtDNA haplotypes CTVT_1B2b1_29 and A1d1a_1 (see Figure 4-23). These data suggest that a tumour ancestor to 559T captured haplotype A1d1a_1 mtDNA from its host. Recombination was initiated between mtDNA haplotypes CTVT_1B2b1_29 and A1d1a_1, and cells containing these recombination products were passed on to host 559H. Alternatively, it is possible that dog 559H received a mixture of both

normal and CTVT cells from its CTVT tumour donor individual, and mtDNA capture and recombination occurred only within 559H dog. It must also be taken into account that the A1d1a_1 haplotype resembles the CTVT clade 3 donor haplotype (Figure 4-14); thus we cannot exclude the possibility that the recombination event that we observe in 559T involved horizontal transfer between clade 1 and clade 3 CTVT tumours that occurred within the same animal.

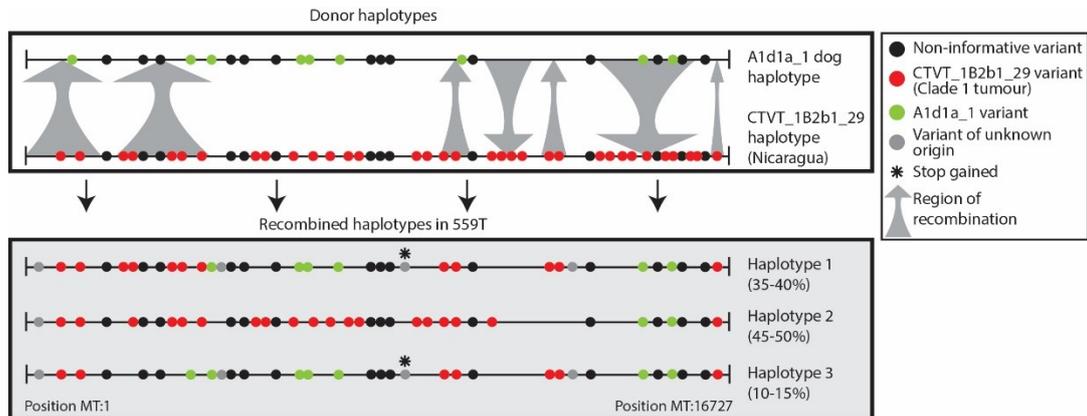


Figure 4-23 Recombination detected in tumour 559T (Nicaragua). The estimated percent contribution of each recombined haplotype to the mtDNA population within 559T CTVT cells is shown, and grey arrows indicate likely sites of recombination. Host contamination is ~5%.

4.3.9 Histopathology analysis of CTVT clades

Capture of distinct mtDNA haplotypes and subsequent genetic divergence of CTVT clades may affect the biological behaviour of various CTVT lineages. To understand phenotypic variation between the two most common clades (clade 1 and clade 2), a histopathology screen was performed. The following histopathology parameters were chosen to be scored, as described in Materials and Methods (section 4.2.12): normal mitotic index, bizarre mitotic index, total mitotic index, bizarre mitotic index as a proportion of total mitotic index, lymphocyte density in parenchyma, lymphocyte density at interface, haemorrhage and necrosis.

4.3.9.1 Mitotic index and lymphocyte density

Mitotic index and lymphocyte density analysis was performed using a Generalised Linear Mixed Effect (GLMER) model implemented in R, taking into account the following confounding factors: host age, host gender and tumour ulceration (see Materials and Methods, section 4.2.12.3)

Differences between clade 1 and clade 2 are displayed in Table 4-13, Table 4-14 and Figure 4-24. Although both normal and total mitotic index were higher in clade 2 in comparison to clade 1 (Table 4-14 and Figure 4-24), these results are insignificant after application of Bonferroni correction for multiple hypothesis testing.

Parameter	GLMER output estimate	Approximate factor by which clade 2 differs from clade 1
Total Mitotic Index	0.218	1.246
Normal Mitotic Index	0.221	1.247
Bizarre Mitotic Index	0.175	1.191
Proportion of Bizarre Mitoses	0.129	1.137
Parenchymal Lymphocyte Density	0.013	-1.013

Table 4-13 GLMER analysis implemented in R (R Core Team 2013). Table showing GLMER output estimate and approximate factor by which clade 2 differs from clade 1.

Parameter	Clade 1 Median (IQR)	Clade 2 Median (IQR)	p-value
Total Mitotic Index	0.9 (0.525-1.175)	1.1 (0.6-1.4)	0.016 (*)
Normal Mitotic Index	0.6 (0.4-0.9)	0.8 (0.5-1.0)	0.011 (*)
Bizarre Mitotic Index	0.2 (0.025-0.4)	0.2 (0-0.5)	0.302
Proportion of Bizarre Mitoses	0.25 (0.089-0.4)	0.25 (0.015-0.397)	0.375
Parenchymal Lymphocyte Density	3.6 (0.8-9.0)	4.2 (2.0-8.0)	0.943

Table 4-14 Median and interquartile range (IQR) for all histopathology parameters. The p-value was generated by GLMER model implemented in R. Significance before Bonferroni correction is indicated by (*). Application of Bonferroni correction for multiple hypothesis testing reduces the p-value required to attain significance from <0.05 to <0.01, thus rendering results insignificant.

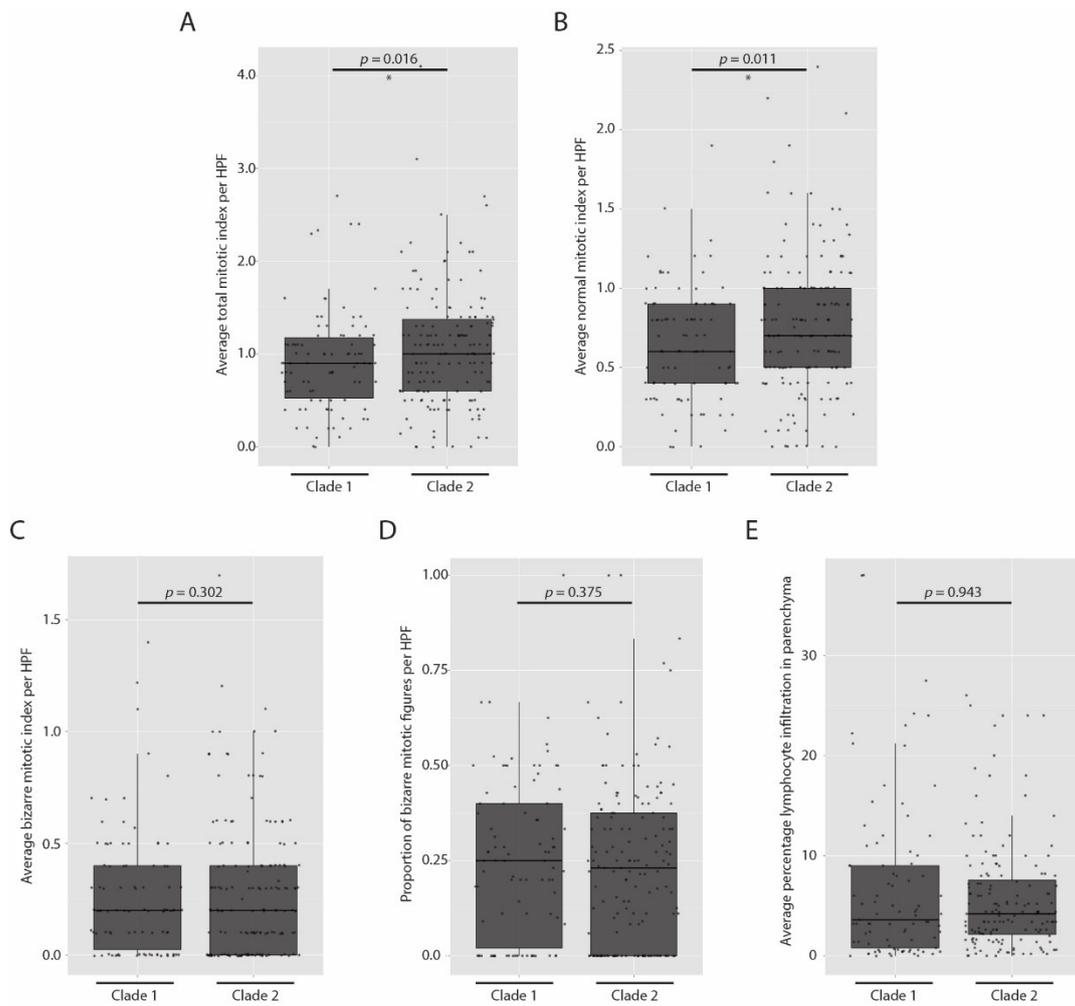


Figure 4-24 Comparison of histopathological parameters between clade 1 and clade 2 (A) total mitotic index (B) normal mitotic index (C) bizarre mitotic index (D) bizarre mitoses as a proportion of total mitoses (E) lymphocyte infiltration in parenchyma. Statistical significance before Bonferroni correction is indicated by *. HPF = high power field (40X). Each individual sample is represented by a dot. Boxes represent the first and third quartiles (inter-quartile range = IQR). Error bars indicate values within 1.5*IQR away from first and third quartiles. P-values were generated by GLMER model (see Table 4-14).

The standard deviation values for each parameter are shown in Table 4-15 and indicate a greater diversity of total/normal/bizarre mitotic figure counts in clade 2.

Parameter	Standard deviation Clade 1	Standard deviation Clade 2
Total Mitotic Index	0.542	1.141
Normal Mitotic Index	0.355	0.750
Bizarre Mitotic Index	0.283	0.489
Proportion of Bizarre Mitoses	0.207	0.214
Parenchymal Lymphocyte Density	8.187	7.031

Table 4-15 Standard deviation for all histopathology parameters.

4.3.9.2 Haemorrhage and necrosis

Frequency of occurrence of haemorrhage and necrosis were not found to significantly differ between clade 1 and clade 2. Percentage of samples with recorded haemorrhage and necrosis is shown in Table 4-16 (two-tailed Fisher Exact Test showing no difference between clades was implemented in R (R Core Team 2013)).

	Haemorrhage		Necrosis	
	Present	Absent	Present	Absent
Clade 1	47 (52.8%)	42 (47.2%)	3 (3.4%)	86 (96.6%)
Clade 2	64 (42.1%)	88 (57.9%)	5 (3.3%)	146 (96.7%)

Table 4-16 Presence of haemorrhage and necrosis in clade 1 and clade 2.

4.4 Discussion

The analysis presented in this Chapter provides evidence to show that CTVT has captured its mtDNA through mtDNA horizontal transfer events at least five times during the history of this cancer lineage, thus delineating five different CTVT clades named clades 1-5. The worldwide distribution and phylogenetic relationships between CTVT samples enabled us to trace and time the historical spread of this disease during the last 2,000 years. Negative selection has been detected to operate in this long-lived cancer, acting to safeguard the essential cellular processes and indicating the importance of mitochondrial function in CTVT. Our work reveals, for the first time, evidence of mtDNA recombination in cancer - we highlighted occasional re-assortment of mtDNA appearing both in ancient CTVT lineages as well as in current CTVT tumours. A histopathology-based screen did not identify any significant phenotypic differences between the mitochondrial clades.

4.4.1 Historical spread of CTVT

Previous work had indicated that the CTVT lineage may have been confined to a single place for most of its history, until it diverged and spread around the world in the last few hundred years (Rebeck et al. 2009, Murchison et al. 2014). The analysis presented in this Chapter allowed us to trace the historical spread of CTVT in the last ~2,000 years and indicated that the disease rapidly spread around the world within a few hundred years. Previously published work likely analysed samples only in CTVT clades 1 and 2 that diverged ~500 years ago (Rebeck et al. 2011). Our analysis reported tumours that belong to 3 additional clades (as well as reporting samples in clades 1 and 2), some of which diverged from each other much longer ago. The oldest clade reported in our analysis, clade 4, arose around ~2,000 years ago. The two most common clades with a worldwide distribution are clades 1 and 2, lending us the opportunity to speculate about the routes of their historical spread (see Figure 4-25).

We detected clade 1 CTVT tumours in Europe, Asia, Australia, Africa, South America and North America. This lineage probably spread approximately 1,000 years ago, most likely from Russia or China, with samples from this region clustering on the ancestral branch of the maximum likelihood phylogenetic tree (see Figure 4-17, Figure 4-25). The second most divergent branch was found in northeast India. Further, our results show that there was probably a single introduction of clade 1 to South America around 500 years ago, most likely with early Colonial contact, after which the lineage has spread down the west coast of South America to Chile and Paraguay. Similarly, our data suggest a single introduction to Australia

after European arrival, just over 100 years ago, with tumours from the mainland of Australia found to be very closely related to CTVT tumours from Samoa. Clade 1 tumours are now found in most parts of the world (Figure 4-16).

Clade 2 captured its mtDNA more recently, approximately 500 years ago, and spread very rapidly around the world as an explosive burst of expansions from a single place. As Europe was the maritime hub at the time, it is possible that this was the place from which clade 2 expanded. The most ancestral branch of the maximum likelihood phylogenetic tree was found in West Africa (The Gambia) (see Figure 4-17), but around the same time clade 2 has spread to other parts of Africa, Brazil, the Caribbean, southern Europe and India. The clade 2 tumours were even introduced to island canine populations such as Reunion or Cape Verde around this time. Based on its worldwide distribution including island populations and timing of its divergence, this clade may have been transported between continents by maritime routes (see Figure 4-25). We can therefore speculate that dogs with CTVT infection would have been travelling on boats as human companions.

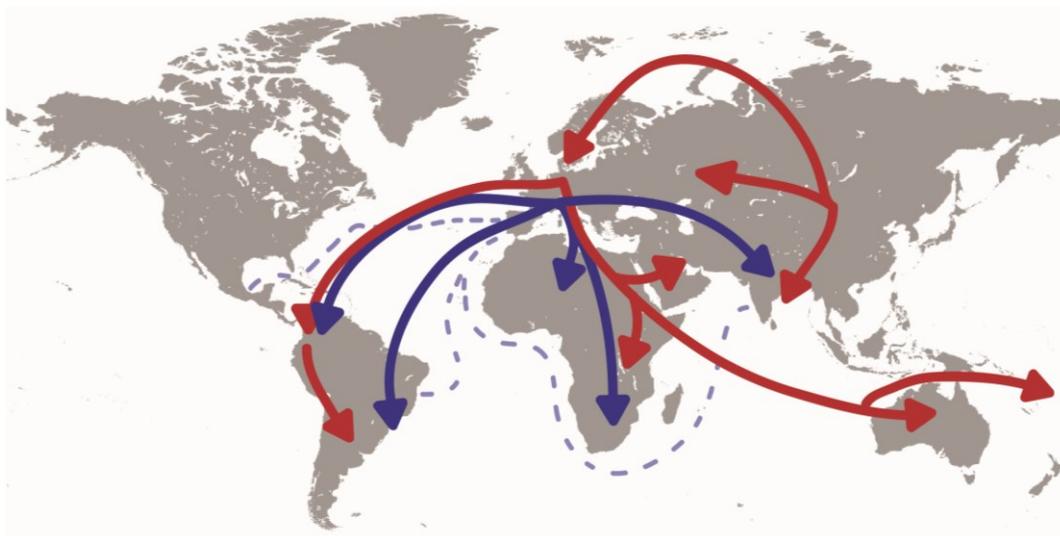


Figure 4-25 A model for how CTVT might have spread around the world. The two most common clades are indicated - red arrows represent clade 1, blue arrows represent clade 2. Dashed line indicates how clade 2 may have been transported by maritime routes.

The clade 3 lineage has spread around the world most recently, only in the last hundred or few hundred years. It was found to be present in South and Central America and India. The distribution patterns of this clade mean that it is less clear how it could have spread to such distant and seemingly unrelated parts of the world (Figure 4-16).

The remaining two clades, clade 4 and clade 5, were only found in two locations – India and Nigeria respectively. Clade 4 was the oldest clade we identified, so it is possible that it was widely present in central Asia in the past and has now been replaced by the most recent clades 1 and 2. Interestingly, the two phylogenetically most divergent groups in clade 4 that diverged around 900 years ago were both found in Jaipur (see Figure 4-17), suggesting that clade 4 may have been present in this area since that time.

The timings of CTVT spread and its worldwide distribution lend themselves to speculate how the disease was transported around the world. It seems very plausible that dogs would have been travelling alongside humans and therefore through understanding how CTVT moved around the world, we can trace the historical movements of dogs. Moreover, the extensive and recent global expansion detected in the CTVT lineage is consistent with signals of widespread admixture observed in worldwide populations of domestic dogs (Shannon et al. 2015), highlighting the extent to which canine companions accompanied human travellers in recent global explorations.

4.4.2 Mechanisms of mtDNA horizontal transfer

The exact mechanism of mtDNA horizontal transfer between cells remains unresolved (Berridge et al. 2015, Tan et al. 2015), although exchange of mtDNA has been observed between human cells *in vitro* as well as between normal and cancer mouse cells (and other mouse models) *in vivo* (Spees et al. 2006, Berridge et al. 2015, Tan et al. 2015, Berridge et al. 2016, Hayakawa et al. 2016, Dong et al. 2017). A recently published study suggested that intact whole mitochondria are selectively transferred between normal and cancer mouse cells *in vivo* (Dong et al. 2017). It has not, however, been yet established whether mitochondria are transferred to the target cells through structures such as tunnelling nanotubes (TNTs) or through uptake of vesicles (Spees et al. 2006, Berridge et al. 2015, Tan et al. 2015, Berridge et al. 2016, Hayakawa et al. 2016, Dong et al. 2017). Considerable evidence, however, supports the view that TNT-like structures are involved (Berridge et al. 2016, Dong et al. 2017). The mechanism and number of mitochondria that are transferred *in vivo* however remains to be elucidated (Berridge et al. 2016, Dong et al. 2017).

The extent of mtDNA horizontal transfer between mammalian cells in different tissues and contexts is currently unknown. Previous studies have shown that mtDNA horizontal transfer *in vitro* is enhanced by some chemotherapeutic agents (Moschoi et al. 2016). Further studies have detected *in vitro* enhancement of chemo-resistance to doxorubicin through acquisition

of mtDNA from endothelial cells (Pasquier et al. 2013). Whether the frequency of mtDNA transfer would be enhanced by chemotherapeutic agents *in vivo* remains to be answered.

Whether whole mitochondria are transferred during mtDNA horizontal transfer in CTVT and whether they may also be transferred by trafficking or perhaps rather by cell fusion still remains as a question, and our current work presented in this Chapter does not yet help us to understand the mechanism further.

4.4.3 Frequency of mtDNA horizontal transfer

Mitochondrial DNA horizontal transfer is usually considered to be a rare phenomenon, but recent studies suggest that it may be more common than previously thought and highlight the particularly evident movement of mtDNA in cancer, where defective mitochondrial respiration could be restored by acquisition of mtDNA (Spees et al. 2006, Tan et al. 2015, Dong et al. 2017)(reviewed in (Berridge et al. 2016)). Our work including 449 CTVT samples from around the world has expanded the identified number of mtDNA horizontal transfer events in CTVT and implied that CTVT has captured mtDNA at least five times during the evolution of this lineage.

Furthermore, a sixth mtDNA horizontal transfer event was detected in sample 559T, undergoing mtDNA recombination. Analysis of sample 559T revealed that a third mtDNA haplotype (apart from the tumour and matched host mtDNA haplotypes) had been acquired from another host dog prior to mtDNA recombination taking place (see section 4.3.8). This finding raises the possibility that the frequency of mtDNA horizontal transfer in CTVT may be even higher than reported, but a proportion of introduced mtDNA haplotypes do not get fixed and therefore could not be detected by our analysis.

It has previously been suggested that replacement of CTVT mtDNA, which presumably was carrying large numbers of possibly deleterious mutations, provided a positive selective advantage to the lineage (Rebeck et al. 2011). In our analysis, we have observed multiple mtDNA horizontal transfer events, which supports the suggestion that mtDNA replacement is indeed a positively selected process. However, our selection analysis was not designed to detect positive selection at the level of a haplotype, and therefore we cannot exclude the possibility that CTVT acquired host mitochondrial DNA via purely neutral processes, or that the acquired mtDNA haplotype became homoplasmic through a selfish replicative advantage.

4.4.4 MtDNA recombination

MtDNA is usually believed to be recombinationally inert. Our work presented in this Chapter has indicated presence of occasional mtDNA recombination in ancient CTVT lineages as well as mtDNA recombination which occurred very recently. With this result in mind, it would be appropriate to consider whether mtDNA recombination might also be occurring in other cancers, including human cancers. In other cancers, the fact that most mtDNAs present within a single cell are clonally inherited would render detection of mtDNA recombination almost impossible using phylogenetic methods. Other approaches may help us to understand whether mtDNA recombination may be happening on a wider scale.

For mtDNA recombination to take place, two mtDNA molecules must come into direct contact which may suggest that a fusion of mitochondria has taken place. This phenomenon may be related to the fact that mitochondria are dynamic organelles, which constantly undergo fusion and fission (Westermann 2010). Further work to understand the mechanisms of mtDNA recombination will be of a future interest.

The nature of the signals that may trigger onset of mtDNA recombination are not clear. Further analysis may determine if DNA damage signalling is involved, as previously suggested (Thyagarajan et al. 1996) - in this context, it is interesting to observe that a truncating nonsense mutation in COX3 was found in some of 559T haplotypes (Figure 4-23).

Although there was no evidence of mtDNA recombination in CTVT beyond what is described in section 4.3.7 and section 4.3.8, we cannot exclude the possibility that recombination is more widespread in CTVT mtDNA than detected. It is possible, therefore, that the phylogenetic, mutation rate and selection analyses presented in this Chapter (sections 4.3.4, 4.3.5 and 4.3.6) have been influenced by an undetected recombination signal. However, the following evidence suggests, that, if such a signal is present, it is at a low level: the presence in all (non-recombining) CTVT mtDNAs of a set of clade-specific markers (Figure 4-14), the absence (beyond 559T) of distinctive phylogenetic outliers (Additional file 4-9), the very low frequency of back-mutation (Additional file 4-14) and the failure of recombination-detection algorithms to detect further recombination.

4.4.5 Negative selection in CTVT

Cancer development is a result of genetic variation and clonal selection. The somatic variants may either be positively selected and act as the 'driver' mutations in the formation of a

tumour (Ostrow et al. 2014) or they may just be propagated to the next generation alongside positively selected mutations as ‘passenger’ mutations with no functional effects. Published evidence shows the potential for functionally relevant somatic mutations in mtDNA to be positively selected in certain types of cancer (Larman et al. 2012, Schon et al. 2012, Wallace 2012, Yuan et al. 2017). We found no evidence for positive selection at the level of the somatic mutations using dN/dS analysis (see Materials and Methods, section 4.2.10.2). However, we cannot exclude the possibility that a positive advantage at the level of the mtDNA haplotype may exist (which would not be detected by our analysis), through providing either a functional or selfish replicative advantage.

So far, there was little evidence in human cancers for existence of negative selection, which would remove the deleterious variants from the pool of somatic variation. Together with a number of recently published reports (Ju et al. 2014, Stewart et al. 2015, Yuan et al. 2017), our analysis provided early evidence for presence of negative selection in cancer mtDNA (Figure 4-18 and Figure 4-19). The extent to which these somatic changes will influence evolution of cancer is normally limited by the life span of the host, hence it may be difficult to detect negative selection in human cancers. Future studies and detection of negative selection in human and other cancers may potentially provide important insights into identifying genes essential for cancer growth.

Although negative selection is operating to prevent accumulation of deleterious mutations in CTVT mtDNA, the long-term accumulation of slightly deleterious somatic mutations may eventually mean that a capture of host mtDNA would nevertheless be more advantageous for CTVT cells as an adaptive mechanism to survive for millennia. This hypothesis is supported by our findings showing that the oldest mtDNA capture took place only ~2,000 years ago, even though the CTVT lineage is approximately 11,000 years old (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014) (but we cannot exclude the possibility that we have not sampled CTVT lineages where mtDNA capture happened much earlier in their evolution).

Overall, these results demonstrate that maintenance of functional mtDNA is important for the biology of CTVT. MtDNA appears to be fulfilling some metabolic requirements of CTVT, perhaps in terms of ATP production or pyrimidine biosynthesis. As previous studies have suggested (Spees et al. 2006, Berridge et al. 2015, Tan et al. 2015, Berridge et al. 2016, Hayakawa et al. 2016, Dong et al. 2017), this indicates a critical role of oxidative phosphorylation in tumourigenesis. Significance of functional mtDNA in other human cancers

has also been shown by other studies, which implicate functional roles of mtDNA mutations in cancer development, especially in certain types of cancers (Yuan et al. 2017).

4.4.6 Phenotypic differences between CTVT clades

A histopathology screen of CTVT tumours in clades 1 and 2 did not reveal any statistically significant differences between the clades (see 4.3.9), and therefore the null hypothesis stating that clade 1 and clade 2 have same histopathology features was not rejected. As some of the results were suggestive of differences between groups (in particular in terms of the mitotic index) it would be interesting to repeat this analysis with a larger set of samples, as it is plausible that the replacement of clade 1 mtDNA by the more recent clade 2 mtDNA, which might be more efficient in ATP generation and biosynthesis of macromolecules like pyrimidines, may allow faster cell cycling.

Interestingly, there was a greater variation in mitotic index (mainly attributable to normal mitoses, see Figure 4-24 Table 4-15) within clade 2 tumours than in clade 1 tumours. This is surprising, since clade 1 samples diverged at an earlier time-point than clade 2 samples, it would be assumed that they had more time to diverge into sub-clades and therefore clade 1 would be expected to be more heterogeneous.

There was no statistically significant difference in lymphocyte density in parenchyma between clade 1 and clade 2. As survival, transmission and propagation of CTVT relies upon its interaction with the host's immune system, it is perhaps not surprising that any inter-clade variation would not have any functional effects. Nevertheless, it would be interesting to see if any of the less common clades might be interacting with the immune system in a slightly different manner.

Although there was no difference between clade 1 and clade 2, it is interesting to note that a significant proportion (~50%) of tumours in both clades scored positively for presence of haemorrhage. It is plausible that presence of haemorrhage may be an advantage that would equally aid tumour transmission of both clades, as it may promote transfer to new hosts (Strakova and Murchison 2015).

Necrosis was present in such a low frequency that statistical power was very low. Further studies with a larger number of samples will be useful in determining whether necrosis might differ between clade 1 and clade 2.

4.4.7 Host genetic variation

Although the primary aim of this Chapter was to perform analysis of CTVT genetic and phenotypic diversity around the world, matched host samples were sequenced as a part of this study. The principal purpose of the host samples was to distinguish between genetic variation arising from the tumour and genetic variation being caused by host contamination.

The host mtDNA sequences were therefore not directly used for analysis presented in this Chapter, however, they provide a large amount of information on the sequences of many in particular street/mixed dogs. Further analysis of mtDNA sequences from host samples may provide novel information about genetic adaptations and other differences between dogs in different parts of the world. This information may be especially useful, as the genetic features of the street/mixed dog population have never been so extensively analysed before. This dataset will prove to be a valuable resource for future studies as well as for the canine genetics community in general (see Chapter 6 for further discussion).

4.4.8 Future directions and next steps

The comprehensive analysis of mtDNA in CTVT from around the world that was presented in this Chapter enabled us to understand the importance of mtDNA functionality in this long-lived lineage. Nevertheless, this work could be expanded on in the future, as discussed in the section below.

As presented and discussed throughout this Chapter, we have identified five independent mtDNA horizontal transfers in CTVT. Many of these mtDNA transfer events were only detected in a very small number of samples, as the majority of CTVT samples belong to clade 1 or clade 2. Even though we have analysed 449 CTVT samples in this Chapter, there is a very likely possibility that there may be other CTVT clades that we have not sampled – either by chance or simply because our sampling was limited in some regions of the world. Further work looking at more CTVT samples from different parts of the world may reveal the presence of other CTVT mtDNA clades, which may have arisen even further back in the past. Specifically, we have not detected the mtDNA haplotype present in the original founder dog which gave rise to CTVT – even though it is very unlikely that it would still persist in the population due to its massive mutational burden over time, we cannot fully exclude this possibility.

It would be interesting to repeat to histopathological screen described in this Chapter with a larger number of samples. Other scoring parameters could be included in future screens, such as for example the following: lymphocyte distribution, eosinophil count, cell pleomorphism, apoptotic fraction, tumour silhouette, epitheliotropism and fibrous tissue distribution. Moreover, a number of individual parameters could be combined into an aggressiveness score that could be compared between samples and clades.

To further understand the phenotypic variation between samples, future studies could include screening for appearance characteristics specific to each clade using photographs of tumours. Suggestions for gross variables to be scored, as concluded from my pilot studies, are listed as follows: gender, location of tumour, tumour distribution, tumour growth pattern and ulceration.

Finally, the data presented in this Chapter enabled us to analyse phylogenetic relationships of CTVT samples within a single clade, as using mtDNA data we can only trace genetic variation as far back in history, as the time when each mtDNA transfer took place. This means that phylogenetic tree for each CTVT clade was rooted with the reconstructed donor mtDNA haplotype (see section 4.2.8.2). In Chapter 5, I will expand on the analysis presented in this Chapter by analysing nuclear exome sequencing data from the same CTVT samples, which will reveal the phylogenetic relationships *between* CTVT mtDNA clades.

5

Recurrent and recent haplotype-specific mitochondrial DNA horizontal transfer in canine transmissible venereal tumour

Summary

In our previous analysis, we detected five mtDNA horizontal transfer events giving rise to five CTVT clades in a set of 449 CTVT tumours. In this Chapter, I will explain how I used exome sequencing data to understand the relationships between the five CTVT mitochondrial clades presented in Chapter 4. Remarkably, this analysis detected that one of the clades, despite clustering on a single phylogenetic branch on the mtDNA tree, clustered on several different branches on the exome phylogenetic tree, thus indicating recurrent haplotype-specific mtDNA horizontal transfer. Further analysis revealed that the mitochondrial haplotype involved in these recurrent transfers belonged to canine A1d1 haplogroup. MtDNAs of this haplogroup have been captured by CTVT on 12 independent occasions, all of which occurred relatively recently. Genetic analysis of A1d1 haplogroup identified a number of candidate genetic variants, which may explain why this haplogroup has been recurrently horizontally transferred. In particular, analysis of homoplasmic variants in CTVT samples undergoing mtDNA recombination highlighted an insertion at position 16660, which may be conferring a selective advantage to the A1d1 haplogroup, possibly by influencing mtDNA transcription or replication.

5.1 Introduction

5.1.1 MtDNA transfer in CTVT and other cancers

Mitochondrial DNA (mtDNA) horizontal transfer is a poorly understood phenomenon, which has rarely been directly observed in mammalian cells. Indeed, in our previous work we detected only six mtDNA horizontal transfer events in a set of 449 tumours; out of which five became fixed and gave rise to CTVT clades and one was detected in a sample undergoing mtDNA recombination (Strakova et al. 2016) (see Chapter 4). The process has been suggested to provide an evolutionary advantage to CTVT cells – as CTVT mtDNA acquire an increasing number of mutations with time, their efficiency may decline over time, and an occasional horizontal transfer of “fresh” mtDNA from the host may provide a functional advantage (Rebbeck et al. 2011).

Transfer of mtDNA between cells of an individual is usually difficult to detect, mainly since both mtDNA haplotypes are usually identical, thus genetic markers for detecting horizontal transfer are usually lacking. Apart from being detected in CTVT, however, exchange of mtDNA has been observed between human cells *in vitro* as well as between normal and cancer mouse cells (and other mouse models) *in vivo* (Spees et al. 2006, Berridge et al. 2015, Tan et al. 2015, Berridge et al. 2016, Hayakawa et al. 2016, Dong et al. 2017). The extent of mtDNA horizontal transfer between mammalian cells in different tissues and contexts is currently unknown, but perhaps shuttling of mitochondria between cells may be more common than previously appreciated. In CTVT, we have evidence for six mtDNA horizontal transfer events; in five of these cases, the captured mtDNA has become fixed, while in one case (in CTVT tumour 559T), the mtDNA was probably captured recently and was not fixed. Perhaps many horizontal transfer events in CTVT do not result in replacement of the endogenous CTVT haplotype, and are therefore not observed (see Chapter 4, (Strakova et al. 2016)).

It has been suggested that movement of mtDNA between cells may be a normal physiological process, but we cannot exclude the possibility that this phenomenon may occur as a response to mtDNA damage (Berridge et al. 2016). In cancer, mtDNA horizontal transfer may act as an adaptation to metabolic stress or other changes in its microenvironment. Previous studies have shown that mtDNA horizontal transfer *in vitro* is enhanced by some chemotherapeutic agents (Moschoi et al. 2016). *In vitro* enhancement of chemo-resistance to doxorubicin through acquisition of mtDNA from endothelial cells to cancer cells has also been detected (Pasquier et al. 2013). The “triggers” which might lead to occurrence or even enhancement

of mtDNA horizontal transfer in CTVT have not, however, been previously considered or discussed in the literature.

5.1.2 Positive selection of mtDNA haplotypes

After mtDNA horizontal transfer, the captured mtDNA haplotype must presumably be present at a low level before increasing in frequency to homoplasmy. There are numerous ways in which homoplasmy of a low level haplotype could be achieved.

Firstly, in cancer, mtDNA horizontal transfer may act as a pathophysiological process to rescue DNA damage, thus providing an evolutionary advantage for the captured mtDNA haplotype via selection for cells with, for instance, more efficient mitochondrial respiratory function which may enhance tumorigenic properties (Polyak et al. 1998, Fliss et al. 2000, Habano et al. 2000). In this scenario, mtDNA haplotype competition at the level of the mtDNA genomes would lead to expansion of the fitter haplotypes selected at the level of the cell and thus purification of less functionally efficient mtDNA genomes (Hill et al. 2014, Ma et al. 2014).

Alternatively, captured mtDNA may become homoplasmic through a haplotype selective advantage, resulting from 'selfish' evolution through enhanced replication of the captured haplotype, which has no advantage for the host cell (Ma and O'Farrell 2016). In this case, preferential replication of an mtDNA haplotype would lead to expansion promoted by a 'selfish' replicative advantage that may increase frequency of less efficient or defective mtDNA haplotypes with no advantage to the cell (Moraes et al. 1999, Samuels et al. 2013, Kang et al. 2016, Ma and O'Farrell 2016).

Finally, it has also been argued that a low-level mtDNA haplotype may become homoplasmic purely by genetic drift, without any physiological or pathological advantage (Coller et al. 2001). It has been proposed that since mutant mtDNAs are arising all the time and either loss of the mutant or drift to homoplasmy are the two stable states, one of these possibilities will be achieved after a sufficient number of cell divisions (Coller et al. 2001). Our knowledge of the frequency at which genetic drift may lead to homoplasmy is limited by the uncertainty in the number of mtDNAs that are horizontally transferred (Dong et al. 2017), thus leading to uncertainty of heteroplasmy levels after the mtDNA horizontal transfer.

Understanding which regions of mtDNA genome may confer selective advantage to particular mtDNA haplotypes would be of a great interest, as these findings could teach us about important roles of mitochondria in CTVT, but also in other canine and human cancers.

5.1.3 Goals of my work

As discussed above, the extent of mtDNA horizontal transfer in different tissues and contexts is currently unknown, but our work presented in Chapter 4 may indicate that this phenomenon is possibly more common than previously believed. Importantly, the “triggers” for movement of mtDNA between cells or situations enhancing occurrence of mtDNA horizontal transfer are yet unclear, even though some chemotherapeutic agents have been suggested to increase its frequency.

In this Chapter, I will explain how I used exome sequencing data to perform analyses to understand the relationships between five CTVT mitochondrial clades presented in Chapter 4 (see Figure 5-1).

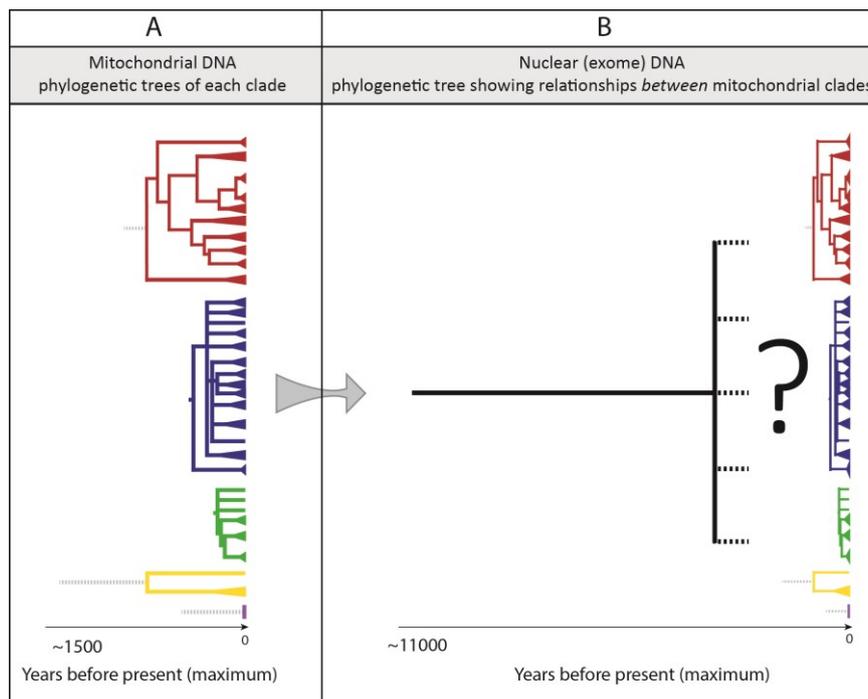


Figure 5-1 Diagram explaining how the analysis of exome sequencing data will allow me to understand the relationships between five CTVT mitochondrial clades analysed in detail in Chapter 4. (A) The phylogenetic tree of each clade stretches only as far back in time as each of the mtDNA horizontal transfer events; (B) while by using the exome data, CTVT nuclear genome can be traced all the way back to the original dog which gave rise to CTVT ~11,000 years ago (Murchison et al. 2014).

Up to this point, all my previous analysis of CTVT mitochondrial clades has been based solely on the mtDNA sequencing data (see Chapter 4), and a clade was defined as a group of CTVT tumours clustering on a single branch of the mtDNA phylogenetic tree, i.e. sharing a common mtDNA evolutionary origin. Remarkably, the analysis presented in this Chapter detected that one of the clades despite clustering on a single phylogenetic branch on the mtDNA tree, clustered on several different branches on the exome phylogenetic tree, thus indicating a recurrent haplotype-specific mtDNA horizontal transfer. The haplogroup undergoing mtDNA horizontal transfer at a significantly higher frequency was a haplogroup in canine mtDNA clade A, specifically the canine A1d1 haplogroup (Fregel et al. 2015).

The finding that the A1d1 haplogroup has undergone recurrent mtDNA horizontal transfer events was unexpected – most importantly, it significantly alters my interpretation from Chapter 4, where all CTVT tumours with A1d1 mtDNA haplogroup were assumed to belong to a single CTVT clade. A1d1 haplogroup is the only haplogroup identified to be captured by CTVT independently multiple times. This initial finding probes a question whether there are any features specific to A1d1 haplogroup that would be responsible for the increased frequency of mtDNA horizontal transfer of A1d1 haplogroup in particular.

With this initial finding in mind, a list of goals for this Chapter is presented below:

- (1)** To understand the relationships between five CTVT mitochondrial clades presented in Chapter 4 using the nuclear (exome) sequencing data from 539 CTVT tumours and 505 matched hosts
- (2)** To determine how many times the A1d1 haplogroup has been independently captured by CTVT, i.e. how many times has A1d1 haplogroup-specific mtDNA horizontal transfer taken place
- (3)** To describe the individual donor mtDNA haplotypes within the A1d1 haplogroup which founded each independent CTVT clade
- (4)** To analyse the distribution and copy number of mtDNA haplogroup A1d1 in the population
- (5)** To analyse the genetic variation specific to A1d1 haplogroup

(6) To investigate two samples where mtDNA recombination involving A1d1 haplogroup mtDNA has taken place (one sample undergoing mtDNA recombination was discussed in Chapter 4)

(7) To describe phenotypic features of A1d1 haplogroup using histopathology data

This work enabled me to conclude that the A1d1 haplogroup has undergone at least 12 independent mtDNA horizontal transfer events into CTVT cells. Analysis of this haplotype highlighted a number of candidate genetic variants, which may be conferring a selective advantage to this haplotype in CTVT, as further discussed in this Chapter.

5.2 Materials and Methods

5.2.1 Sample collection and DNA extraction

Samples were collected and DNA extracted as described in Chapter 3. All tumours were diagnostically confirmed to be CTVT either by amplification of a CTVT specific *LINE-MYC* rearrangement (see Chapter 3) or by histopathology. In total, 1060 tumour and matched host samples were sequenced as part of this study (551 tumours and 509 hosts).

5.2.2 CanFam3.1. pull down kit design

The exome pull-down kit for reference genome CanFam3.1 was designed in collaboration with Agilent Technologies, using the Agilent SureSelect XT Protocol, with the intention to be made publicly available. After selecting the CanFam3.1 protein coding regions to be sequenced, the pull-down probes of length 120bp were designed by the Agilent Technologies team.

Repeat masking on the set of probes designed by Agilent Technologies was performed both by the manufacturers and internally. The repeat masking was performed based on the following criteria:

- Probes with more than 35bp of repetitive sequence present in the genome more than thousand times should be discarded
- Probes with 20bp - 35bp of repetitive sequence may or may not need to be discarded (details from Agilent Technologies are confidential, but criteria are based on GC content, melting temperature and uniqueness of each probe)

Based on the criteria above, the probes were split into 20bp overlapping windows and 35bp overlapping windows and each window was checked for presence in the CanFam3.1 reference genome (filtering was done in collaboration with Dr German Tischler, Wellcome Trust Sanger Institute). Three probes were discarded from the Agilent Technologies design:

- One probe with 35bp window hitting the genome more than 1000 times
- Two probes with more than twenty-one 20bp windows hitting the genome more than 1000 times

5.2.3 DNA sequencing

Whole genome sequencing libraries with insert size of 100 to 400 base pairs (bp) were generated using standard methods according to the Wellcome Trust Sanger Institute exome library preparation protocol. The total number of exome pull-down reactions performed was 1060. Whole exomes were sequenced with 75bp paired end reads on an Illumina HiSeq2000 instrument, using the V3 and V4 sequencing chemistry (Illumina, San Diego, CA). The plexing strategy was as follows: 4 tumour samples per lane, 5.3 host samples per lane. Tumour exomes were sequenced at an average exome depth of coverage $\sim 132X$ (plexed as 4 samples per lane) and host samples were sequenced at an average exome depth of coverage $\sim 104X$ (plexed as 5.3 samples per lane).

5.2.4 Performance of exome pull-down and sequencing

The percentage of exome with at least 30X coverage was used as a metric to determine pull-down and sequencing performance. The average percentage of exome with at least 30X coverage was 85.8% for tumours and 77.5% for hosts (with values above 70% being generally considered acceptable) (see Figure 5-2).

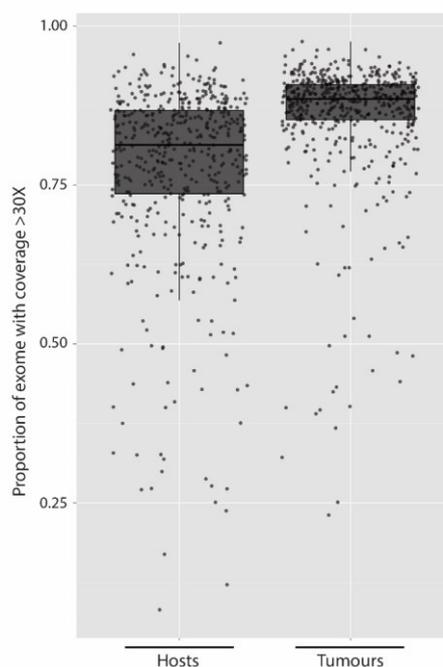


Figure 5-2 Proportion of exome with at least 30X coverage in 509 host samples and 551 CTVT tumour samples. Each individual sample is represented by a dot. Boxes represent the first and third quartiles (inter-quartile range = IQR). Error bars indicate values within 1.5*IQR away from first and third quartiles.

Overall performance of the Agilent SureSelect CanFam3.1 exome pull-down kit (43Mb design) was comparable to the Agilent SureSelect Human pull-down kit currently available. The on-target rate was ~57% (increased to ~72% when expanded to +/-200 bp). For tumour samples (plexed 4 samples per lane), we obtained on average 9.1 Gb of sequencing for 43Mb design. For host samples (plexed 5.3 samples per lane), we obtained on average 6.8 Gb of sequencing for 43Mb design. An example screen shot showing coverage across protein coding regions on chromosome 12 is shown in Figure 5-3.

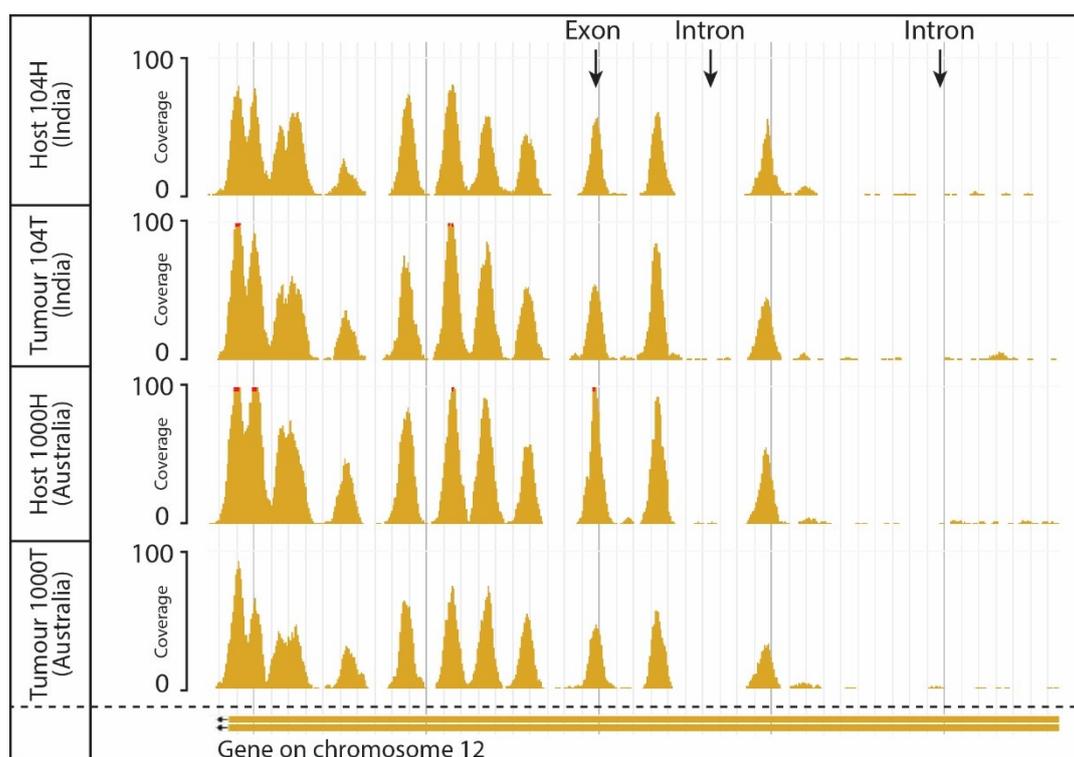


Figure 5-3 Coverage depth across protein coding regions (exons). Data was viewed using JBrowse (Skinner et al. 2009).

5.2.5 Sample validation

All 1060 samples were validated and checked for sample quality and CTVT diagnosis. A number of samples was discarded as they were not CTVT tumours, or removed from the set due to extremely low sequencing coverage and/or low DNA quality. The final set of samples used for analyses presented in this Chapter consisted of 539 CTVT tumour samples and 505 matched host samples.

5.2.6 Substitution calling

Substitutions in mtDNA genomes from the whole exome sequencing data were called and filtered using Mitotypus variant calling pipeline by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group), as previously described (<https://github.com/MaireNiL/mitotypus>, see section 4.2.5, (Strakova, Ni Leathlobhair et al. 2016)). Substitutions in mitochondrial genomes of CTVT samples in clades 5, 8-18 and additional CTVT samples collected in Belize (see section 5.4.3) were called using both Mitotypus and SAMtools mpileup (Li et al. 2009, Li 2011). Substitutions in nuclear genomes from whole exome sequencing data were called and filtered using Somatypus variant calling pipeline by Adrian Baez-Ortega (PhD student in the Transmissible Cancer Group) (Baez-Ortega 2017b).

5.2.7 Indel calling

Small insertions and deletions (indels) in mtDNA genomes were extracted from whole exome sequencing data using SAMtools mpileup (Li et al. 2009, Li 2011). Filtering was only performed in samples, which belong to clades 5, 8-18, as described below (section 5.2.9.6).

5.2.8 Phylogenetic analyses

The mtDNA phylogenetic tree represented in Figure 5-4A was constructed using a maximum likelihood method implemented in RAxML (Stamatakis 2014) by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group). The nuclear DNA (exome) phylogenetic tree represented in Figure 5-4B was constructed using a maximum likelihood method implemented in RAxML (Stamatakis 2014) by Adrian Baez-Ortega (PhD student in the Transmissible Cancer Group) (Baez-Ortega 2017a).

I used both mitochondrial and nuclear phylogenetic trees to define the total number of recurrent mtDNA horizontal transfers of haplotypes from the A1d1 haplogroup. Each individual horizontal transfer gave rise to a single clade. Each separate clade arising from the A1d1 haplogroup was defined as follows:

- (1) CTVT tumours from the same clade arising from the A1d1 haplogroup cluster on a single branch within the green 'clade 3' cluster on mitochondrial DNA phylogenetic tree represented in Figure 5-4A.

- (2) CTVT tumours from the same clade arose as a result of a single horizontal transfer event of an mtDNA donor haplotype from the A1d1 haplogroup (section 5.2.9) and share the same set of germline substitutions which was inferred to be present on the donor mtDNA haplotype.
- (3) CTVT tumours from the same clade cluster on a single branch of the nuclear DNA phylogenetic tree represented in Figure 5-4B.

Should points 1) and 2) above hold true, while crucial condition 3) would not be fulfilled, the samples would be classed as separate clades.

5.2.9 Analysis of A1d1 haplogroup

The haplogroup naming system was adapted from the cladistics canine mtDNA phylogeny nomenclature proposed by Fregel et al. (Fregel et al. 2015).

5.2.9.1 *Donor haplotypes*

Current population of 338 canine host samples was analysed for presence of mtDNA haplotypes within A1d1 haplogroup. The most closely related mtDNA haplotype from the normal dog population of 557 hosts is presented in Table 5-4A. The 'donor haplotype' was reconstructed for each clade arising from the A1d1 haplogroup, representing the inferred donor mtDNA haplotype in each individual horizontal transfer event (note that the donor haplotypes all belonged to the A1d1 haplogroup and therefore are very similar). The donor haplotypes were reconstructed from the clade-defining germline substitutions and the clade-defining potential somatic substitutions as previously described (see Chapter 4, section 4.2.8.2) and are shown in Additional file 5-1.

5.2.9.2 *A1d1 haplogroup frequency and distribution*

Frequency and worldwide distribution of A1d1 haplogroup was analysed within a population of 338 canine host samples. Host samples and the locations they were collected from are shown on a map (Additional file 4-13B).

5.2.9.3 *A1d1 haplogroup copy number*

To calculate the mtDNA copy number, I used the following equation: $(\text{mtCOV}/\text{nuclCOV}) * P$, where mtCOV = average coverage across the mitochondria, nuclCOV = average coverage across the nuclear genome and P = ploidy. The ploidy used in my calculations was 2 for both

CTVT tumours and CTVT hosts. Tumour fraction (mtDNA and nuclear) was taken into account when performing mtDNA copy number calculations in CTVT tumours. MtDNA tumour fraction was estimated from variant allele fraction (VAF) plots (see Chapter 4, section 4.2.5.3 for further information about VAF plots). Nuclear tumour fraction was estimated from whole exome sequencing data by Adrian Baez-Ortega (PhD student in the Transmissible Cancer Group) using the following equation: $(\text{mode of T-VAF}) \times 2$, where T-VAF = somatic tumour variant allele fraction.

5.2.9.4 *A1d1 haplogroup long-term heteroplasmy analysis*

Whole genome sequencing data from thirteen CTVT samples were analysed for variants specific to A1d1 haplogroup. Seven of these samples were not of A1d1 haplogroup, neither did they have host of A1d1 haplogroup, and so were used for final analysis. Whole genome sequencing libraries with insert size of 400 to 600 base pairs (bp) and clean-up to ensure peak at size 450bp were generated using standard methods according to manufacturer's instructions and sequenced with 150bp paired end reads on an Illumina HiSeq X Ten instrument (Illumina, San Diego, CA) at the Wellcome Trust Sanger Institute, Hinxton. Two lanes were sequenced for tumour samples and one lane for host samples. Coverage across each base was obtained for the following A1d1 specific mtDNA positions: 2683, 2962, 3196, 5367, 5444, 6065, 6401, 7593, 8281, 8368, 8807, 10992, 13299, 15214, 15627, 15639 and 15814. Average coverage across the mtDNA genome in these samples was calculated using average coverage in all thirteen samples across the positions listed above. Average number of mtDNA copies in CTVT was obtained from our previous analysis (see Chapter 4, section 4.2.4 and (Strakova et al. 2016)). The expected coverage across a base, should there be one mtDNA copy of A1d1 haplogroup, was calculated using the following equation: $\text{COV}(\text{mtDNA})/\text{CN}(\text{mtDNA})$, where $\text{COV}(\text{mtDNA})$ = average coverage across mtDNA, $\text{CN}(\text{mtDNA})$ = number of mtDNA copies present in CTVT.

5.2.9.5 *A1d1 haplogroup substitutions*

Substitution variants which were common and unique to all samples in haplogroup A1d1 were identified through filtering against other samples. Substitutions defining A1d1 haplogroup and sub-haplogroups A1d1a and A1d1a1 were highlighted as potential candidate variants. Due to possibility that some substitutions have been lost as a result of mtDNA recombination, substitutions missing from samples 559T and 1315T (samples undergoing mtDNA recombination) have also been included in further analysis as potential candidate

variants. Changes in protein structure as a result of non-synonymous substitution at position 7593 were analysed using RaptorX (Kallberg et al. 2012).

5.2.9.6 *A1d1a haplogroup indels*

Small insertions and deletions (indels) which were common and unique to all samples in haplogroup A1d1 were identified and filtered against indels present in the population. Indels defining haplogroup A1d1 and sub-haplogroups A1d1a and A1d1a1 were highlighted as potential candidate variants and validated using IGV (Robinson et al. 2011, Thorvaldsdottir et al. 2013).

5.2.9.7 *D-loop repeat polymerase chain reaction*

Length variation of the mtDNA control region caused by polymorphisms in copy number of a 10bp repeat unit (GTACACGT(A/G)C) was assessed using a polymerase chain reaction (PCR) assay. Each PCR reaction was performed along with a positive control (see Chapter 4, section 4.2.9.3) using an Eppendorf Mastercycler Nexus GSX1 instrument, with primers and conditions specified in Table 5-1 and Table 5-2.

Primer		Sequence
MT-repeat primers (spanning the 10bp repeat) (obtained from (Gundry et al. 2007))	Forward	CTC ACG CAT AAA ATC AAG GTG
	Reverse	GGG TTT GGC GGG ACA TAA
MT-short previously tested positive control primers (MT region positions: 4450-5024) (section 4.2.9.3)	Forward	AGC ATA CTC CTC CAT TGC CC
	Reverse	TCA GGG GTT AAG TGT GTA GCA

Table 5-1 Primers used to assess length variation of mtDNA control region caused by polymorphisms in copy number of a 10bp unit.

PCR master mix reagents	Volume per reaction (µl)
10X PCR buffer	2
dNTP mix (2.5mM of each dNTP, 10mM total)	1.6
Qiagen Taq DNA polymerase (5 units/µl)	0.1
Primer forward (10µM)	1.5
Primer reverse (10µM)	1.5
DNA (WGA, ~40ng/µl)	2
Distilled water	11.3
Total volume	20

Table 5-2 Polymerase chain reaction conditions used to assess length variation of mtDNA control region caused by polymorphisms in copy number of a 10bp unit.

5.2.9.8 *mtDNA recombination analysis*

5.2.9.8.1 *Illumina data analysis*

The mitochondrial DNA phylogenetic tree constructed by Máire Ní Leathlobhair (PhD student in the Transmissible Cancer Group) (see section 5.2.8) was searched for outliers as previously described (see Chapter 4, section 4.2.9.2). Variant allele fraction (VAF) plots were plotted for samples 559T and 1315T using the VAF values obtained from the Illumina sequencing data during variant calling (see section 5.2.6) and checked for regions of homoplasmy.

5.2.9.8.2 *Long read sequencing*

A genomic library was created directly using 5µg of genomic DNA from samples 559T and 1315T, not utilising shearing or amplification techniques, as previously described (Chapter 4, section 4.2.9.4, see also (Strakova et al. 2016)). PacBio (Pacific Biosciences) data reads aligned to CanFam3.1 mtDNA genome were viewed in SMRT view (Pacific Biosciences) and in Integrative Genomics Viewer (IGV) (Robinson et al. 2011, Thorvaldsdottir et al. 2013). The three or two most common haplotypes in 559T and 1315T respectively were phased, as shown in Figure 5-13. Additional haplotypes present at a low level (less than 5%) were also identified in both samples and are not shown in Figure 5-13. Indels were not included due to uncertainty in making calls using PacBio data.

5.2.10 Somatic tumour substitutions and timing analysis

CTVT tumour substitutions in clades 5, 8-18 were classified as follows as previously described (see Chapter 4, section 4.2.7.2.1, see also (Strakova et al. 2016)): tumour germline clade-defining substitutions (substitutions present on most closely related dog mtDNA haplotype), tumour potential somatic substitutions (substitutions absent from most closely related dog mtDNA haplotype but common to all tumours within a clade) and tumour somatic substitutions (substitutions inferred to have arisen after the horizontal transfer event). Number of somatic substitutions was used to estimate the age of clades 5, 8-18, using previously reported timing estimates (see Chapter 4, section 4.2.7.3 for timing calculations), assuming a constant accumulation of mutations within and between clades, (Strakova et al. 2016). Clade 18 was excluded from the timing calculations as both samples have undergone mtDNA recombination.

5.2.11 Phenotypic analyses of tumours in A1d1 haplogroup

Phenotypic features of A1d1 haplogroup were analysed through a histopathology screen.

5.2.11.1 *Histopathology scoring*

A blinded histopathology scoring study was performed. Thirty-one validated histology samples from tumour clades 5, 8-18 were randomly chosen, together with thirty-one control tumour samples from other clades. Control samples were chosen to be the most closely related tumour samples on the nuclear exome tree (see Figure 5-4B), which did not arise from the A1d1 haplogroup. Histology slides were scanned using the Hamamatsu Nanozoomer 2.0-HT slide scanner (C9600) at the Wellcome Trust Sanger Institute, Hinxton and were viewed using the NDP.view2 software (Hamamatsu).

Two scores were chosen for the screen – total number of mitoses per 10 high-power (40X) fields as a measure of proliferative capacity and percentage of lymphocytes within the parenchyma which had previously been reported to increase in vincristine-induced regression in CTVT (Gonzalez et al. 2000). Scoring field selection criteria and the scoring criteria for each variable are described in Chapter 4, section 4.2.12.2.

5.3 Results

5.3.1 Samples analysed in this Chapter

The number of samples submitted for exome sequencing was 1060 tumour and host samples. The final set of samples after validation that was used for analysis presented in this Chapter consisted of 539 confirmed CTVT tumours (338 of which were also included in analysis presented in Chapter 4) and 505 matched hosts (see Materials and Methods, section 5.2.5). We started the analysis by assigning samples to CTVT mtDNA clades. We identified 207 samples in clade 1, 285 samples in clade 2, 36 samples in clade 3, 6 samples in clade 4, 2 samples in clade 5. In addition, two new CTVT clades were discovered: 2 samples in clade 6 (found in Armenia), 1 sample in clade 7 (found in Mexico).

5.3.2 Recurrent and recent haplotype-specific mtDNA horizontal transfer

In order to determine how the five CTVT mtDNA clades presented in Chapter 4 relate to each other, analysis of the maximum likelihood phylogenetic tree constructed using 539 CTVT exome samples was performed (see section 5.2.8). Remarkably, this analysis revealed that one of the clades reported in Chapter 4 (clade 3), despite clustering on a single branch of the mtDNA tree (Figure 5-4A), fell in 11 different locations on the nuclear (exome) DNA tree (highlighted by arrows in Figure 5-4). This finding indicates that a single canine mtDNA haplogroup, originally believed to only give rise to clade 3, has recurrently undergone multiple independent horizontal transfer events. In contrast to my interpretation from Chapter 4, this indicates that clade 3 is not a single CTVT mtDNA tumour clade, but 11 independent mtDNA tumour clades - these were now renamed to clades 8-18 (with the name 'clade 3' discarded from the naming system) (see Figure 5-4).

We had previously noted (see Chapter 4), that clade 3 tumours are not monophyletic with respect to haplotypes found in normal dogs (Figure 4-13 and Additional file 4-12). However, given that the same mutational processes operate on mtDNA in the dog germline and in CTVT, germline and somatic variants often occur at the same sites. Therefore, we interpreted clade 3 as a single clonal lineage that had acquired somatic mutations that were also found in germline haplotypes closely related to the donor haplotype.

CTVT clade 5, described in Chapter 4, maps close to clade 3 on the mtDNA tree (Figure 4-13, Figure 5-4A and Additional file 4-12). However, we previously considered that this was likely

to be a separate clade, as it (1) did not share key germline variants with the clade 3 donor haplotype, suggesting that it was derived from a separate donor haplotype and (2) clade 5 carried considerably more somatic mutations than other tumours in clade 3 (see Table 4-11), suggesting an earlier and separate horizontal transfer event. However, given that our exome analysis indicated that 'clade 3' tumours actually belong to 11 distinct horizontal transfer groups, we considered the possibility that the similarity in donor haplotype between clade 5 and the donor haplotypes in the tumours previously grouped as 'clade 3' may suggest that the clade 5 donor haplotype may carry the same positively selected genetic variants as those haplotypes that founded clades 8 – 18. Clade 5 tumours are considered alongside clades 8 – 18 in the following discussion.

As indicated from the nuclear DNA phylogenetic tree (see Figure 5-4B) and from the map in Figure 5-5 showing geographical distribution of mtDNA clades 5, 8-18; tumours in these CTVT clades originated from a large number of countries around the world – indeed, we detected clades 5, 8-18 in Belize, Chile, Colombia, Grenada, India, Nicaragua, Nigeria, Paraguay and The Gambia. We had previously noted that the worldwide distribution of clades 5, 8-18 (in contrast to clades 1 and 2, as discussed in Chapter 4) cannot be easily explained by historical travel routes; this difficulty is now explained by the finding that these tumours have 12 independent origins.

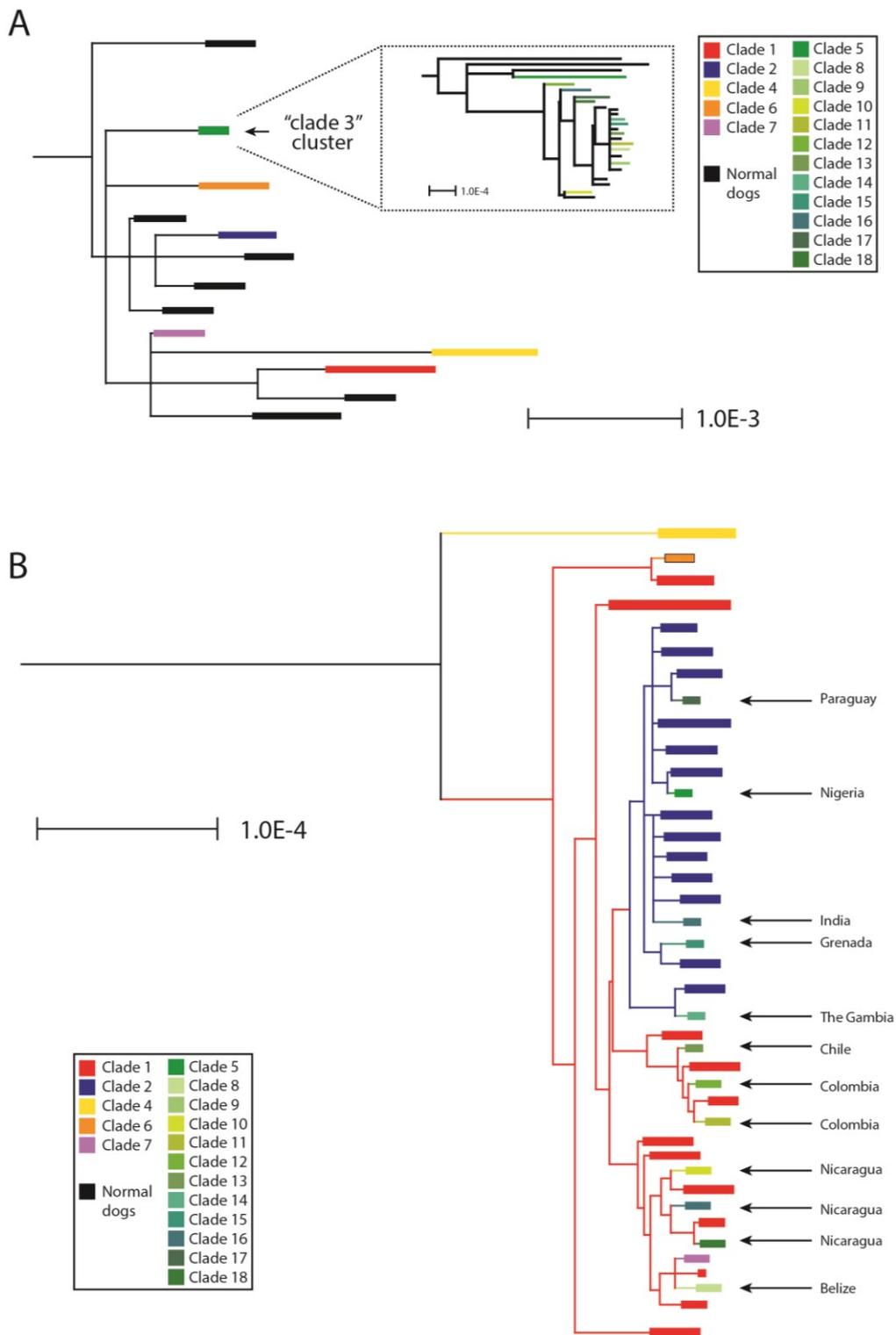


Figure 5-4 Recurrent and recent haplotype specific mtDNA horizontal transfer. Simplified representation of maximum likelihood phylogenetic trees for (A) mtDNA (with 'clade 3' zoom-in in a separate box). Scale bar indicates base substitutions per site. Note that only canine haplogroup A is shown (see Figure 4-13). Name 'clade 3' was discarded from the naming system and replaced by 'clades 8-18'. (B) nuclear DNA (based on exome sequencing). Samples with A1d1/A1d1a/A1d1a1 mitochondrial haplotype are indicated by shades of green (note that clade 5 is included within this group). Scale bar indicates base substitutions per site.

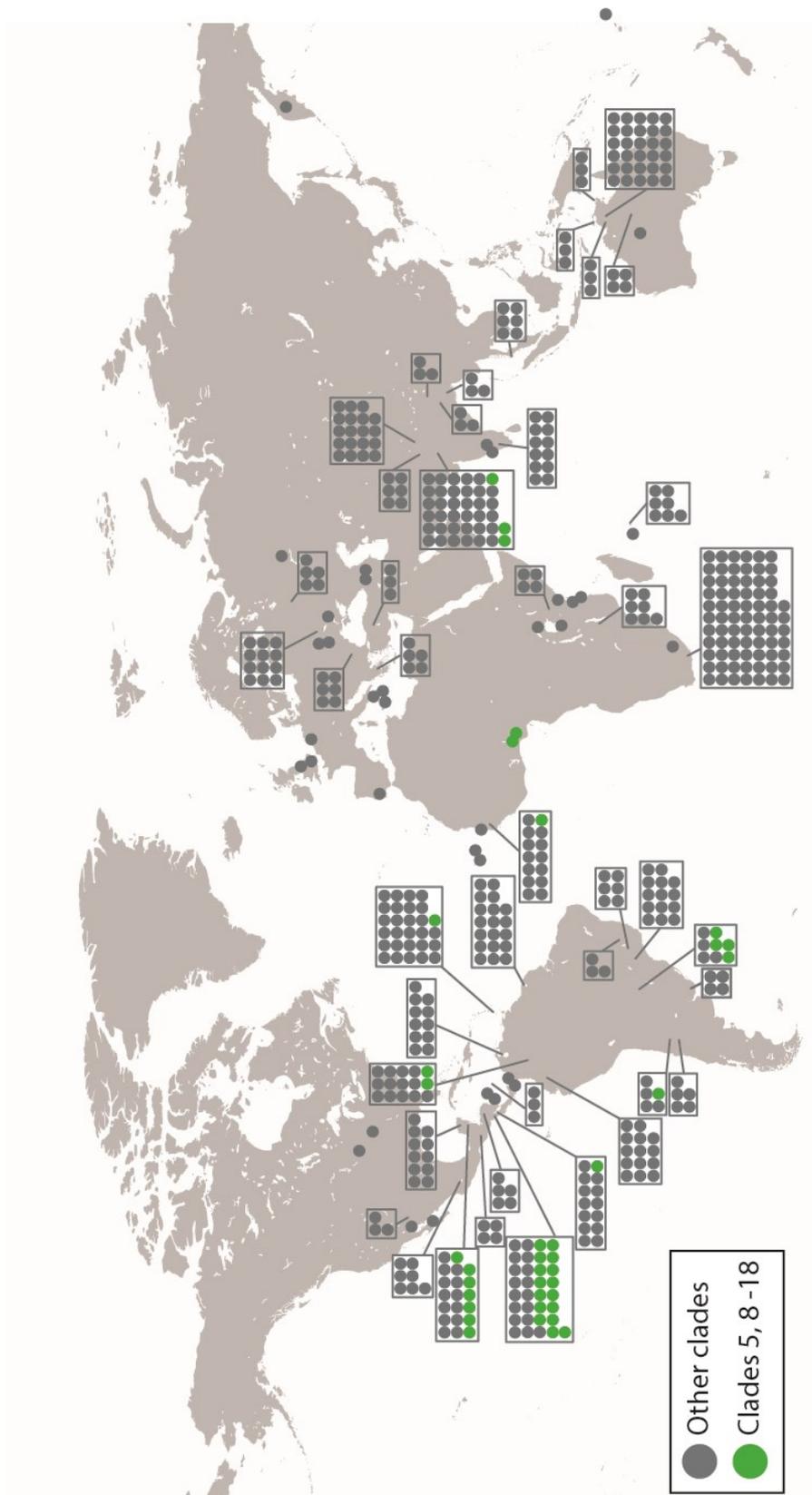


Figure 5-5 Worldwide distribution of CTVT mtDNA clades 5, 8-18. Clades 5, 8-18 are represented by green dots, other clades are represented by grey dots. Total number of CTVT samples is 539.

5.3.3 Analysis of the A1d1 haplogroup

Based on the cladistic canine mitochondrial DNA phylogeny, the mtDNA haplogroup that has undergone multiple independent mtDNA horizontal transfers falls into the A1d1 mtDNA haplogroup (Fregel et al. 2015). I therefore next searched for characteristics specific to A1d1. I assessed the frequency of A1d1 in normal dog population, analysed A1d1 mtDNA copy number in hosts and tumours, considered the possibility of long-term A1d1 heteroplasmy and investigated various genetic features of this mtDNA haplotype.

5.3.3.1 Identification of donor haplotypes

The A1d1 haplogroup is composed of a number of closely related sub-haplogroups, which are displayed in Figure 5-6, with sub-haplogroups involved in independent mtDNA horizontal transfers highlighted in green. The frequency of each sub-haplogroup in current dog population is shown in Table 5-3.

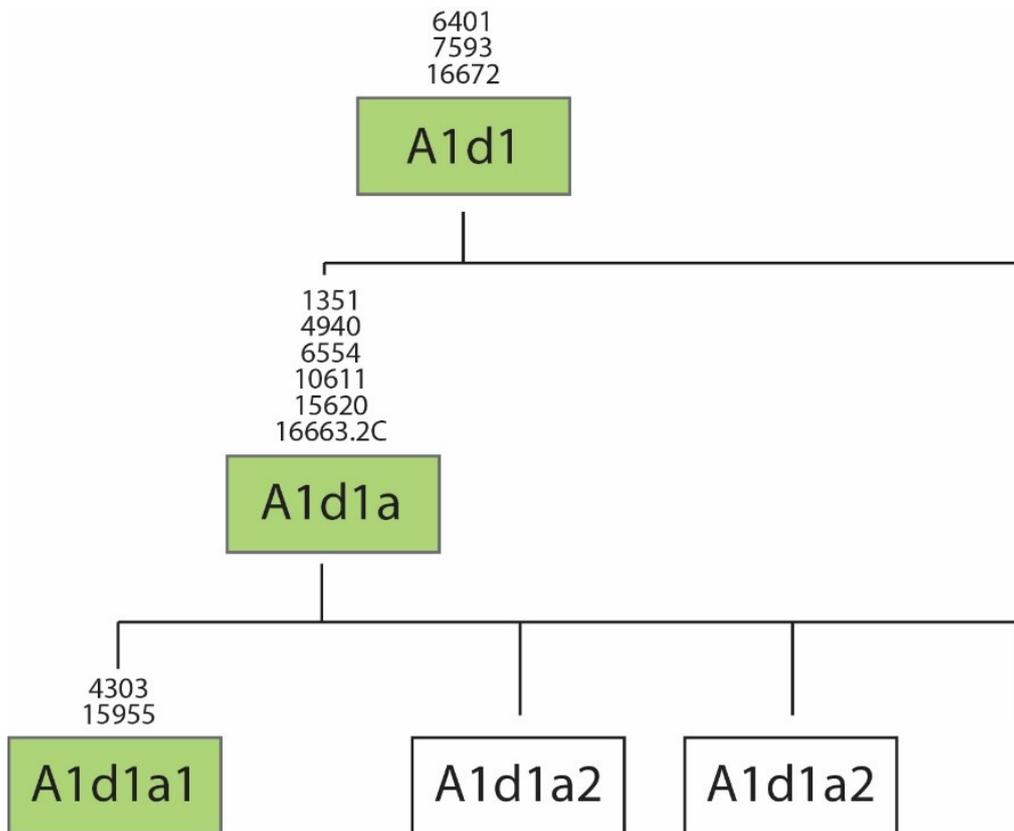


Figure 5-6 Schematic diagram of the A1d1 haplogroup phylogenetic tree adapted from Fregel et al. (Fregel et al. 2015). Sub-haplogroups are displayed only up to the level with six letters/digits. Sub-haplogroups involved in recurrent mtDNA horizontal transfer are highlighted in green and their diagnostic variants are displayed.

Haplogroup name	Number of times haplogroup found in host population	Number of CTVT clades founded by this haplogroup, with list of clade numbers in brackets
A1d1	1	1 (clade 5)
A1d1a	20	5 (clades 10, 12, 16, 17, 18)
A1d1a1	51	6 (clades 8, 9, 11, 13, 14, 15)

Table 5-3 Sub-haplogroups of the A1d1 haplogroup. Frequency in the current dog population (n = 338 hosts) and number of CTVT mtDNA clades founded by each sub-haplogroup are shown.

We observed 41 distinct host mtDNA haplotypes within our set of canine hosts, which belong to the A1d1 haplogroup. The most phylogenetically closely related mtDNA donor haplotypes, which founded each of clades 5 and 8-18, were identified in host population and are presented in Table 5-4.

CTVT mtDNA clade	Clade location	Number of samples belonging to this clade	Phylogenetically closest haplotype in the host population	Number of times host haplotype found in host population
Clade 5	Nigeria	2	A1d1_2	1
Clade 8	Belize	7	A1d1a1_19	1
Clade 9	Nicaragua	14	A1d1a1_6	1
Clade 10	Nicaragua	1	A1d1a_15	1
Clade 11	Colombia	1	A1d1a1_19	1
Clade 12	Colombia	1	A1d1a_16	1
Clade 13	Chile	1	A1d1a1_8	1
Clade 14	The Gambia	1	A1d1a1_17	1
Clade 15	Grenada	1	A1d1a1_27	2
Clade 16	India	3	A1d1a_3	6
Clade 17	Paraguay	4	A1d1a_1	2
Clade 18	Nicaragua	2	A1d1a_1	2

Table 5-4 Locations and number of samples in clades 5, 8-18 and their phylogenetically closest haplotype present in the current dog population. Names of individual host haplotypes are based on workings in Strakova et al. (Strakova et al. 2016).

5.3.3.2 Worldwide distribution

I analysed the distribution of the A1d1 haplogroup in the normal canine population, as increased frequency of A1d1 haplogroup in areas where recurrent haplotype specific mtDNA horizontal transfer has taken place could explain the increased number of mtDNA horizontal transfer events. My analysis has shown, however, that the A1d1 mtDNA haplogroup is found in host populations around the world (see Figure 5-7). Interestingly, the A1d1 haplotype, which founded clade 5, found in Nigeria, is only found in West Africa.

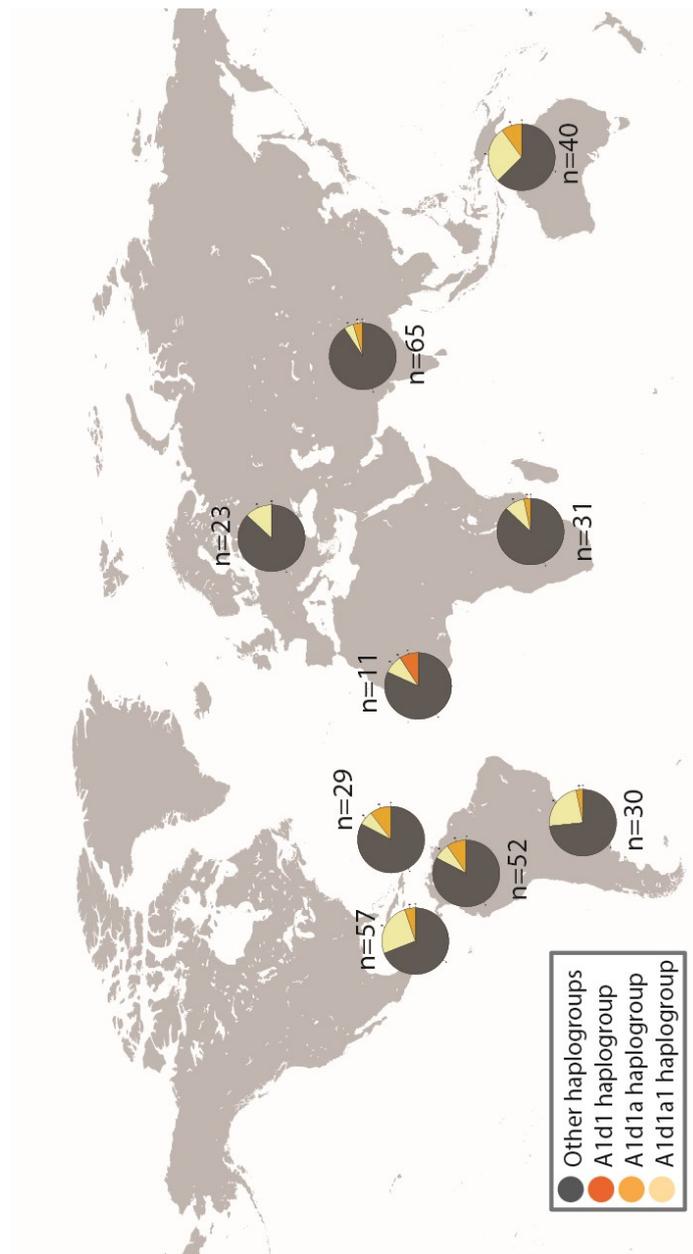


Figure 5-7 Distribution of A1d1 haplogroup in the current normal canine population around the world (n = 338 host samples).

5.3.3.3 mtDNA copy number

Increased copy number of A1d1 haplogroup mtDNA within the donor host tissue may increase the chances of an mtDNA haplotype specific horizontal transfer taking place; therefore explaining the increased number of mtDNA horizontal transfer events of A1d1 haplogroup. Analysis of mtDNA copy number of hosts with A1d1 haplotype and of tumours in clades 5, 8-18 is discussed below.

5.3.3.3.1 mtDNA copy number in hosts

Analysis of mtDNA copy number in the ovarian and testicular tissue of canine host samples, and comparison between A1d1 haplogroup and other haplogroups, showed no significant differences (see Figure 5-8). It should, however, be noted that CTVT tumours grow within stromal compartments, rather than within the internal ovarian or testicular tissue. Increased mtDNA copy number in stromal compartments would not be captured by this analysis.

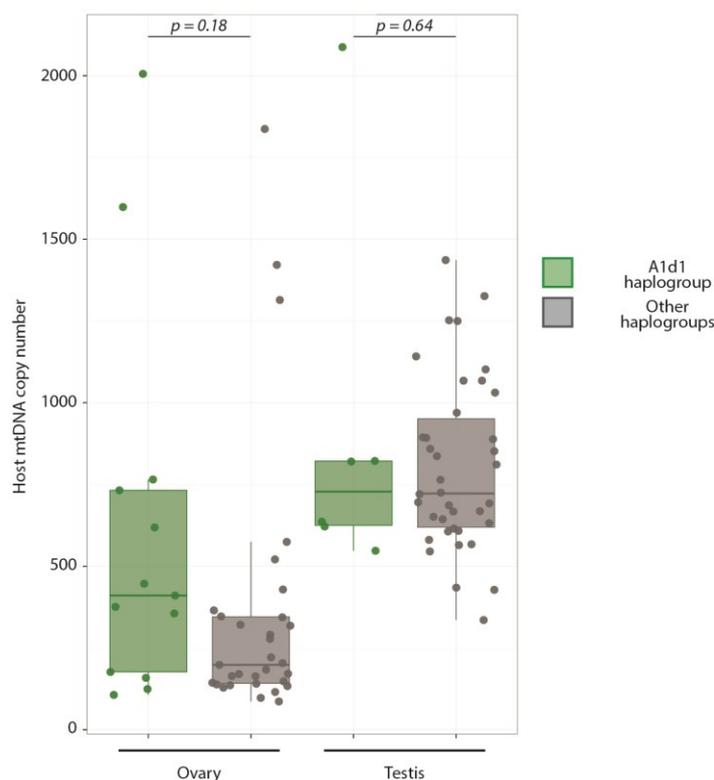


Figure 5-8 mtDNA copy number of A1d1 haplogroup versus other haplogroups in normal ovarian and testicular tissue. Each individual value is represented by a dot. Boxes represent the first and third quartiles (inter-quartile range = IQR). Error bars indicate values within 1.5*IQR away from first and third quartiles. P-values were calculated using two-sided student's t-test. Only host samples where tissue of origin is known have been included in this analysis (n = 88 host samples).

5.3.3.3.2 mtDNA copy number in tumours

Comparison of mtDNA copy number (normalised for host contamination, see section 5.2.9.3) between CTVT clades indicated that mtDNA copy number of clades 5, 8-18 did not significantly differ to mtDNA copy number of other clades (see Figure 5-9). Clade 4 has been indicated to significantly differ in mtDNA copy number, but this result should be treated with caution due to the very small sample size ($n = 3$).

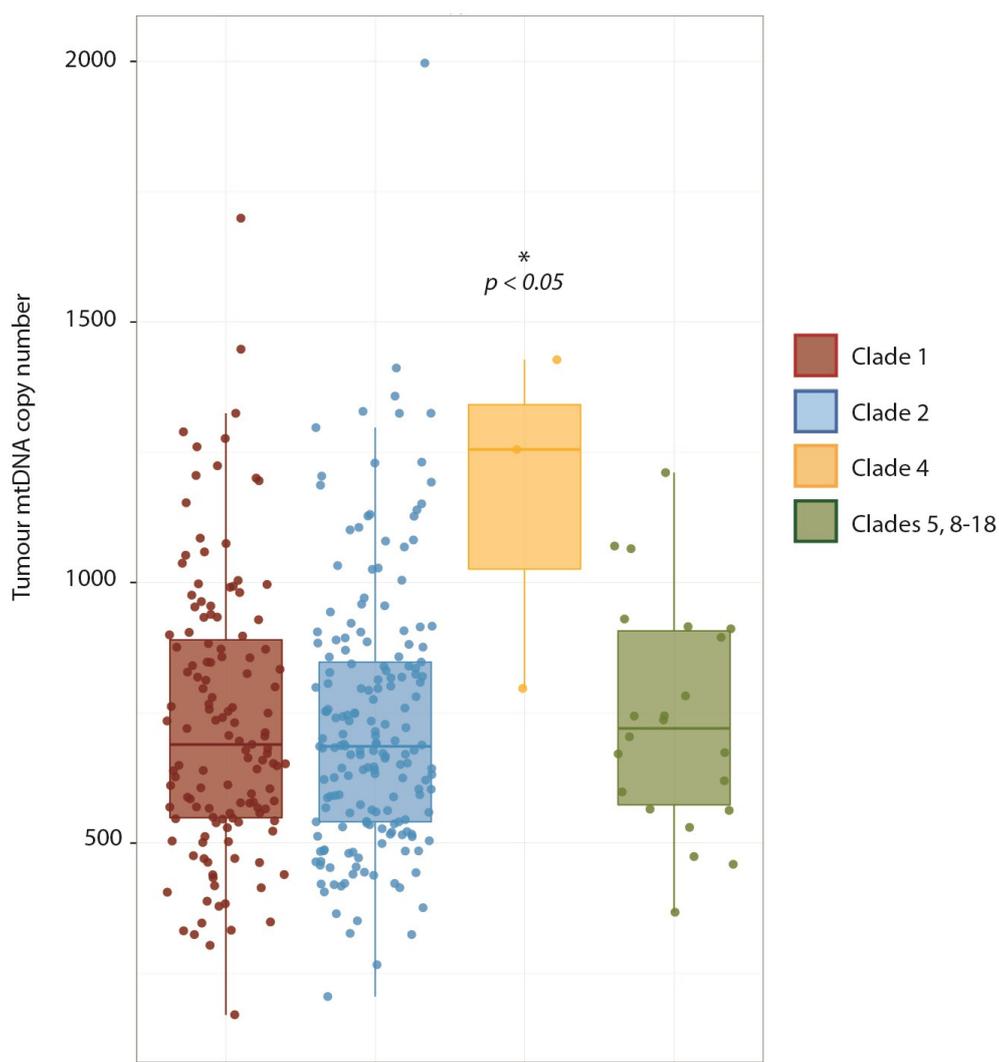


Figure 5-9 Tumour mtDNA copy number adjusted for host contamination in different CTVT mitochondrial clades. Each individual value is represented by a dot. Boxes represent the first and third quartiles (inter-quartile range = IQR). Error bars indicate values within $1.5 \times \text{IQR}$ away from first and third quartiles. P-values were calculated using ANOVA in R (R Core Team 2013), to investigate clade effect on copy number. Only CTVT tumours sequenced using whole genome low coverage sequencing were used for this analysis ($n=447$).

5.3.3.4 Long-term heteroplasmy

We have considered the possibility that A1d1 haplogroup mtDNA may be maintained in all tumours of the CTVT lineage at low levels, and under certain conditions become the most dominant mtDNA haplotype. To test whether the A1d1 haplogroup may be present in all CTVT tumours at low levels, and thus ruling out my interpretation of recurrent haplotype specific mtDNA horizontal transfer events, I analysed whole genomes of CTVT samples sequenced at high coverage. Given the average coverage of 14,335X across the mtDNA in these samples (see section 5.2.9.4) and given that each CTVT cell carries approximately 470 mtDNA copies (see Chapter 4 and (Strakova et al. 2016)), the expected coverage we would see if one mtDNA copy of A1d1 haplogroup was present would be approximately 31X. Analysis of tumours in clades 1, 2 and 4, with corresponding hosts not in the A1d1 haplogroup, did not identify any low level A1d1 haplogroup mtDNA in these samples (see Table 5-5).

Sample name	399T	401T	423T	556T	608T	609T	683T
Average coverage across mtDNA	~11060	~11680	~9820	~17530	~7870	~9600	~19430
Read count: 6401 C>T	6	8	12	18	5	5	14
Read count: 7593 T>C	2	1	2	2	1	2	8
Control: 5444, base T	5	2	2	3	2	3	1
Control: 6065, base A	5	6	7554	20	2	7	13

Table 5-5 Long-term low-level heteroplasmy of A1d1 haplogroup. Table showing read counts across two mtDNA positions uniquely defining of the A1d1 haplogroup, indicating sequencing errors (position 6401 and 7593). Read counts across control positions 5444 and 6065 are shown in bottom rows, indicating sequencing errors. Control position 6065 in sample 423T shows an example of a real variant in this sample (read count does not reach average coverage across mtDNA in this sample, due to absence of this variant from matched host sample). Average mtDNA coverage for each sample is displayed.

5.3.3.5 Genetic features

As demonstrated above, I did not identify any differences in worldwide distribution or copy number of A1d1 haplogroup, and I ruled out the possibility that this phenomenon could be a result of long-term low-level heteroplasmy of A1d1 haplogroup in CTVT. This leaves the possibility that the specific mtDNA sequence of A1d1 haplogroup is responsible for the

recurrent and A1d1 haplogroup specific mtDNA horizontal transfer. I therefore carefully analysed the genetic variation specific to A1d1 haplogroup, as presented below.

5.3.3.5.1 Substitutions

Substitutions shared by all mtDNA haplotypes within A1d1 haplogroup and not present in any other haplogroups (i.e. specific to A1d1 haplogroup) are presented in Table 5-6 (see also Figure 5-6). Predicted functional consequences of substitutions which were common to all CTVT samples in clades 5, 8-18 with donor haplotypes in haplogroup A1d1 or sub-haplogroup A1d1a (see Table 5-4) are shown in Figure 5-10.

Haplogroup name	Substitutions shared by all haplotypes in this haplogroup
A1d1	6401 C>T, 7593 T>C
A1d1a	6401 C>T, 7593 T>C, 1351 A>G, 4940 T>C, 6554 T>C, 10611 A>T, 15620 T>C
A1d1a1	6401 C>T, 7593 T>C, 1351 A>G, 4940 T>C, 6554 T>C, 10611 A>T, 15620 T>C, 4303 A>G, 15955 C>T

Table 5-6 Substitutions shared by all mtDNA haplotypes in haplogroups A1d1, A1d1a, A1d1a1.

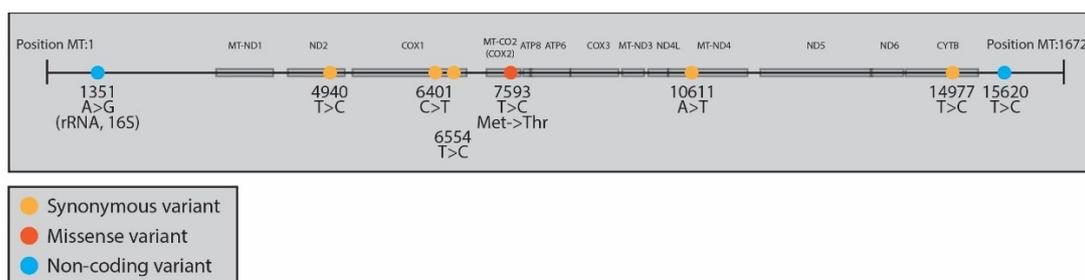


Figure 5-10 Predicted functional consequences of variants shared by all CTVT samples in clades 5, 8-18 with donor haplotypes in haplogroup A1d1 or sub-haplogroup A1d1a. MtDNA genes are represented in the diagram. Base change is displayed below each variant position.

The only missense substitution (position 7593 T>C) (see Figure 5-10) was a methionine to threonine change found within the C-terminal of the cytochrome c oxidase subunit II (COX2) gene, not leading to any apparent changes in the tertiary structure of the protein (see Materials and Methods, section 5.2.9.5). This substitution is equivalent to a variant found in

the normal human population at position 8145 (Marzuki et al. 1991, Lott et al. 2013), with C being the more commonly present base.

The two non-coding variants (positions 1351 A>G and 15620 T>C) were found within the 16S subunit of rRNA and within the D-loop of the mtDNA control region respectively. Variant at position 1351 A>G was in a region conserved between dogs and humans, equivalent to a known variant at position 1935 A>G in human mtDNA (Lott et al. 2013). Variant at position 15620 T>C was in a region partly conserved between dogs and humans, equivalent to position 16247 in human mtDNA, where the reference base in humans was A instead of T, with the most common variant in humans being A>G (Lott et al. 2013).

5.3.3.5.2 Indels

There were no small insertions or deletions (indels) common to the A1d1 haplogroup. However, a single indel was common to a proportion of haplotypes in A1d1 haplogroup - all haplotypes in A1d1a sub-haplogroup (see Table 5-7 and Figure 5-11). The only indel was an insertion of 2Cs at position 16660, which was very close to a non-unique variant at 16672 C>T, leading to the following DNA sequence change in this region: 16660 CCCTTTTTTTTCCC > CCCCTTTTTTTTCC. This indel is found in the mtDNA control region, which is not conserved between dogs and humans. Nevertheless, it is clear that the indel is present within the region between CSB-III (conserved sequence block III) and tRNA^{Phe}. Since HSP (heavy-strand promoter) and LSP (light-strand promoter) have not been specifically defined within the canine genome, the exact position of this indel with respect to HSP and LSP is unclear.

Haplogroup name	Indels shared by all haplotypes in this haplogroup
A1d1	None
A1d1a	16660 CCCTTTTTTTTCCC > CCCCTTTTTTTTCC
A1d1a1	16660 CCCTTTTTTTTCCC > CCCCTTTTTTTTCC

Table 5-7 Indels shared by all mtDNA haplotypes in haplogroups A1d1, A1d1a, A1d1a1.

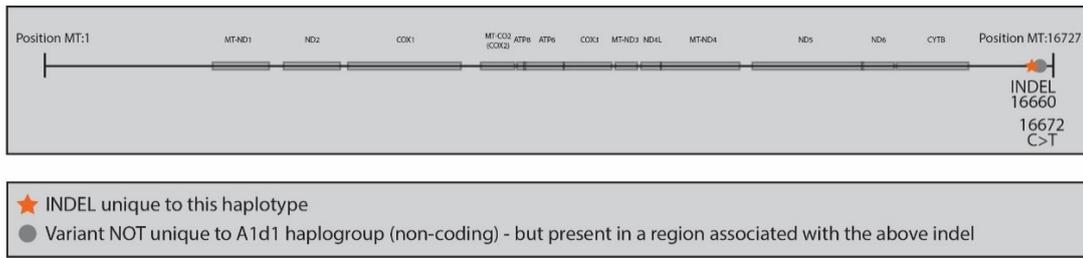


Figure 5-11 Graphical illustration of indel common to all haplotypes in A1d1a haplogroup.

5.3.3.5.3 Copy number variation of 10bp repeat unit in the mtDNA control region

Previous literature has reported copy number variation of a 10bp repeat unit (5'-GTACACGT(A/G)C-3') between positions 16130-16429 of the canine mtDNA control region (Gundry et al. 2007). As the mtDNA control region has been highlighted as identical between recombining samples, I next examined the copy number of this repeat unit, further searching for A1d1 haplogroup specific features. PCR analysis (see Materials and Methods, section 5.2.9.7) has shown that the length of this repeat region is not conserved amongst the A1d1 haplogroup (see Figure 5-12).

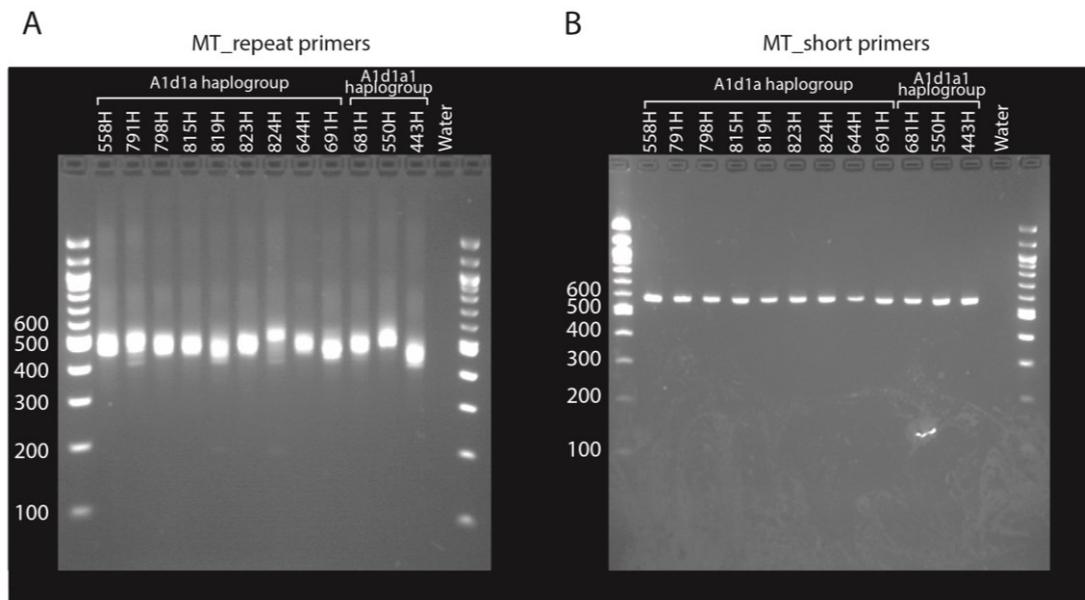


Figure 5-12 Copy number variation of the 10bp repeat unit within the canine mtDNA control region. Polymerase chain reaction showing (A) length variation of region spanning the 10bp repeat unit in samples in A1d1 haplogroup (A1d1a and A1d1a1 are sub-haplogroups of A1d1) (B) reaction with a control set of mtDNA primers in the same set of samples.

5.3.3.5.4 mtDNA recombination

MtDNA recombination involving A1d1 haplotypes

Although mtDNA is usually considered not to undergo recombination, our previous analysis of mtDNA in CTVT presented the first evidence of mtDNA recombination in cancer (see Chapter 4, (Strakova et al. 2016)). Analysis of additional samples sequenced in this Chapter (see section 5.3.1, for recombination analysis see Materials and Methods, section 5.2.9.8 and Chapter 4) revealed presence of mtDNA recombination in another sample from Nicaragua, sample 1315T. The two CTVT samples undergoing recombination were therefore samples 559T and 1315T. Strikingly, mtDNA recombination in both samples involved mtDNA haplotypes from the A1d1 haplogroup (see Table 5-4, CTVT clade 18).

Analysis of genetic variants shared by these two samples provides an opportunity to understand which variants from the A1d1 haplogroup, if any, had risen to higher frequency within the recombining samples. Such analysis may highlight genetic variation conferring advantage to the A1d1 haplogroup and thus may help to explain why the A1d1 haplogroup has undergone recurrent mtDNA horizontal transfer.

I used long-read sequencing (see section 5.2.9.8, Materials and Methods) to phase mtDNA variants and reconstruct haplotypes present within each of recombining samples 559T and 1315T, as shown in Figure 5-13. The recombination products were all derived from inferred donor haplotypes CTVT_1B2b1 (clade 1 tumour haplotype) and A1d1a_1 (host haplotype) (Figure 5-13). Analysis of regions common to all haplotypes in both samples highlighted a ~1400bp region spanning the mtDNA control region (Figure 5-13).

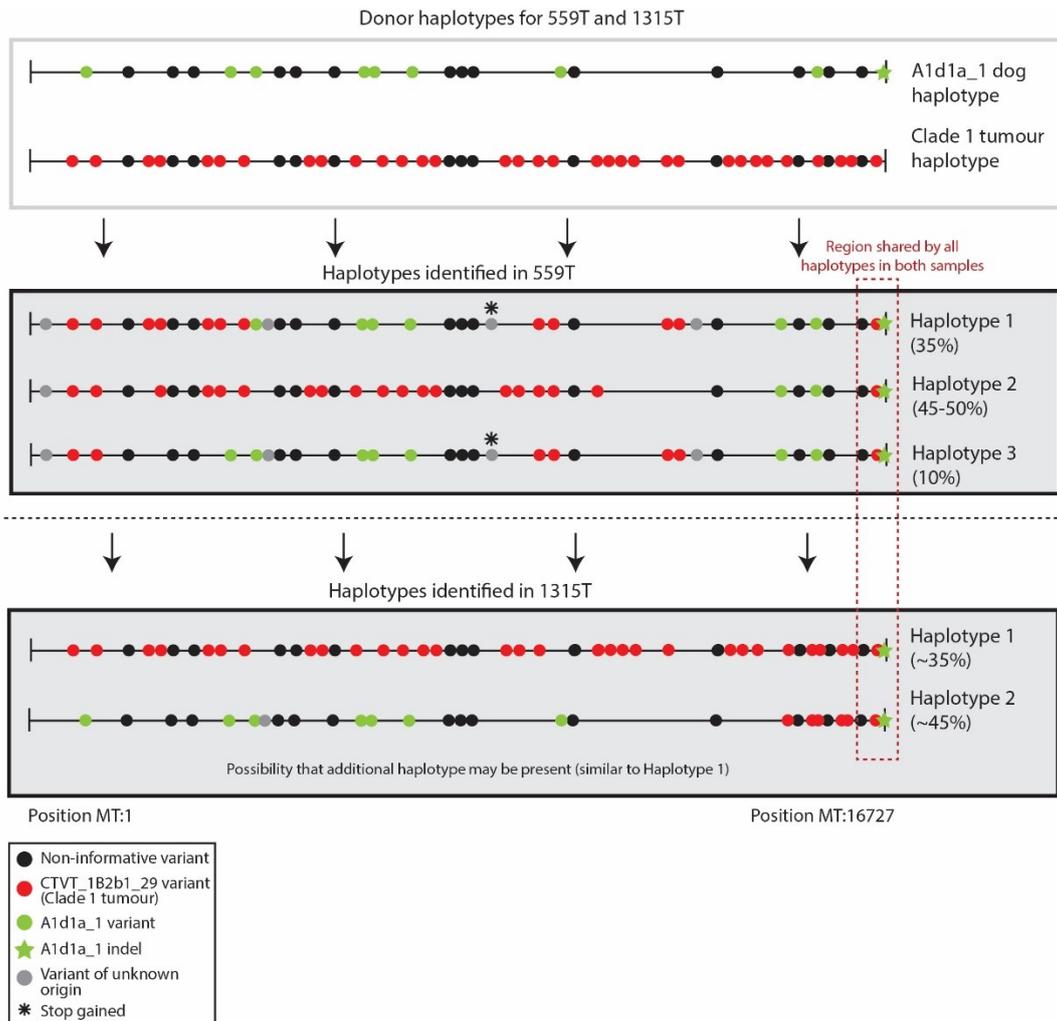


Figure 5-13 MtDNA haplotypes detected using long-read sequencing in 559T and 1315T. The region common to all haplotypes in both samples is indicated by a dotted box. The estimated percentage contribution of each recombined haplotype is shown.

Three genetic changes were identified within the ~1400bp region identical to both samples: 15814 C>T, 16025 T>C and indel 16660 CCCTTTTTTTTCCC > CCCCTTTTTTTTTTCC, out of which only the indel was a defining feature of A1d1a sub-haplogroup (see Table 5-7 and Figure 5-11). Further analysis of variant allele fraction confirmed the homoplasmic nature of this indel in both recombining samples (see section 5.2.9.8).

Homoplasmic presence of indel 16660 CCCTTTTTTTTCCC > CCCCTTTTTTTTTTCC in the mtDNA control region of both recombining samples confirms that this indel had risen to a higher frequency in comparison to other A1d1 defining genetic variation, and suggests this indel as a possible candidate genetic variation which would be interesting to look at further.

Relationship between tumours 559T and 1315T

Since the donor haplotypes in 1315T are indistinguishable to donor haplotypes in 559T based on the available data, this presents us with two possible scenarios (although we cannot exclude the possibility that the donor haplotypes may differ in regions where variants have been lost by recombination in 559T and 1315T). It is possible that a tumour ancestor to 559T captured haplotype A1d1a_1 mtDNA from its host, recombination was initiated between mtDNA haplotypes CTVT_1B2b1_29 and A1d1a_1, and cells containing various recombination products were passed on to hosts 559H and 1315H independently (Figure 5-14A). Alternatively, recombining mtDNA haplotypes in 559T have undergone different stages of recombination and a proportion of cells was passed on to 1315H (Figure 5-14B). It is unlikely that cells were passed in the other direction, i.e. from 1315T to 559H, as tumour sample 559T was collected in July 2013, while tumour sample 1315T (that was of a similar size as tumour 559T at the time of sample collection) was collected at a later stage, in March 2015.

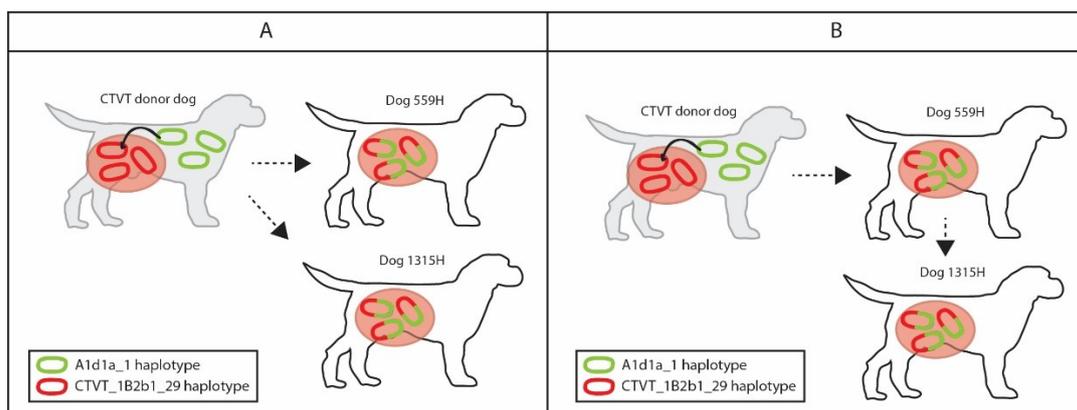


Figure 5-14 Suggested scenarios explaining relationship between CTVT tumours 559T and 1315T (although we cannot exclude the possibility that donor haplotypes for each of tumours 559T and 1315T are different and that the two events occurred independently). See text for further detail.

5.3.4 Somatic variant analysis

Genetic variation arising *after* each of the mtDNA horizontal transfer events is called somatic variation. The number of somatic mutations in each of CTVT clades 5 and 8-18, can tell us about the likely timing of the individual mtDNA horizontal transfer events. The analysis of somatic substitutions (without indels) in CTVT clades 5, 8-18 is presented below.

5.3.4.1 Timing analysis using somatic substitutions

In order to estimate the relative time since each mtDNA horizontal transfer event that gave rise to clades 5 and 8-18 took place, I estimated the number of mtDNA somatic substitutions acquired by each clade since mtDNA capture (see Table 5-8). Potential somatic substitutions represent substitutions absent from phylogenetically closest dog mtDNA haplotypes (donor haplotypes, see Table 5-4) but common to all tumours within a clade; thus the early somatic or rare germline status of these variants is unknown.

CTVT mtDNA clade	Average number of somatic substitutions per clade	Average number of potential somatic substitutions per clade	Minimum time since mtDNA uptake (years)	Maximum time since mtDNA uptake (years)
Clade 5	0	10	0	487
Clade 8	3	0	33	146
Clade 9	0.64	0	7	31
Clade 10	0	4	0	195
Clade 11	0	2	0	98
Clade 12	0	1	0	49
Clade 13	0	5	0	244
Clade 14	0	1	0	49
Clade 15	0	1	0	49
Clade 16	1.33	0	15	65
Clade 17	2	1	22	146

Table 5-8 Average number of somatic and potential somatic substitutions in clades 5, 8-18 and estimated timing of mtDNA horizontal transfer events. Somatic substitutions indicate substitutions arising after mtDNA horizontal transfer event, potential somatic substitutions represent substitutions absent from phylogenetically closest dog mtDNA haplotypes but common to all tumours within a clade; thus the early somatic or rare germline status of these variants is unknown.

Based on analysis presented in Chapter 4, section 4.2.7.3.1 and section 4.3.4, the estimated minimum (somatic substitutions only) and maximum (somatic and potential somatic substitutions) timing of the mtDNA horizontal transfer events for clades 5, 8-18 is shown in Table 5-8 (see Materials and Methods, section 5.2.10). Thus this analysis suggests, that all clades 5, 8-18 arose relatively recently in comparison to all the other CTVT clades (see Figure 5-15).

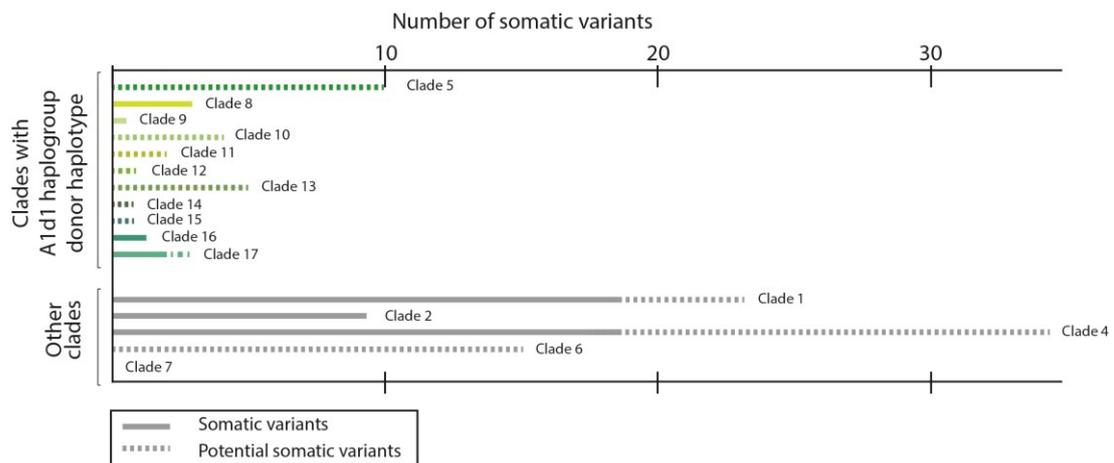


Figure 5-15 Number of somatic variants acquired by each clade after the horizontal transfer event. Continuous lines represent somatic substitutions arising after the horizontal transfer event; discontinuous lines represent potential somatic substitutions absent from normal dog donor mtDNA haplotypes but common to all tumours within a clade (see Table 5-8). Clades shown in grey represent clades 1, 2, 4, 6, 7 (n = 467 tumours).

5.3.5 Phenotypic analysis

Variation specific to the A1d1 haplogroup may not just be on the genetic level, but may also be detectable on the phenotypic level. I therefore took a histopathology approach to identify any phenotypic, and potentially functional, features specific to A1d1 haplogroup.

5.3.5.1 Histopathology analysis

Scoring of two distinct features of histopathology was performed - mitotic activity and percentage of lymphocytic infiltration. A set of tumour samples from clades 5, 8-18 was compared to matched control set from other clades (see Materials and Methods, section 5.2.11). Neither number of mitotic figures nor the percentage of lymphocyte infiltration differed significantly between clades 5, 8-18 and other clades (see Figure 5-16).

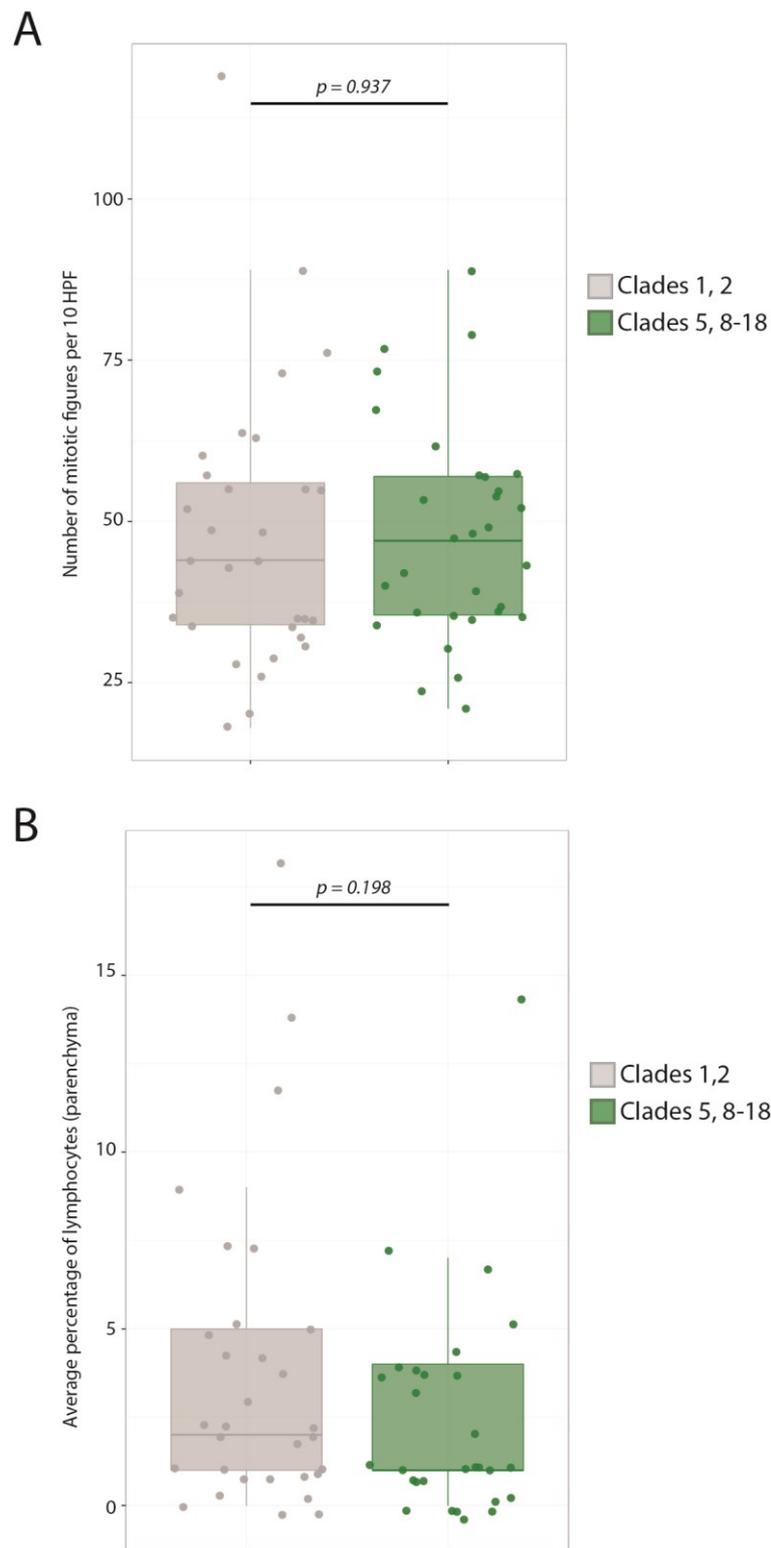


Figure 5-16 Histopathology analysis of CTVT clades 5, 8-18 in comparison with other CTVT clades. (A) Number of mitotic figures per 10 high-power fields (HPF) (B) Average percentage of lymphocyte infiltration in parenchyma. Each individual value is represented by a dot. Boxes represent the first and third quartiles (inter-quartile range = IQR). Error bars indicate values within 1.5*IQR away from first and third quartiles. P-value was calculated using two-sided student's t-test.

5.4 Discussion and future directions

5.4.1 Recurrent and recent haplotype specific mtDNA horizontal transfer

My analysis provides evidence for the occurrence of 12 independent and recent horizontal transfers of A1d1 mtDNA haplogroup in CTVT tumours. The A1d1 haplogroup is the only mtDNA haplogroup, which we have seen to be recurrently and independently transferred. Apart from these 12 mtDNA horizontal transfers, we have detected only five other mtDNA horizontal transfers of non-A1d1 mtDNA haplotypes that have been fixed, all of which are distinct from each other. These included clades 1, 2 and 4, previously identified in Chapter 4, together with two new clades found in the additional set of samples not previously analysed in Chapter 4, named clade 6 (Armenia) and clade 7 (Mexico). These findings highlight the unusual frequency of the recurrent horizontal transfer of the A1d1 haplogroup, in particular as all of the A1d1 mtDNA haplogroup horizontal transfers have occurred very recently.

Analysis of features of the A1d1 haplogroup has shown that neither the worldwide distribution within the normal population, nor the mtDNA copy number of this haplogroup in host/tumour samples differ from other mtDNA haplogroups in the canine population. It should, however, be noted that the worldwide distribution of A1d1 haplogroup was analysed only within the CTVT host population and we cannot rule out that this haplotype may be highly overrepresented in this population. Additionally, evidence from high coverage whole genome sequencing data indicates that long-term low-level heteroplasmic maintenance of the A1d1 haplotype in CTVT cells is very unlikely. This suggests that neither of the above would explain the increased number of detected horizontally transferred mtDNA A1d1 haplotypes. Therefore, a feature specific to the A1d1 haplogroup is more likely to be responsible for the high frequency of horizontal transfer observed to involve this haplotype.

Genetically, a number of candidate variants specific to the A1d1 haplogroup has been identified through my analysis, as follows:

- Indel at position 16660 CCCTTTTTTTTCCC > CCCCTTTTTTTTTCC
- Non-synonymous variant at position 7593 T>C
- Non-coding variants at position 1351 A>G and 15620 T>C

In addition, a number of features in the A1d1 haplogroup were not analysed in this Chapter and should therefore still be considered as candidate variations:

- Point variation within the 10bp repeat unit (GTACACGT(**G/A**)C) in the canine mtDNA control region (as described before by Gundry et al. 2007 (Gundry et al. 2007))
- Presence of variable-length polyC homopolymer tract in the canine mtDNA control region (Fregel, Suarez et al. 2015)

Given the data presented in this Chapter, the insertion at position 16660 CCCTTTTTTTTCCC > CCCCTTTTTTTTCC appears to be the most promising candidate for conferring functional or competitive advantage to the A1d1 haplogroup. Importantly, however, this insertion is specific to the A1d1a/A1d1a1 sub-haplogroups (i.e. present in all tumours from clades 8-18 and in all hosts with the A1d1a/A1d1a1 mtDNA haplotypes), as detected through my analysis (see section 5.3.3.5.2) as well as reported previously (Fregel et al. 2015). Moreover, it is the only identified A1d1a/A1d1a1 sub-haplogroups-specific variant found to be present homoplasmically in all CTVT mtDNA haplotypes within samples undergoing mtDNA recombination - 559T and 1315T (see section 5.3.3.5.4). This finding highlights the indel as a possible genetic variation conferring advantage to the A1d1 haplogroup and thus may contribute to our understanding of why the A1d1 haplogroup has undergone recurrent mtDNA horizontal transfer.

Although the insertion at position CCCTTTTTTTTCCC > CCCCTTTTTTTTCC appears as a good candidate, the fact that it is not present in clade 5 (which is the only haplotype with A1d1-only haplotype) may disregard its relevance. On the other hand, it is possible that clade 5 may not be part of the same recurrent horizontal transfer phenomenon. This possibility would be supported by the fact that clade 5 is the only CTVT clade with the A1d1 donor haplotype and can be argued to be more genetically distinct on a phylogenetic tree to clades 8-18, which arose from A1d1a/A1d1a1 sub-haplogroups (see Figure 5-17). It is therefore possible, that the A1d1 haplotype mtDNA capture which gave rise to clade 5 happened by chance as a result of the known occasional mtDNA horizontal transfer process occurring in CTVT (Rebeck et al. 2011, Strakova et al. 2016). This hypothesis would be supported by the fact that we identified two additional mtDNA clades, clade 6 and clade 7, that were also only found within a single country. This may suggest that mtDNA transfer in CTVT indeed is more common than previously expected and that the A1d1 mtDNA haplotype capture giving rise to clade 5 may have happened just by chance, independently of the recurrent A1d1a/A1d1a1 horizontal transfer phenomenon. Neither of the possibilities above, however, can be ruled out.

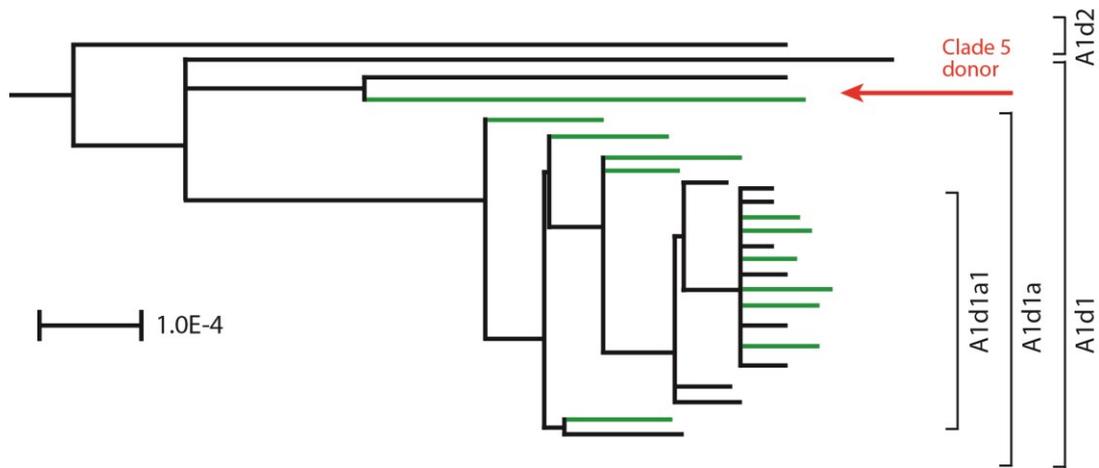


Figure 5-17 Phylogenetic structure of A1d1 haplogroup. Locations of donor haplotypes for clades 5, 8-18 are shown in green. Clade 5 donor mtDNA haplotype is highlighted by red arrow.

The functional consequences of the insertion at position 16660 CCCTTTTTTTTCCC > CCCCTTTTTTTTTTCC are not clear. The insertion appears within the canine mtDNA control region which is not conserved in humans, making the identification of its exact position in respect to the functional elements in the control region difficult. Nevertheless, it appears to be located between the conserved sequence block III (CBS-III) and tRNA^{Phe}. It is unclear however, whether the insertion may be present within either the heavy strand promoter (HSP), the light strand promoter (LSP) or between them, as their exact position has not been defined in the canine genome (see Figure 5-18). It has however been previously reported that, in human mtDNA, variants associated with the mtDNA control region may confer a replicative advantage to specific mtDNA haplotypes (Liou et al. 2010, Kang et al. 2016). Therefore, there is potential for insertion at position 16660 CCCTTTTTTTTCCC > CCCCTTTTTTTTTTCC in the canine mtDNA control region to have a functional effect.

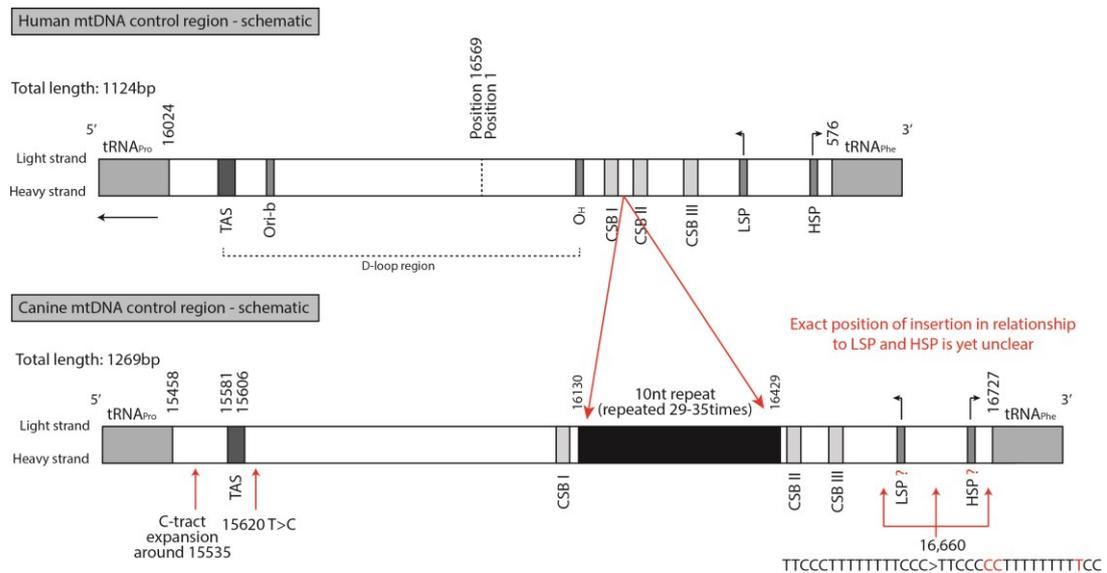


Figure 5-18 Schematic comparison between the human and canine mtDNA control region. Insertion at position 16660 CCCTTTTTTTTCCC > CCCCTTTTTTTTTTCC is indicated on the diagram, but its exact position in respect to functional elements in the mtDNA control region is unclear.

Other genetic variants specific to the A1d1 haplogroup were also detected through this analysis and should be considered as potential candidates for conferring functional or competitive advantage to the A1d1 haplogroup. The only missense variant, identified at position 7593 T>C, was found at a position conserved between dogs and humans - since this variant was previously detected to appear as a single nucleotide polymorphism in the normal human population (Marzuki et al. 1991, Lott et al. 2013), it would be very unlikely to have any functional effects in dogs (see section 5.3.3.5.1). Two non-coding variants were identified at positions 1351 A>G and 15620 T>C - similarly to above, both were found at positions conserved between dogs and humans, with single nucleotide polymorphisms detected in the normal human population (see section 5.3.3.5.1) (Lott et al. 2013). Five of these variants were synonymous, thus not leading to an amino acid sequence change. Even though we cannot rule out the relevance of the genetic variants discussed above, it is important to note that none of the substitution variants had risen to higher frequency within the two recombining samples (see section 5.3.3.5.4), therefore making it less likely for these variants to confer advantage to the A1d1 haplogroup.

5.4.2 Hypotheses and possible link to Vincristine treatment

As discussed in section 5.3.4.1, all of the independent mtDNA horizontal transfers of A1d1 haplogroup have occurred relatively recently. Considering the changes taking place in CTVT

management within the last few decades, perhaps the most striking change is the introduction of use of chemotherapy treatment. More specifically, vincristine has gradually become the treatment of choice for CTVT, but other chemotherapy agents such as doxorubicine or vinblastine are sometimes also used (see Chapter 2, Figure 2-11) (Brown et al. 1980, Calvert et al. 1982, Singh et al. 1996, Nak et al. 2005, Strakova and Murchison 2014, Ramadinha et al. 2016). Given this significant change in the management of CTVT, it is possible that chemotherapy treatment could be involved in influencing the high frequency of horizontal transfer observed to involve the A1d1 haplogroup.

With the possibility that chemotherapy or other external agents might be involved, I formulated four different hypotheses that might explain the recurrent, A1d1 haplogroup-specific horizontal transfer phenomenon that we have observed. The four hypotheses are outlined below as well as summarised in Figure 5-19:

- (1)** Exposure to chemotherapy or other factors increases the mtDNA horizontal transfer of all mtDNA haplotypes. After the mtDNA transfer, A1d1 haplogroup has a selective advantage over the CTVT-endogenous haplotypes.
- (2)** Exposure to chemotherapy or other factors increases the likelihood of mtDNA horizontal transfer specifically for the A1d1 mtDNA haplogroup.
- (3)** All haplotypes undergo mtDNA horizontal transfer equally frequently, and chemotherapy or other factors do not influence frequency of mtDNA horizontal transfer. After the mtDNA transfer, A1d1 mtDNA haplogroup has a selective advantage specifically in the presence of chemotherapy or other factors.
- (4)** CTVT cells carrying the A1d1 haplogroup mtDNA do not persist - A1d1 lineages have always arisen with higher frequency, but they do not contribute to CTVT expansion over the long-term. A1d1 haplotype might be conferring a short-term advantage for CTVT cells or may be particularly prone to horizontal transfer.

Considering the four hypotheses outlined above, I will review them in light of the information already published in the literature and discuss to what extent my data support the formulated hypotheses.

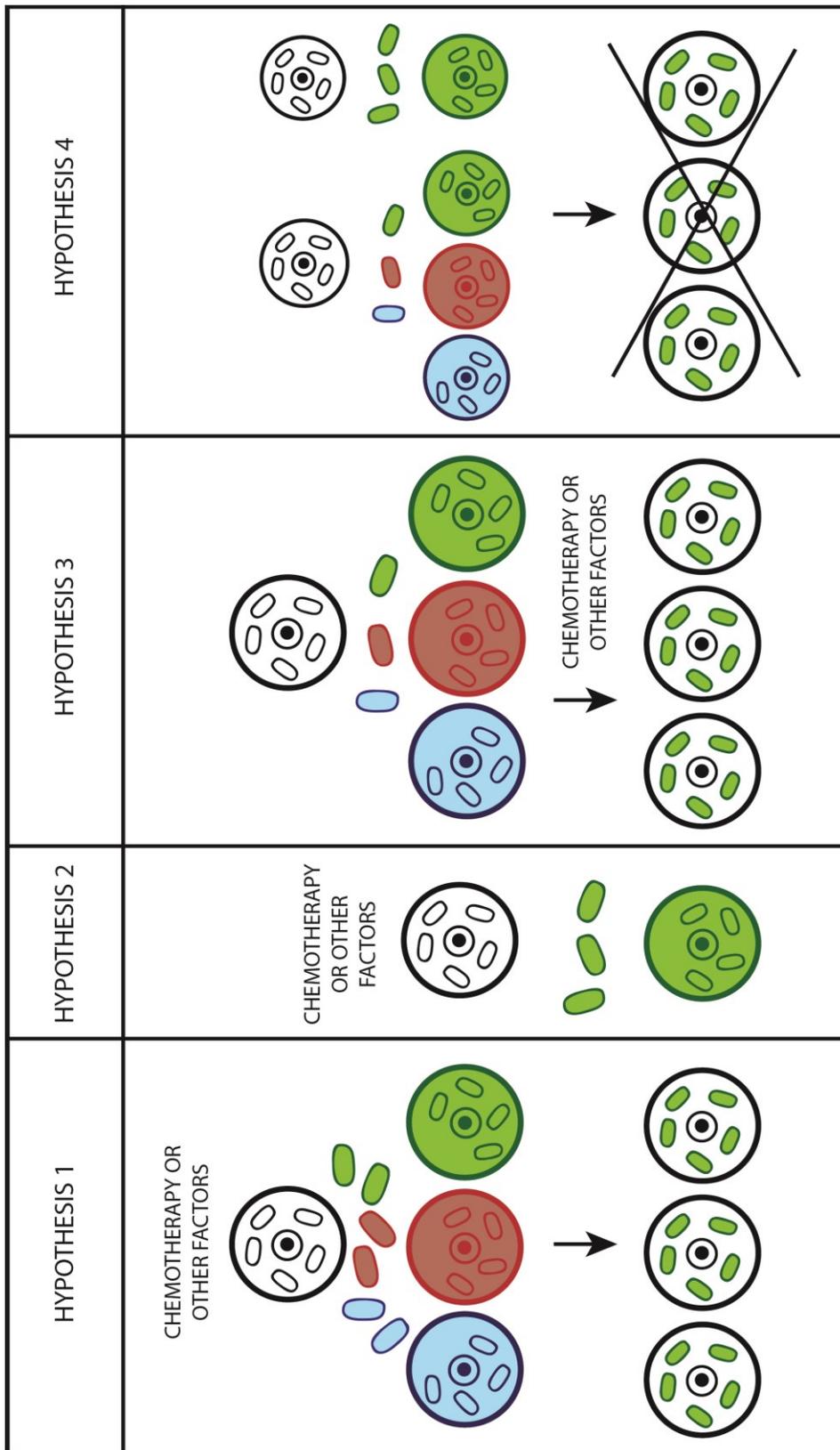


Figure 5-19 Diagrams representing the different hypotheses which may explain the recurrent, recent and A1d1 haplogroup-specific mtDNA horizontal transfer. Green cells and mitochondria represent CTVT mitochondrial clades with A1d1 haplogroup donor mtDNA (CTVT clade 5, 8-18). Red and blue cells represent other CTVT mitochondrial clades. See text for further details.

As discussed above, my analysis of the worldwide distribution of A1d1 haplogroup and the host/tumour copy number of A1d1 haplogroup have shown no differences between haplogroups. Similarly, my data have suggested that long-term heteroplasmic maintenance of the A1d1 haplogroup in CTVT is very unlikely. I will therefore focus on the genetic and phenotypic features of A1d1 haplogroup.

Taking into account the data presented in this Chapter, the insertion at position 16660 CCCTTTTTTTCCC > CCCCTTTTTTTTCC appears to be the most promising candidate for conferring functional or competitive advantage to the A1d1 haplogroup (even though we cannot rule significance of other genetic variants out). As eluded to, variants associated with the regulatory part of the mtDNA control region may confer a replicative advantage to specific mtDNA haplotypes (Kang et al. 2016). One possible explanation may therefore be that A1d1 haplogroup could outcompete the CTVT-endogenous haplotypes through a replicative, rather than a functional, advantage. The A1d1 haplogroup would only have an advantage in direct competition with other CTVT-haplotypes, thus supporting hypothesis (1).

To further support hypothesis (1), we detected evidence for possible appearance of new, recently arising CTVT clades, which would be expected with a general increase in mtDNA horizontal transfer frequency. As presented in Chapter 4 and in section 5.3.1 of this Chapter, we have identified five additional CTVT clades 1, 2, 4, 6 and 7 (apart from clades with A1d1 mtDNA haplogroup), arising from distinct non-A1d1 mtDNA haplotypes. One of these clades (clade 7, see Figure 5-15) has a very low number of somatic mutations, thus indicating that it arose very recently. Moreover, we only found a single CTVT tumour which belongs to this clade. This finding would fit with our prediction in hypothesis (1), where an increase in mtDNA horizontal transfer of all mtDNA haplotypes would lead to random appearance of other new CTVT clades.

The phenotypic histopathology analysis did not show any differences between A1d1 haplogroup and other haplotypes (see section 5.3.5), however, it should be highlighted that all samples were collected before vincristine treatment. This suggests that there is no difference in histopathology appearance of tumours before vincristine treatment, which may make hypothesis (2) less likely. However, further studies looking at the morphology of mitochondria in different haplogroups using techniques such as transmission electron microscopy (preliminary data discussed in section 5.4.3.2) may provide further information as to whether a morphological difference might increase the likelihood of mtDNA horizontal transfer for the A1d1 mtDNA haplogroup specifically.

During the analysis of a set of 449 CTVT tumours (Chapter 4), I did not detect low levels of non-host mtDNA in CTVT tumours, which would have been expected if all haplotypes were to undergo mtDNA horizontal transfer equally frequently, as suggested in hypothesis (3). This finding indicates that hypothesis (3) is not supported by my current data. Further studies into linking A1d1 haplogroup CTVT clades and chemo-resistance (preliminary data discussed further in section 5.4.3) will provide additional information as to whether A1d1 haplogroup may be conferring a selective advantage in the presence of vincristine.

Further considering hypotheses (1), (2), and (3), both chemotherapy as well as other factors may be involved. However, in light of the published literature, and noting that all CTVT clades with A1d1 donor haplogroup arose relatively recently, chemotherapy is a strong candidate factor. Notably, previous studies have reported that chemotherapy enhances the probability of mtDNA horizontal transfer, which confers advantage to neoplastic cells (Moschoi et al. 2016). Increasing chemotherapy use in CTVT may therefore be leading to increased numbers of mtDNA horizontal transfer events, as implicated by my hypothesis (1).

Should the theoretically increased frequency of mtDNA horizontal transfer in CTVT be chemotherapy-mediated (hypothesis (1)), a crucial enabling factor would be occurrence of CTVT cases with the ability to transmit the tumour during chemotherapy treatment (i.e. tumour would have to be transmitted by a dog that had commenced chemotherapy treatment, but has not yet completely resolved). This means that CTVT affected dogs in treatment would have to be either in direct contact with other dogs during treatment (thus providing an opportunity for CTVT transmission) or dogs may be released back to streets/not brought back by their owners before the treatment was fully completed. Preliminary studies focusing on this question are further discussed in section 5.4.3.3.

Although we cannot exclude any of the hypotheses above at this stage, overall, based on the available literature and the current data, the most feasible hypothesis appears to be hypothesis (1): chemotherapy (or other factor)-linked enhancement of mtDNA horizontal transfer frequency, followed by a competitive advantage of the A1d1 haplogroup, possibly linked to insertion at position 16660 CCCTTTTTTTTCCC > CCCCTTTTTTTTCC which may influence mtDNA transcription or replication.

5.4.3 Exploratory analyses and future directions

Even though hypothesis (1) may appear as the most feasible hypothesis with my current data, we cannot exclude any of these four hypotheses discussed above. In order to confirm which

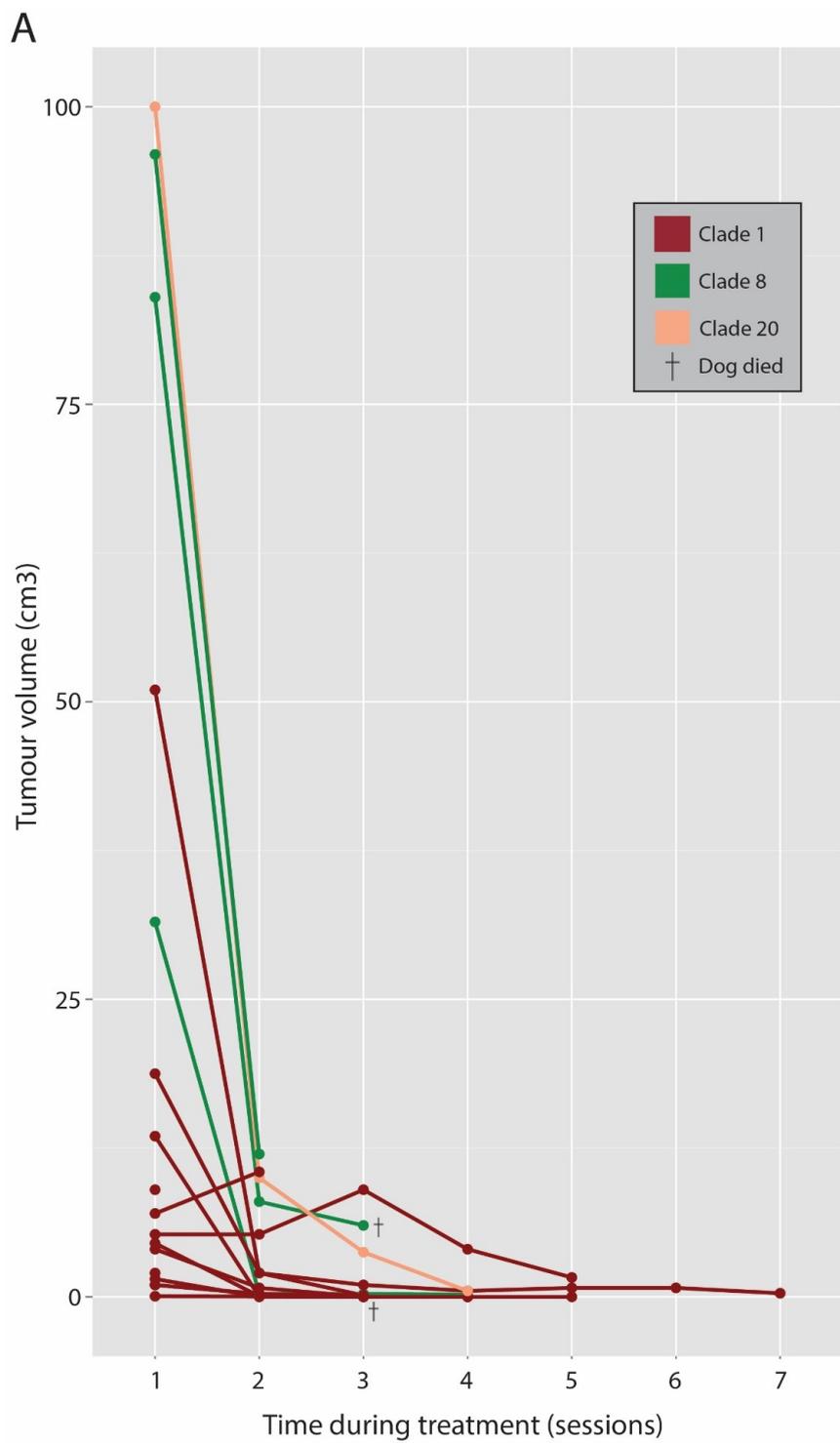
hypothesis (if any) might best explain the increased frequency of A1d1 haplogroup mtDNA horizontal transfer, further analyses will need to be performed. In this section, I discuss three pilot studies which have recently been in progress, and other possible future directions. Overall, further studies will be of a great interest, as understanding the mechanism for recurrent A1d1 haplogroup-specific mtDNA horizontal transfer in CTVT may also teach us about important roles of mitochondria in other canine and human cancers. Importantly, should this phenomenon be linked to chemotherapy, we may learn valuable information about mitochondrial chemotherapy resistance mechanisms. Finally, understanding the mechanism of mtDNA haplotype specific advantage and expansion would be of interest in three-parent embryo biology, where diseased mtDNA haplotype is known to be detected at a low level, but future development is unclear (Zhang et al. 2017).

The three types of exploratory analyses, which I have already performed are: clinical study of vincristine treatment response in different CTVT clades, transmission electron microscopy analysis and questionnaire-based approach to understand CTVT management during treatment, as discussed below.

5.4.3.1 *Clinical study of vincristine treatment response*

Clinical studies following the response to chemotherapy treatment in A1d1 clades in comparison to other clades will help us to identify any potential links between increased frequency of A1d1 haplogroup mtDNA horizontal transfer and vincristine treatment response. During my PhD, Alexander Sampson (Part II pathology veterinary student in the Transmissible Cancer Group) and I were able to perform a preliminary clinical study in Belize, during which we followed eighteen clinical CTVT cases for the duration of their vincristine treatment. We collected fourteen samples which belonged to CTVT clade 1, three samples which belonged to CTVT mitochondrial clade 8 (clade A1d1 donor haplogroup, see Table 5-4) and one sample which belonged to another new distinct clade, named clade 20 (note that the new clade 20 is not further discussed in this Thesis). Tumour volume was recorded during the progression of vincristine treatment. Reduction in tumour size during vincristine treatment was seen for all 18 cases (Figure 5-20A and Figure 5-20B).

The results, however, were inconclusive due to the limited number of cases followed-up as well as due to complications in the field conditions (death of dogs due to car accidents or other unexplained reasons, inaccessibility of places due to a hurricane), which meant that we were not able to fully follow-up a significant number of the clinical cases. Further follow-up work will be performed in the future.



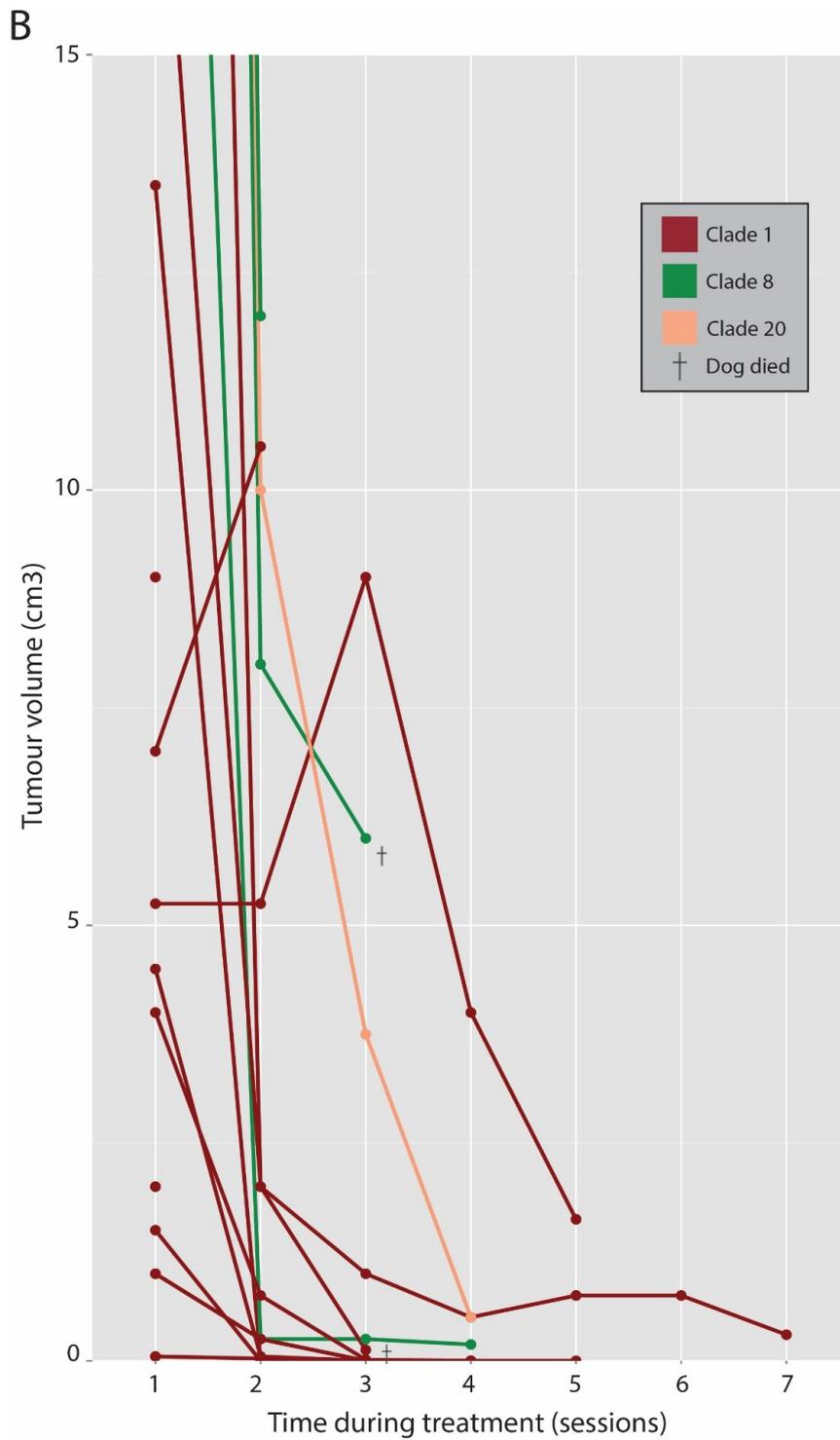


Figure 5-20 Clinical study of vincristine treatment response in Belize. CTVT tumour volume during treatment progression. (A) Full sized graph (B) zoom-in of the graph with the maximum on y axis = 15cm³.

5.4.3.2 *Transmission electron microscopy analysis of A1d1 mitochondria*

Transmission electron microscopy (TEM) is a technique which achieves high resolution and magnification of a sample and has the potential to uncover information about the structural differences between mitochondria. By comparing the TEM data from clades 5, 8-18 and other clades, there is potential to learn about variation in number, density and distribution of mitochondria or structural differences and shape abnormalities. By understanding the shape and distribution of A1d1 mitochondria with a cell, we may be able to decide whether the likelihood of mtDNA transfer in A1d1 haplogroup specifically may be increased, thus considering my hypothesis (2). Moreover, to my knowledge, this would be the first time that a naturally occurring CTVT tumour has been analysed using transmission electron microscopy.

As part of my PhD work, samples for transmission electron microscopy analysis were collected during a field-trip to Belize by myself and Alexander Sampson (Part II pathology veterinary student in the Transmissible Cancer Group) using a reusable steel biopsy punch (size 1mm) directly into fixative (2% glutaraldehyde, 2% formaldehyde, 0.05M PIPES buffer). The sample was kept in the fixative at 4°C for 12-24 hours. After fixation, the sample was washed 3-5 times in saline (for 5 minutes each time). The samples were processed by the team at the Cambridge Advanced Imaging Centre.

To date I have performed the qualitative analysis of the TEM images to confirm the high-quality of fixation, with further quantitative analysis to be performed in the future (see Figure 5-21).

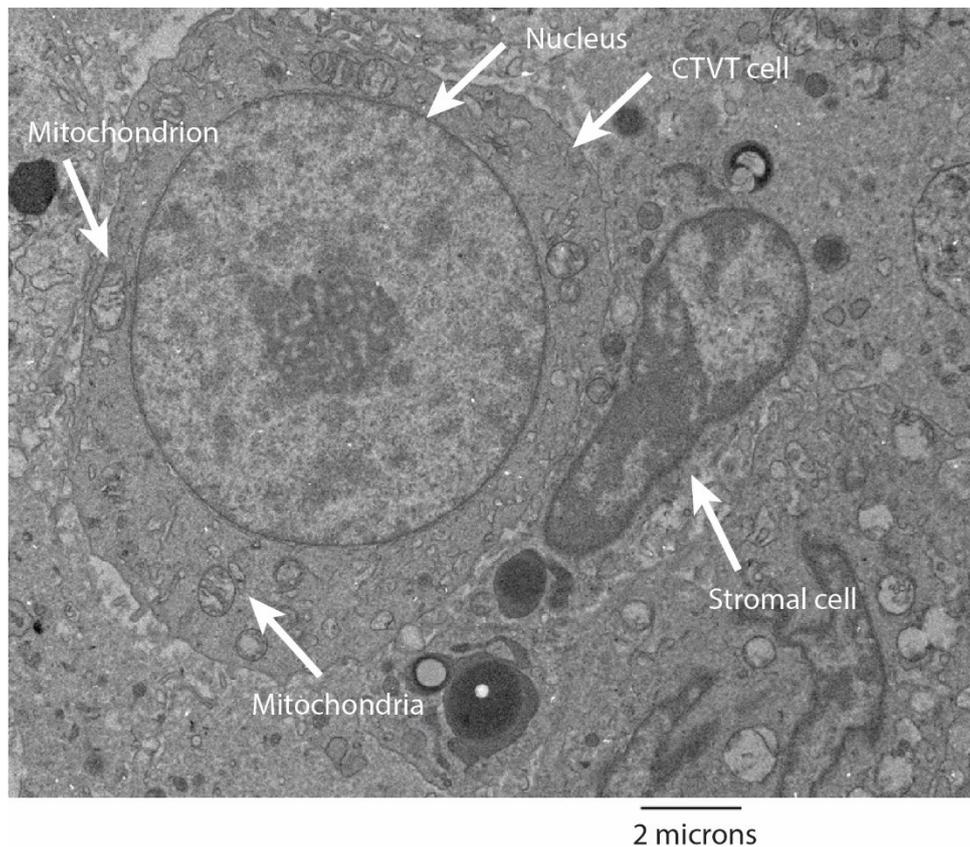


Figure 5-21 Transmission electron microscope image of a CTVT tumour sample (magnification 1700x). The stromal cell and CTVT cell + nucleus + mitochondria are labelled using arrows. Scale bar is shown below the image.

5.4.3.3 CTVT management during treatment

As discussed previously, if any of the hypotheses discussed above would be chemotherapy-mediated, the tumour would have to be transmitted by a dog that had commenced chemotherapy treatment, but has not yet completely resolved. In order to understand the management of CTVT during treatment in different veterinary practices around the world, I designed and distributed an internet-based questionnaire, with 10 predominantly multiple-choice questions about CTVT treatment strategies, to more than 70 veterinary collaborators around the world (<https://www.surveymonkey.co.uk/r/G3ZJXJW>). I received a total of 63 responses from 39 countries around the world.

Moreover, this questionnaire may help us to understand any possible links between A1d1 haplogroup and vincristine resistance - should A1d1 haplogroup be more resistant to vincristine, we would expect a higher number of vincristine doses given in locations with clades 5, 8-18. The questionnaire-based approach may therefore provide information about hypotheses (1) - (3) above.

The results from this questionnaire were currently inconclusive, due to the small number of responses received so far. Further follow-up work may be carried out in the future.

5.4.3.4 *Future directions*

During my analysis described in this Chapter, I have been able to make the following conclusions:

- I have reported that the A1d1 mtDNA haplogroup has undergone 12 independent and recent horizontal transfer events, giving rise to 12 independent CTVT mitochondrial clades.
- I have shown that the A1d1 worldwide distribution and the A1d1 haplogroup host/tumour mtDNA copy number did not differ from other mtDNA haplogroups in the canine population. My analysis has indicated that long-term low-level heteroplasmic maintenance of the A1d1 haplotype is very unlikely.
- My genetic analysis has highlighted a number of candidate variants responsible for the increased frequency of A1d1 mtDNA horizontal transfer, and I have speculated that the indel at position position 16660 CCCTTTTTTTTCCC > CCCCTTTTTTTTTCC appears to be the most promising candidate, which may confer an advantage to the A1d1 haplogroup through influencing mtDNA transcription or replication.
- My phenotypic analysis has indicated that the number of mitotic figures, nor the infiltration of lymphocytes, differ between A1d1 haplogroup and other haplogroups.

Apart from my preliminary pilot work described above, there is a number of other studies that could be performed to further understand the phenomenon of A1d1 haplogroup specific mtDNA horizontal transfer. Other future possible studies are discussed below and summarised in Figure 5-22.

One of the most informative future analyses might be analysis of RNAseq expression data from CTVT tumours with A1d1a/A1d1a1 haplotype in comparison to other CTVT tumours. I have already sequenced a pilot set of 9 CTVT tumours – 3 of which are of the A1d1a/A1d1a1 haplotype, and the remaining 6 tumours belong to mitochondrial clades 1 and 2. Stranded RNAseq library with ribozero rRNA depletion was made and then sequenced using Illumina HiSeq 4000 (paired-end reads, 75bp). Further analysis of this dataset might inform us about variable RNA expression between A1d1 haplogroup CTVT tumours and other CTVT tumours,

and may also be useful in determining whether the insertion at position 16660 CCCTTTTTTTTCCC > CCCCTTTTTTTTTTCC might have an effect on the heavy or light strand promoters (HSP or LSP).

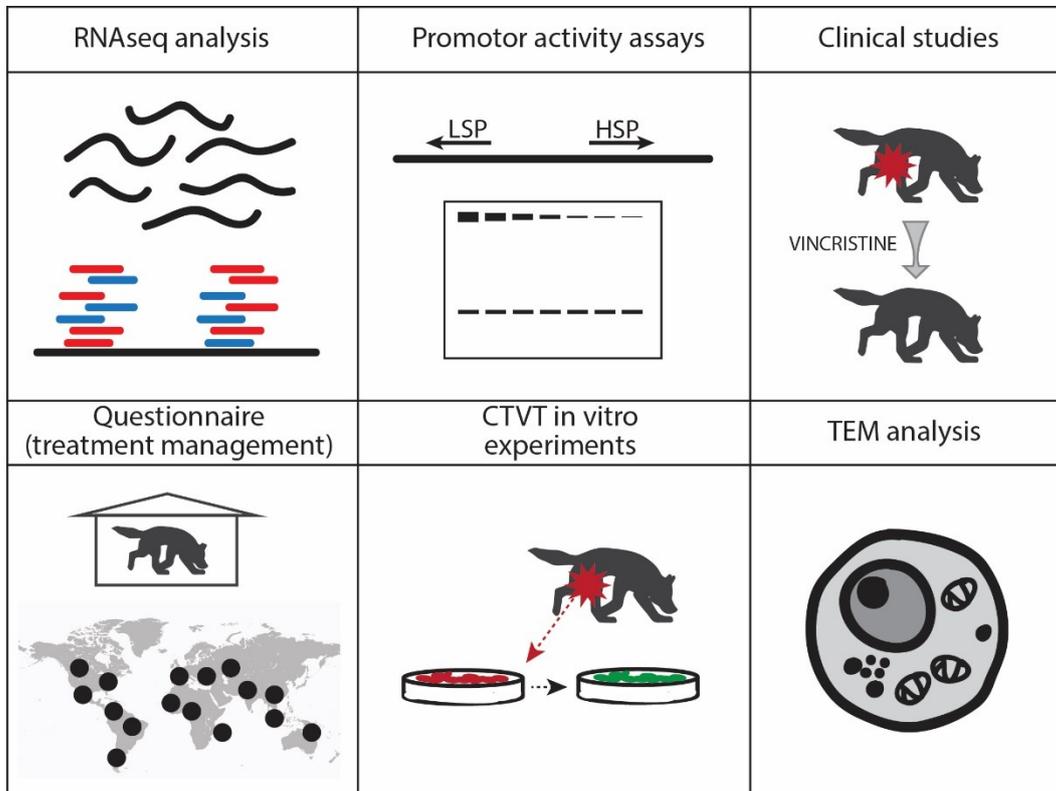


Figure 5-22 Summary of future experiments which could be performed to understand the phenomenon of A1d1 haplogroup specific recurrent and recent mtDNA horizontal transfer event.

Functional effects of insertion at position 16660 could also be in the future analysed through a promoter activity assay, as reported before (Connor et al. 2017), to understand whether the competitive advantage that the A1d1 haplogroup would appear to have, assuming hypothesis (1), may be linked to mtDNA heavy or light strand promoter activity and transcription.

Finally, there are many more possibilities to analyse the A1d1 haplogroup specific recurrent horizontal transfer phenomenon and the possible link to vincristine treatment with CTVT cell line experiments. It would be very interesting to see whether vincristine would indeed lead to enhancement of mtDNA horizontal transfer in CTVT cell lines followed by a competitive advantage of the A1d1 haplotype (thus confirming hypothesis (1)) – these questions could

be answered for example through fluorescent labelling experiments. There is, however, no CTVT cell line being passaged in vitro at this moment and therefore these experiments will be dependent on cultivation of CTVT cell lines.

Overall, further studies to understand the specific features of A1d1 haplogroup which lead to recurrent mtDNA transfer would be worth pursuing further. Through learning about selective advantages of particular mtDNA haplotypes, we may gain further understanding about the role of mitochondria in cancer, but also about feasibility of mitochondrial replacement therapies and role of mitochondria in three-parent embryos. Another important area of interest may be mtDNA-linked chemotherapy resistance. Studying mtDNA horizontal transfer in CTVT gives us a unique opportunity to identify each mitochondrial capture through phylogenetic analysis (which may not be possible in other systems when the mtDNA transfer is genetically undetectable). CTVT may therefore be a useful model for further studies of mitochondrial function in cancer.

6

Final discussion and future studies

Summary

This Chapter summarises the scientific approaches and results presented throughout the Thesis. A summary of the Thesis describes the overall goals of my work, followed by an explanation of how I tackled each of them. CTVT is a unique biological entity and the longest living cancer known in nature. Studying the genetics and biology of CTVT may therefore enable us to study processes that may also occur in normal human and non-transmissible cancers, but may just be very difficult to detect. Accumulation of genetic changes that have been taking place in CTVT over thousands of years may make some processes more prominent and detectable in scientific studies. The direct implications of my studies are relevant to both clinical veterinarians, but also have applications in cancer biology research, understanding the role of mitochondria in cancer and finally my results are relevant for the field of canine genetics. The work presented in this Thesis is always limited by the number and diversity of CTVT samples available, however, at the same time, this provides scope to continue to expand this analysis with new CTVT samples in the future. Other future projects may focus on understanding the functional relevance of haplotype-specific mtDNA transfer and on expanding the phenotypic screen of CTVT tumours and linking the phenotypic changes to the underlying genetic changes. Overall, the conclusions from my work bridge the different branches of CTVT research to further our understanding of CTVT evolution and genetic diversity around the world.

6.1 Thesis summary

Canine transmissible venereal tumour (CTVT) is a contagious cancer that is naturally transmitted between dogs by the allogeneic transfer of living cancer cells during coitus. The disease first arose in a single dog living several thousand years ago and since then has spread as a clonal lineage through dog populations worldwide. The majority of previous studies on CTVT have focused on clinical features, treatment options and case studies of this disease, with only a small number investigating the genetics of CTVT (Murgia et al. 2006, Rebbeck et al. 2009, Rebbeck et al. 2011). However, at the time I started my PhD, the first two CTVT genomes were sequenced, which, apart from further confirming the clonal nature of the disease, indicated that CTVT was confined to a single place for most of its history, until it colonised the worldwide dog population relatively recently in its evolution (Rebbeck et al. 2009, Murchison et al. 2014). These studies also alluded to the genetic variation present between tumours from different parts of the world. To understand how this extraordinary cell lineage evolved and to understand its genetic diversity, the overall goals of this Thesis were:

- (1)** *To gain further understanding of CTVT prevalence around the world and to decipher historical patterns of global CTVT spread*
- (2)** *To determine the frequency of mitochondrial horizontal transfer and search for evidence of mitochondrial genome recombination*
- (3)** *To understand the effects of negative selection acting on mitochondrial genomes in CTVT*
- (4)** *To map the phenotypic diversity of CTVT tumours around the world, and explain phenotypic traits by underlying genetic changes*
- (5)** *To understand phylogenetic relationships between individual CTVT mtDNA clades*

The work presented in Chapters 2-5 describes the approaches I used to tackle the goals presented above and a short summary is outlined below.

To understand the distribution patterns of CTVT, I obtained information from 645 veterinarians and animal health workers in 109 countries, and generated a snapshot of the locations in which this disease is found (Chapter 2). Additionally, as preparation for further genetic analysis, I collected samples from over a thousand CTVT cases from more than 50 countries, optimised methods for high-throughput DNA extraction and quantification and

optimised a qPCR-based assay for CTVT diagnosis and host contamination detection (Chapter 3).

With the goal of tracing the historical spread of CTVT and learning about the genetic diversity of this disease, I sequenced complete mitochondrial genomes of 449 CTVT tumours and their matched hosts (Chapter 4). The analysis of the CTVT mitochondrial diversity revealed that CTVT has captured mitochondrial DNA (mtDNA) through horizontal transfer events at least five times during the history of the lineage, delineating five tumour clades. CTVT appears to have spread rapidly around the world within the last 2,000 years, perhaps transported by dogs travelling along historic maritime trade routes (Chapter 4). This work indicated that negative selection has operated to prevent accumulation of deleterious mutations in captured mtDNA, and that recombination has caused occasional mtDNA re-assortment (Chapter 4). A histology-based screen of CTVT clades did not show any significant phenotypic differences between groups (Chapter 4).

In order to determine how the five mtDNA clades relate to each other, I analysed data from 539 CTVT exomes (Chapter 5). This work revealed that a single canine mtDNA haplogroup has recurrently and recently undergone multiple horizontal transfer events (Chapter 5). Analysis of this haplotype highlighted a number of candidate genetic variants which may be conferring a selective advantage to this haplotype in CTVT, possibly by influencing mitochondrial transcription or replication (Chapter 5).

Overall, I have combined and linked information obtained from a number of different approaches to understand the genetic and phenotypic diversity, evolution and spread of CTVT around the world (see Figure 6-1). The different approaches utilised during my PhD work constitute the following:

- (1)** Mapping and studying the worldwide distribution of CTVT
- (2)** Collecting and validating CTVT samples from around the world
- (3)** Utilising a number of DNA sequencing strategies to obtain genetic data (low coverage whole genome sequencing, exome sequencing, PacBio long-read sequencing, deep coverage whole genome sequencing and capillary sequencing)
- (4)** Tracing historical spread of CTVT using phylogenetics

- (5) Analysing CTVT mitochondrial genetic diversity around the world and determining frequency of mtDNA horizontal transfer
- (6) Performing a large-scale histopathology screen
- (7) Using CTVT exome sequencing data to understand relationships between CTVT mitochondrial lineages

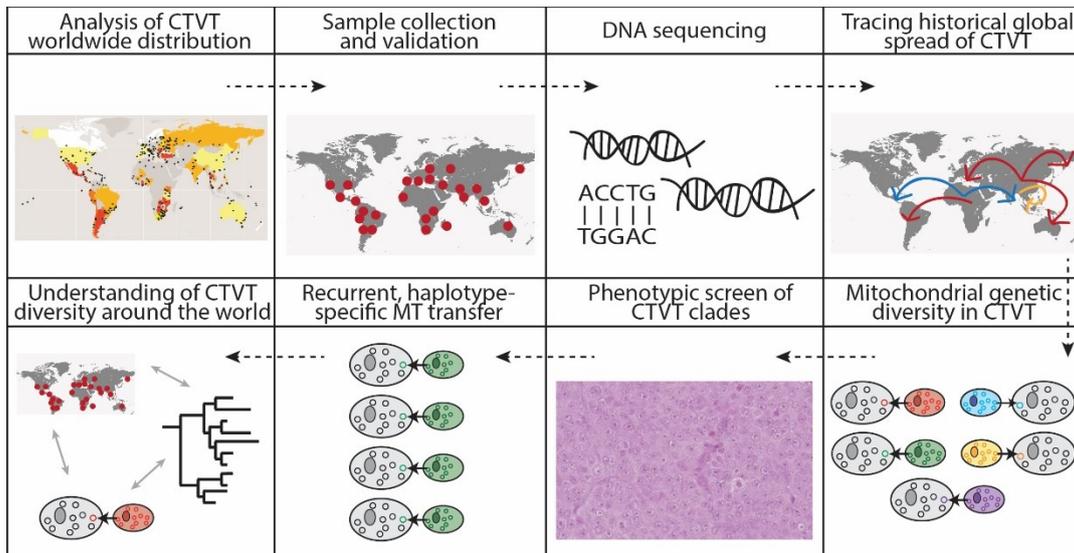


Figure 6-1 Summary of approaches used in my PhD thesis to understand evolution and genetic diversity of CTVT around the world.

6.2 Implications of this research

6.2.1 Study of a transmissible cancer

In this section, I explain why I believe that through understanding the biology of CTVT, a unique transmissible cancer, we may learn about essential cellular processes occurring also in other cancers.

- (1)** CTVT, as discussed above, is a long-lived cancer, which is a few thousand years old (Murgia et al. 2006, Rebbeck et al. 2009, Murchison et al. 2014). In contrast, cancers that are not transmissible, and die with their host, are only a few decades old. This implies that CTVT has had substantially more time to evolve and cancer evolutionary processes may be more prominent and so more detectable than in other cancers. For example, in Chapter 4, I presented evidence for negative selection acting on deleterious mutations in CTVT mtDNA – this may be a process that is more widespread in human and other cancers, but the signal may be more difficult to detect (Ju et al. 2014, Yuan et al. 2017).
- (2)** As discussed in Chapter 4 and Chapter 5, CTVT has been shown to capture its mtDNA a number of times in its history through mtDNA horizontal transfer events from transient canine hosts (Rebbeck et al. 2011, Strakova et al. 2016). The transfer of mtDNA in CTVT can be robustly detected using phylogenetics, as the donated mtDNA is genetically distinct (Chapter 4 and Chapter 5). Horizontal transfer of mtDNA may be more widespread than previously appreciated, especially as in human and other non-transmissible cancers, the host donor mtDNA haplotype would be genetically almost identical to the cancer mtDNA haplotype. It is possible that this process may also be occurring in human and other cancers, without being detected, and it has also been suggested to be a normal physiological process (Berridge et al. 2016). However, we cannot exclude the possibility that mtDNA transfer may be specific to CTVT, since shuttling of mitochondria may only be driven by a significant evolutionary advantage for the cell. In CTVT, this evolutionary advantage of a horizontal transfer of a host mtDNA haplotype would arise from the large mutational burden carried by the CTVT mtDNA.
- (3)** After each mtDNA transfer in CTVT, a transient period where two distinct mtDNA haplotypes are present in each cell arises – this is a situation, which would not occur in non-transmissible cancers, as both the donor and cancer haplotype would be

almost identical. This situation enables us to look at processes which may be very difficult to detect in human and other cancers, specifically recombination of mtDNA haplotypes (Chapter 4). As we have detected both ancient and recent mtDNA recombination in CTVT, it may be possible that this process is also more widespread in human and other non-transmissible cancers, but would be impossible to detect using phylogenetic methods. Through studying mtDNA in CTVT, we can learn about mitochondrial processes, knowledge of which may also be applied to understanding cancer biology in general.

- (4)** Our understanding of how CTVT escapes the detection by the immune system is still limited (see Chapter 1, section 1.2.5). Although this avenue has not been directly explored in this Thesis, study of mutations shared by all CTVT tumours may help us to understand how CTVT has adapted to be tolerated as a completely genetically unmatched graft. Understanding the interaction of CTVT with the immune system would have extensive applications for transplant biology specifically, but also for cancer immunology in general.

6.2.2 Relevance of my PhD work

Studying CTVT as a transmissible cancer has many broad implications and advantages, which I have highlighted and discussed above, in section 6.2.1. Here, I discuss the direct implications of the work presented in this Thesis.

- (1)** Questionnaire-based study highlighting the historical as well as current prevalence and distribution of CTVT around the world will serve as a useful resource for veterinarians and veterinary workers worldwide (Chapter 2). Information distributed to veterinarians worldwide through the survey as well as information leaflets may help with providing advice about the disease and its management.
- (2)** Survey of factors influencing CTVT prevalence highlighted the factors that continue to modify CTVT's prevalence around the world together with implicating free-roaming dogs as a reservoir for the disease (Chapter 2). Importantly, this study may assist policy-makers and veterinarians in the development of measures to more effectively control and reduce CTVT prevalence and prevent further spread of the disease.

- (3)** Collection of over thousand CTVT samples from more than 50 countries around the world (Chapter 3) has enabled me to perform the CTVT genetic analysis presented in Thesis (Chapter 4 and Chapter 5). Moreover, this set of samples has proven to be a great resource for other members of the Transmissible Cancer Group, as well as for future analyses.
- (4)** Sequencing of CTVT tumours using a number of DNA sequencing approaches (Chapter 4 and Chapter 5) has enabled me to perform the CTVT genetic analysis presented in this Thesis (Chapter 4 and Chapter 5). Moreover, the genetic data are being used by other members of the Transmissible Cancer Group and will continue to be a valuable resource to be analysed in the future.
- (5)** Phylogenetic analysis of CTVT mtDNA from over 50 countries around the world enabled us to trace the historical spread of the disease in the last 2,000 years, highlighting that the disease has been transported between continents by ancient maritime trade routes (Chapter 4). Tracking the spread of CTVT revealed information about historical movement of dogs, providing a resource for historians focusing on history of science.
- (6)** Identification of a number of mtDNA horizontal transfer events as well as detection of mtDNA recombination between mtDNA haplotypes (Chapter 4) may illuminate processes that may be happening on a much larger scale in human and non-transmissible cancers. This work may add to the current knowledge of cancer biology and highlight processes, which had not yet been detected to act in human cancer.
- (7)** The finding that a specific mtDNA haplogroup in CTVT has undergone recurrent mtDNA horizontal transfer may have further implications for understanding mtDNA function in cancer and it may be of an interest for the mitochondrial community. Through learning about selective advantages of particular mtDNA haplotypes, we may gain further understanding about the role of mitochondria in cancer, but also about feasibility of mitochondrial replacement therapies.
- (8)** Although the primary aim of this Thesis was to sequence the CTVT tumour samples, matched host (canine) samples were also sequenced as a part of the study. The primary purpose of these sequences was to exclude contaminating host alleles (see section 4.2.5.2 and section 4.2.6.2), however, nevertheless, these data will be of considerable interest for dog population genetics. The mitochondrial and the exome

sequences have provided a large amount of information on the sequences of mainly street/mixed dogs. Further analysis of the phylogenetic tree of the host samples may provide novel information about the genetic differences and adaptations of dogs in different parts of the world, especially because genomes of street/mixed breed dogs were previously rarely sequenced (Parker et al. 2004, Freedman et al. 2014).

6.3 Limitations of this research

In this section, I discuss the limitations of the work presented in this Thesis.

CTVT is a disease which is present all over the world, as evidenced in Chapter 2. The questionnaire-based survey has confirmed that CTVT is endemic in at least 90 countries worldwide across all inhabited continents. Although I have obtained responses from 109 countries worldwide, there is still a significant number of countries where we do not have CTVT prevalence data from (such as countries in sub-Saharan Africa and central Asia). This study was limited by availability of responses from each country and some countries were more difficult to reach than others. Overall, CTVT prevalence data presented in this Thesis are limited by the ability to contact respondents from each country.

In order to study genetic diversity of CTVT around the world, my aim was to collect samples from all six inhabited continents, to represent the worldwide genetic diversity (Chapter 3). Subsequent genetic analysis was always limited by the availability of CTVT samples collected. Further sample collection from regions and countries with no representation in the sample set would be desirable (see Chapter 3, Figure 3-10).

Analysis of CTVT mtDNA samples from around the world has identified at least 17 mtDNA horizontal transfer events which took place during the history of the cancer lineage and have subsequently been fixed (Chapter 4 and Chapter 5). It should be considered that with an increased sample size we may, very likely, detect additional mtDNA horizontal transfer events. The number of mtDNA horizontal transfers identified was therefore limited by our sampling coverage across the world. Further genetic analysis of additional CTVT samples collected in the future may reveal a number of additional CTVT mtDNA clades.

A single mtDNA haplogroup (A1d1) has been found to recurrently undergo multiple mtDNA horizontal transfer events. Analysis of the A1d1 haplogroup highlighted a number of candidate genetic variants which may be conferring a selective advantage to this haplotype in CTVT. Specifically, an indel at position 16660 CCCTTTTTTTTCCC > CCCCTTTTTTTTCC has been suggested to possibly influence mitochondrial transcription or replication. However, no functional evidence for the role of indel at position 16660 CCCTTTTTTTTCCC > CCCCTTTTTTTTCC has yet been provided, thus limiting the conclusions which could be made from the work presented in Chapter 5. Further functional analysis will be required before any conclusions can be made, as discussed in section 5.4.3. The main limitation of the

current functional pilot studies was the challenge of obtaining robust clinical data, as discussed further in section 5.4.3.

6.4 Future directions

The following projects are priorities for future work, which would expand on the work carried out in this Thesis:

(1) Understanding genetic features of A1d1 haplogroup which lead to recurrent mtDNA horizontal transfer

As discussed in Chapter 5, understanding the genetic features that lead to recurrent mtDNA transfer of the A1d1 haplogroup would be of a great interest to the mitochondrial community. Understanding this phenomenon may allow us to learn about selective advantages of particular mtDNA haplotypes. Ultimately, we may not just gain further understanding about the role of mitochondria in cancer, but also about feasibility of mitochondrial replacement therapies and role of mitochondria in three-parent embryos.

The future studies that could be performed are described in detail in section 5.4.3, and in short presented here: analysis of RNAseq expression data from CTVT tumours with A1d1 haplogroup in comparison to other CTVT tumours; promoter activity assays; clinical follow-up studies to understand progression during treatment; CTVT treatment questionnaire to understand differences in treatment response and dog management during treatment, transmission electron microscopy (TEM) analysis to identify structural differences in distinct mtDNA haplotypes and finally, in vitro fluorescent labelling experiments.

(2) Performing a large-scale phenotypic screen on all CTVT samples and linking the results to the underlying genetic features

Phenotypic analysis of clade 1 and clade 2 CTVT samples has been performed through a histopathology screen approach, as discussed in Chapter 4. Even though no statistically significant results have been obtained, the data indicated that there may be differences in specific histopathology features between individual CTVT clades. It would be desirable to repeat the histopathology screen with a larger number of CTVT samples. Moreover, including additional scoring parameters in the screen may reveal further histopathology differences, which have not yet been highlighted (see section 4.4.8 for discussion of additional scoring parameters).

Additionally, future studies could include a screen of gross characteristics of CTVT tumours (scored using photos), as described in section 4.4.8. The appearance of CTVT

has also been noted to differ between tumours and it would be interesting to follow up on the current preliminary work.

Finally, after performing a more extensive histopathology screen as well as a screen of gross characteristics, the ultimate aim would be to combine all phenotypic features and link these to the underlying genetic features identified from the DNA sequencing data. So far, the comparisons in this Thesis were only made between CTVT clades (see section 4.3.9). In the future, phenotypic features could be compared between individual branches of the CTVT nuclear tree constructed using CTVT exome sequencing data (see Figure 5-4B) to identify sub-groups of CTVT samples sharing particular phenotypic traits and to link them to branch-specific genetic changes.

(3) Expanding the set of CTVT samples to include samples from more countries

As presented in Chapter 3, as a part of my PhD work, I have collected over 1400 CTVT samples from over 50 different countries around the world. The locations where these samples have been collected from (Chapter 3, Figure 3-10), however, will nevertheless remain the limiting factor (as discussed in section 6.3). Through further expanding the set of CTVT samples included in the analysis, we may be able to:

- Identify new, previously unknown CTVT clades
- Locate the place of CTVT origin

Identification of new CTVT mtDNA clades has already been discussed in section 6.3. The place of CTVT origin has not been discussed before – with the current data presented in this Thesis, the place of CTVT origin remains currently unknown. The current data may suggest origin in South-East Asia: (1) the oldest clade (clade 4, see Chapter 4) was found in India and (2) the greatest genetic diversity of CTVT, indicating that CTVT has been present in this region for a very long time, was also located in India. This data, however, does not provide evidence for CTVT origin in India and further genetic analysis of samples collected in remote regions of the globe may shed more light on this question in the future.

(4) Analysis of host DNA sequences

As discussed in section 6.2.2, over 500 host DNA sequences have been generated as part of this PhD study. Future analysis of the genetic data may be of a great interest, especially as DNA from street/mixed breed dogs was previously rarely sequenced (Parker et al.

2004, Freedman et al. 2014). Potential implications for studies involving host DNA sequences would be two-fold: (1) the data may provide novel information about the genetic differences and adaptations of dogs in different parts of the world and (2) analysis of host DNA sequences of dogs infected with CTVT may lead to new insights about interactions between CTVT and its host.

6.5 Conclusions

Overall, in this Thesis I have come to conclusions that may bridge the different branches of CTVT research and further our understanding of transmissible cancer evolution, CTVT spread around the world and adaptations to survival as a long-lived lineage. The main findings arising from this work are listed below:

- (1)** CTVT is a canine infectious disease endemic in at least 90 countries worldwide and affects a significant proportion of dog population around the world.
- (2)** Free-roaming dogs were implicated as a reservoir for the disease, while dog spaying and neutering were associated with reduced CTVT prevalence.
- (3)** Analysis of mtDNA sequencing data has revealed that mtDNA horizontal transfer has occurred a number of times in the history of CTVT, delineating individual tumour clades.
- (4)** CTVT appears to have spread rapidly around the world within the last 2,000 years, perhaps transported by dogs travelling along historic maritime trade routes.
- (5)** Negative selection has been detected to operate in CTVT to prevent accumulation of deleterious mutations in captured mtDNA
- (6)** MtDNA recombination has caused occasional mtDNA re-assortment in CTVT.
- (7)** Analysis of CTVT exome sequencing data has revealed that a single haplogroup (A1d1) has undergone recurrent haplotype-specific mtDNA horizontal transfer events.

In summary, this study combined questionnaire-based approaches, sample collection and validation followed by DNA extraction and DNA sequencing, analysis of mtDNA data from a diverse set of CTVT tumours and analysis of exome sequencing data from a diverse set of CTVT tumours to understand the genetic diversity and evolution of the longest living cancer.

REFERENCES

- Abuom, T. O. (2006). "Transmissible Venereal Tumor with Subcutaneous and Bone Metastasis in a Dog." *Kenya Vet* **30**(1): 11-13.
- Adams, E. W. and Slaughter, L. J. (1970). "A canine venereal tumor with metastasis to the brain." *Pathol Vet* **7**(6): 498-502.
- Adey, A., Burton, J. N., Kitzman, J. O., Hiatt, J. B., Lewis, A. P., Martin, B. K., Qiu, R., Lee, C. and Shendure, J. (2013). "The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line." *Nature* **500**(7461): 207-211.
- Ahmad, T., Mukherjee, S., Pattnaik, B., Kumar, M., Singh, S., Kumar, M., Rehman, R., Tiwari, B. K., Jha, K. A., Barhanpurkar, A. P., Wani, M. R., Roy, S. S., Mabalirajan, U., Ghosh, B. and Agrawal, A. (2014). "Miro1 regulates intercellular mitochondrial transport & enhances mesenchymal stem cell rescue efficacy." *EMBO J* **33**(9): 994-1010.
- Ajello, P. (1939). *Osservazioni sul tumore genitale contagioso del cane* Doctoral thesis, Facolta do medicina veterinaria, Universita di Messina.
- Ajello, P. (1960). "Trasmissione del tumore di Sticker con materiale acellulare." *Nuova Vet* **36**: 179-183.
- Ajello, P. (1980). "Il tumore di Sticker." *Annali Facolta Medicina Veterinaria Mes* **17**: 289-339.
- Albanese, F., Poli, A., Millanta, F. and Abramo, F. (2002). "Primary cutaneous extragenital canine transmissible venereal tumour with Leishmania-laden neoplastic cells: a further suggestion of histiocytic origin?" *Vet Dermatol* **13**(5): 243-246.
- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Borresen-Dale, A. L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjord, J. E., Foekens, J. A., Greaves, M., Hosoda, F., Hutter, B., Ilicic, T., Imbeaud, S., Imielinski, M., Jager, N., Jones, D. T., Jones, D., Knappskog, S., Kool, M., Lakhani, S. R., Lopez-Otin, C., Martin, S., Munshi, N. C., Nakamura, H., Northcott, P. A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J. V., Puente, X. S., Raine, K., Ramakrishna, M., Richardson, A. L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T. N., Span, P. N., Teague, J. W., Totoki, Y., Tutt, A. N., Valdes-Mas, R., van Buuren, M. M., van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L. R., Australian Pancreatic Cancer Genome Initiative, IcgC Breast Cancer Consortium, IcgC Mmml-Seq Consortium, IcgC PedBrain, Zucman-Rossi, J., Futreal, P. A., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S. M., Siebert, R., Campo, E., Shibata, T., Pfister, S. M., Campbell, P. J. and Stratton, M. R. (2013). "Signatures of mutational processes in human cancer." *Nature* **500**(7463): 415-421.
- Algarra, I., Collado, A. and Garrido, F. (1997). "Altered MHC class I antigens in tumors." *Int J Clin Lab Res* **27**(2): 95-102.
- Alvar, J., Velez, I. D., Bern, C., Herrero, M., Desjeux, P., Cano, J., Jannin, J., den Boer, M. and WHO Leishmaniasis Control Team (2012). "Leishmaniasis worldwide and global estimates of its incidence." *PLoS One* **7**(5): e35671.
- Amariglio, E. N., Hakim, I., Brok-Simoni, F., Grossman, Z., Katzir, N., Harmelin, A., Ramot, B. and Rechavi, G. (1991). "Identity of rearranged LINE/c-MYC junction sequences specific for the canine transmissible venereal tumor." *Proc Natl Acad Sci U S A* **88**(18): 8136-8139.
- Amber, E. I., Henderson, R. A., Adeyanju, J. B. and Gyang, E. O. (1990). "Single-drug chemotherapy of canine transmissible venereal tumor with cyclophosphamide, methotrexate, or vincristine." *J Vet Intern Med* **4**(3): 144-147.
- Angell, J. W., Duncan, J. S., Carter, S. D. and Grove-White, D. H. (2014). "Farmer reported prevalence and factors associated with contagious ovine digital dermatitis in

- Wales: A questionnaire of 511 sheep farmers." Prev Vet Med **113**(1): 132-138.
- Applied Biosystems. (2004). "Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR." Retrieved 5th October 2015, from [http://www.gu.se/digitalAssets/1125/1125331_ABI - Guide Relative Quantification using realtime PCR.pdf](http://www.gu.se/digitalAssets/1125/1125331_ABI_Guide_Relative_Quantification_using_realtime_PCR.pdf).
- Avery, O. T., Macleod, C. M. and McCarty, M. (1944). "Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii." J Exp Med **79**(2): 137-158.
- Baez-Ortega, A. (2017a). "mutree: A pipeline for phylogenetic tree inference and recurrent mutation discovery." doi:10.5281/zenodo.583634.
- Baez-Ortega, A. (2017b). "Somatypus: A Platypus-based variant calling pipeline for cancer data." doi:10.5281/zenodo.268310.
- Balmain, A. (2001). "Cancer genetics: from Boveri and Mendel to microarrays." Nat Rev Cancer **1**(1): 77-82.
- Banfield, W. G., Woke, P. A., Mackay, C. M. and Cooper, H. L. (1965). "Mosquito transmission of a reticulum cell sarcoma of hamsters." Science **148**: 1239-1240.
- Barron, C. N., Saunders, L. Z., Seibold, H. R. and Heath, M. K. (1963). "Intraocular Tumors in Animals. V. Transmissible Venereal Tumor of Dogs." Am J Vet Res **24**: 1263-1270.
- Barski, G. and Cornefert-Jensen, F. (1966). "Cytogenetic study of Sticker venereal sarcoma in European dogs." J Natl Cancer Inst **37**(6): 787-797.
- Bastan, A. (2008). "Uterine and ovarian metastasis of transmissible venereal tumor in a bitch." Turk J Vet Anim Sci **32**(1): 65-66.
- Batista, J. S., Soares, H. S., Pereira, R. H. M. A., Petri, A. A., Sousa, F. D. N. and Nunes, F. C. R. (2007). "Canine transmissible venereal tumor with intra-ocular localization and spleen metastasis (Tumor venereo transmissivel canino com localizacao intra-ocular e metastase no baco)." Acta Vet Brasilica **1**: 45-48.
- Beebe, S. P. (1907). "The growth of lymphosarcoma in dogs-summary of recent observations." JAMA **49**(18): 1492-1493.
- Beebe, S. P. and Ewing, J. (1906). "A Study of the so-called infectious Lymphosarcoma of Dogs." J Med Res **15**(2): 209-228 205.
- Belkin, P. V. (1959). "Extragenital venereal granuloma in the abdominal organs of a dog." J Am Vet Med Assoc **135**: 575-576.
- Belpomme, D., Irigaray, P., Hardell, L., Clapp, R., Montagnier, L., Epstein, S. and Sasco, A. J. (2007). "The multitude and diversity of environmental carcinogens." Environ Res **105**(3): 414-429.
- Bergell, P. and Sticker, A. (1907). "Ueber Pathogenese und uber den spezifischen Abbau der Krebsgeschwulste." Dtsch Med Wochenschr **2**(38): 1521-1522.
- Bergthorsson, U., Adams, K. L., Thomason, B. and Palmer, J. D. (2003). "Widespread horizontal transfer of mitochondrial genes in flowering plants." Nature **424**(6945): 197-201.
- Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J. S., Dobson, J., Urashima, M., Mc Henry, K. T., Pinchback, R. M., Ligon, A. H., Cho, Y. J., Haery, L., Greulich, H., Reich, M., Winckler, W., Lawrence, M. S., Weir, B. A., Tanaka, K. E., Chiang, D. Y., Bass, A. J., Loo, A., Hoffman, C., Prensner, J., Liefeld, T., Gao, Q., Yecies, D., Signoretti, S., Maher, E., Kaye, F. J., Sasaki, H., Tepper, J. E., Fletcher, J. A., Taberero, J., Baselga, J., Tsao, M. S., Demichelis, F., Rubin, M. A., Janne, P. A., Daly, M. J., Nucera, C., Levine, R. L., Ebert, B. L., Gabriel, S., Rustgi, A. K., Antonescu, C. R., Ladanyi, M., Letai, A., Garraway, L. A., Loda, M., Beer, D. G., True, L. D., Okamoto, A.,

- Pomeroy, S. L., Singer, S., Golub, T. R., Lander, E. S., Getz, G., Sellers, W. R. and Meyerson, M. (2010). "The landscape of somatic copy-number alteration across human cancers." *Nature* **463**(7283): 899-905.
- Berridge, M. V., Dong, L. and Neuzil, J. (2015). "Mitochondrial DNA in tumor initiation, progression, and metastasis: role of horizontal mtDNA transfer." *Cancer Res* **75**(16): 3203-3208.
- Berridge, M. V., McConnell, M. J., Grasso, C., Bajzikova, M., Kovarova, J. and Neuzil, J. (2016). "Horizontal transfer of mitochondria between mammalian cells: beyond co-culture approaches." *Curr Opin Genet Dev* **38**: 75-82.
- Blaine, D. P. (1810). *A domestic treatise on the diseases of horses and dogs.*, London. T. Boosey.
- Bloom, F., Paff, G. H. and Noback, C. R. (1951). "The transmissible venereal tumor of the dog; studies indicating that the tumor cells are mature end cells of reticulo-endothelial origin." *Am J Pathol* **27**(1): 119-139.
- Bogenhagen, D. and Clayton, D. A. (1974). "The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial deoxyribonucleic acid." *J Biol Chem* **249**(24): 7991-7995.
- Boni, M. F., Posada, D. and Feldman, M. W. (2007). "An exact nonparametric method for inferring mosaic structure in sequence triplets." *Genetics* **176**(2): 1035-1047.
- Borrel, M. A. (1907). "Lymhosarcome du chien." *Semaine Medicale (Paris)* **27**: 94-95.
- Boscós, C. (1988). "Canine transmissible venereal tumor: clinical observations and treatment." *Anim Familiaris* **3**: 10-15.
- Botigue, L. R., Song, S., Scheu, A., Gopalan, S., Pendleton, A. L., Oetjens, M., Taravella, A. M., Seregely, T., Zeeb-Lanz, A., Arbogast, R. M., Bobo, D., Daly, K., Unterlander, M., Burger, J., Kidd, J. M. and Veeramah, K. R. (2017). "Ancient European dog genomes reveal continuity since the Early Neolithic." *Nat Commun* **8**: 16082.
- Boveri, T. (1914). "Concerning the Origin of Malignant Tumours."
- Boyko, A. R., Quignon, P., Li, L., Schoenebeck, J. J., Degenhardt, J. D., Lohmueller, K. E., Zhao, K., Brisbin, A., Parker, H. G., vonHoldt, B. M., Cargill, M., Auton, A., Reynolds, A., Elkahloun, A. G., Castelano, M., Mosher, D. S., Sutter, N. B., Johnson, G. S., Novembre, J., Hubisz, M. J., Siepel, A., Wayne, R. K., Bustamante, C. D. and Ostrander, E. A. (2010). "A simple genetic architecture underlies morphological variation in dogs." *PLoS Biol* **8**(8): e1000451.
- Brandon, M., Baldi, P. and Wallace, D. C. (2006). "Mitochondrial mutations in cancer." *Oncogene* **25**(34): 4647-4662.
- Brindley, D. C. and Banfield, W. G. (1961). "A contagious tumor of the hamster." *J Natl Cancer Inst* **26**: 949-957.
- Bronden, L. B., Flagstad, A. and Kristensen, A. T. (2007). "Veterinary cancer registries in companion animal cancer: a review." *Vet Comp Oncol* **5**(3): 133-144.
- Brown, G. K., Kreiss, A., Lyons, A. B. and Woods, G. M. (2011). "Natural killer cell mediated cytotoxic responses in the Tasmanian devil." *PLoS One* **6**(9): e24475.
- Brown, N. O., Calvert, C. and MacEwen, E. G. (1980). "Chemotherapeutic management of transmissible venereal tumors in 30 dogs." *J Am Vet Med Assoc* **176**(10 Pt 1): 983-986.
- Calvert, C. A., Leifer, C. E. and MacEwen, E. G. (1982). "Vincristine for treatment of transmissible venereal tumor in the dog." *J Am Vet Med Assoc* **181**(2): 163-164.
- Campbell, P. J., Pleasance, E. D., Stephens, P. J., Dicks, E., Rance, R., Goodhead, I., Follows, G. A., Green, A. R., Futreal, P. A. and Stratton, M. R. (2008). "Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing." *Proc Natl Acad Sci U S A* **105**(35): 13081-13086.

- Catone, G., Marino, G., Poglayen, G., Gramiccia, M., Ludovisi, A. and Zanghi, A. (2003). "Canine transmissible venereal tumour parasitized by *Leishmania infantum*." Vet Res Commun **27**(7): 549-553.
- Cella, F. (1939). "Sopra un secondo caso di granuloma venereo del cane (c.d. sarcoma di sticker) osservato nell'Emilia - metastasi splenica." Nuova Veterinaria **12**: 217-220.
- Chandler, J. P. and Yang, T. J. (1981). "Canine transmissible venereal sarcoma: distribution of T and B lymphocytes in blood, draining lymph nodes and tumours at different stages of growth." Br J Cancer **44**(4): 514-521.
- Chapman, A. L., Darton, T. C. and Foster, R. A. (2013). "Managing and monitoring tuberculosis using web-based tools in combination with traditional approaches." Clin Epidemiol **5**: 465-473.
- Chen, X. J. (2013). "Mechanism of homologous recombination and implications for aging-related deletions in mitochondrial DNA." Microbiol Mol Biol Rev **77**(3): 476-496.
- Chikweto, A. (2013). "Genital and Extragenital Canine Transmissible Venereal Tumor in Dogs in Grenada, West Indies." Open J Vet Med **3**: 111-114.
- Chinnery, P. F., Thorburn, D. R., Samuels, D. C., White, S. L., Dahl, H. M., Turnbull, D. M., Lightowers, R. N. and Howell, N. (2000). "The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both?" Trends Genet **16**(11): 500-505.
- Choi, Y., Ishiguro, N., Shinagawa, M., Kim, C. J., Okamoto, Y., Minami, S. and Ogihara, K. (1999). "Molecular structure of canine LINE-1 elements in canine transmissible venereal tumor." Anim Genet **30**(1): 51-53.
- Choi, Y. K. and Kim, C. J. (2002). "Sequence analysis of canine LINE-1 elements and p53 gene in canine transmissible venereal tumor." J Vet Sci **3**(4): 285-292.
- Chu, R. M., Lin, C. Y., Liu, C. C., Yang, S. Y., Hsiao, Y. W., Hung, S. W., Pao, H. N. and Liao, K. W. (2001). "Proliferation characteristics of canine transmissible venereal tumor." Anticancer Res **21**(6A): 4017-4024.
- Clark, M. A. and Shay, J. W. (1982). "Mitochondrial transformation of mammalian cells." Nature **295**(5850): 605-607.
- Cockrill, J. M. and Beasley, J. N. (1979). "Transmission of transmissible venereal tumor of the dog to the coyote." Am J Vet Res **4**: 409-410.
- Cohen, D. (1973). "The biological behaviour of the transmissible venereal tumour in immunosuppressed dogs." Europ J Cancer **9**: 253-258.
- Cohen, D. (1985). "The canine transmissible venereal tumor: a unique result of tumor progression." Adv Cancer Res **43**: 75-112.
- Cohen, D., Shalev, A. and Krup, M. (1984). "Lack of beta 2-microglobulin on the surface of canine transmissible venereal tumor cells." J Natl Cancer Inst **72**(2): 395-401.
- Cohen, D. and Steel, G. G. (1972). "Thymidine labelling studies in a transmissible venereal tumour of the dog." Br J Cancer **26**(5): 413-419.
- Coller, H. A., Khrapko, K., Bodyak, N. D., Nekhaeva, E., Herrero-Jimenez, P. and Thilly, W. G. (2001). "High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection." Nat Genet **28**(2): 147-150.
- Connor, T. M., Hoer, S., Mallett, A., Gale, D. P., Gomez-Duran, A., Posse, V., Antrobus, R., Moreno, P., Sciacovelli, M., Frezza, C., Duff, J., Sheerin, N. S., Sayer, J. A., Ashcroft, M., Wiesener, M. S., Hudson, G., Gustafsson, C. M., Chinnery, P. F. and Maxwell, P. H. (2017). "Mutations in mitochondrial DNA causing tubulointerstitial kidney disease." PLoS Genet **13**(3): e1006620.

- Cooper, H. L., Mackay, C. M. and Banfield, W. G. (1964). "Chromosome studies of a contagious reticulum cell sarcoma of the Syrian hamster." J Natl Cancer Inst **33**: 691-706.
- Cotchin, E. (1954). "Neoplasia in the dog." Vet Rec **66**: 879-885.
- Cruz, J. C. (2010). "Canine transmissible venereal tumor in the metropolitan area of Mexico City." Rev Cient **20**(4): 362-366.
- Das, U. and Das, A. K. (2000). "Review of canine transmissible venereal sarcoma." Vet Res Commun **24**(8): 545-556.
- Dass, L. L. and Sahay, P. N. (1989). "Surgical treatment of canine transmissible venereal tumour - a retrospective study." Indian Veterinary Journal **66**: 255-258.
- Dobson, J. M. (2013). "Breed-predispositions to cancer in pedigree dogs." ISRN Vet Sci **2013**: 941275.
- Dong, L. F., Kovarova, J., Bajzikova, M., Bezawork-Geleta, A., Svec, D., Endaya, B., Sachaphibulkij, K., Coelho, A. R., Sebkova, N., Ruzickova, A., Tan, A. S., Kluckova, K., Judasova, K., Zamecnikova, K., Rychtarcikova, Z., Gopalan, V., Andera, L., Sobol, M., Yan, B., Pattnaik, B., Bhatraju, N., Truksa, J., Stopka, P., Hozak, P., Lam, A. K., Sedlacek, R., Oliveira, P. J., Kubista, M., Agrawal, A., Dvorakova-Hortova, K., Rohlena, J., Berridge, M. V. and Neuzil, J. (2017). "Horizontal transfer of whole mitochondria restores tumorigenic potential in mitochondrial DNA-deficient cancer cells." Elife **6**: 10.7554/eLife.22187.
- Drake, A. G. and Klingenberg, C. P. (2010). "Large-scale diversification of skull shape in domestic dogs: disparity and modularity." Am Nat **175**(3): 289-301.
- Duncan, J. R. and Prasse, K. W. (1979). "Cytology of canine cutaneous round cell tumors. Mast cell tumor, histiocytoma, lymphosarcoma and transmissible venereal tumor." Vet Pathol **16**(6): 673-679.
- Dungern, V. (1912). "Zue Biologie des Runderzellensarkom des Hundes." Munch Med Wochenschr **5**: 238.
- Duplay, S. (1894). "Tumeurs expérimentales chez les animaux." Atti dell'XI Congresso Medico Internazionale, Roma **2**: 103-104.
- Epstein, R. B. and Bennett, B. T. (1974). "Histocompatibility typing and course of canine venereal tumors transplanted into unmodified random dogs." Cancer Res **34**: 788-793.
- Eze, C. A., Anyanwu, H. C. and Kene, R. O. C. (2007). "Review of Canine Transmissible Venereal Tumour (TVT) in dogs." Nigerian Vet J **28**(1): 54-70.
- Fay, J. C. and Wu, C. I. (2003). "Sequence divergence, functional constraint, and selection in protein evolution." Annu Rev Genomics Hum Genet **4**: 213-235.
- Feldman, W. H. (1929). "So called infectious sarcoma of the dog in an unusual anatomic situation." Am J Pathol **5**: 183-195.
- Ferreira, A. J., Jaggy, A., Varejao, A. P., Ferreira, M. L., Correia, J. M., Mulas, J. M., Almeida, O., Oliveira, P. and Prada, J. (2000). "Brain and ocular metastases from a transmissible venereal tumour in a dog." J Small Anim Pract **41**(4): 165-168.
- Fliss, M. S., Usadel, H., Caballero, O. L., Wu, L., Buta, M. R., Eleff, S. M., Jen, J. and Sidransky, D. (2000). "Facile detection of mitochondrial DNA mutations in tumors and bodily fluids." Science **287**(5460): 2017-2019.
- Folkman, J. (2002). "Role of angiogenesis in tumor growth and metastasis." Semin Oncol **29**(6 Suppl 16): 15-18.
- Fouad, Y. A. and Aanei, C. (2017). "Revisiting the hallmarks of cancer." Am J Cancer Res **7**(5): 1016-1036.
- Foulds, L. (1954). "The experimental study of tumor progression: a review." Cancer Res **14**(5): 327-339.
- Frantz, L. A., Mullin, V. E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., Linderholm, A., Mattiangeli, V., Teasdale, M. D., Dimopoulos, E. A., Tresset,

- A., Duffraisse, M., McCormick, F., Bartosiewicz, L., Gal, E., Nyerges, E. A., Sablin, M. V., Brehard, S., Mashkour, M., Balasescu, A., Gillet, B., Hughes, S., Chassaing, O., Hitte, C., Vigne, J. D., Dobney, K., Hanni, C., Bradley, D. G. and Larson, G. (2016). "Genomic and archaeological evidence suggest a dual origin of domestic dogs." *Science* **352**(6290): 1228-1231.
- Freedman, A. H., Gronau, I., Schweizer, R. M., Ortega-Del Vecchyo, D., Han, E., Silva, P. M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., Beale, H., Ramirez, O., Hormozdiari, F., Alkan, C., Vila, C., Squire, K., Geffen, E., Kusak, J., Boyko, A. R., Parker, H. G., Lee, C., Tadisotla, V., Siepel, A., Bustamante, C. D., Harkins, T. T., Nelson, S. F., Ostrander, E. A., Marques-Bonet, T., Wayne, R. K. and Novembre, J. (2014). "Genome sequencing highlights the dynamic early history of dogs." *PLoS Genet* **10**(1): e1004016.
- Fregel, R., Suarez, N. M., Betancor, E., Gonzalez, A. M., Cabrera, V. M. and Pestano, J. (2015). "Mitochondrial DNA haplogroup phylogeny of the dog: Proposal for a cladistic nomenclature." *Mitochondrion* **22**: 75-84.
- French, C. (1906). *Surgical diseases and surgery of the dog*, Washington D.C.
- Fujinaga, T., Yamashita, M., Yoshida, M. C., Mizuno, S., Okamoto, Y., Tajima, M. and Otomo, K. (1989). "Chromosome analysis of canine transmissible sarcoma cells." *Zentralbl Veterinarmed A* **36**(7): 481-489.
- Gandotra, V. K. (1993). "Occurrence of canine transmissible venereal tumor and evaluation of two treatments." *Ind Vet J* **70**: 854-857.
- Ganguly, B., Das, U. and Das, A. K. (2013). "Canine transmissible venereal tumour: a review." *Vet Comp Oncol*: DOI: 10.1111/vco.12060.
- Gantenbein, B., Fet, V., Gantenbein-Ritter, I. A. and Balloux, F. (2005). "Evidence for recombination in scorpion mitochondrial DNA (Scorpiones: Buthidae)." *Proc Biol Sci* **272**(1564): 697-704.
- Gartner, H. V., Seidl, C., Luckenbach, C., Schumm, G., Seifried, E., Ritter, H. and Bultmann, B. (1996). "Genetic analysis of a sarcoma accidentally transplanted from a patient to a surgeon." *N Engl J Med* **335**(20): 1494-1496.
- Gibbs, M. J., Armstrong, J. S. and Gibbs, A. J. (2000). "Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences." *Bioinformatics* **16**(7): 573-582.
- Giles, R. E., Blanc, H., Cann, H. M. and Wallace, D. C. (1980). "Maternal inheritance of human mitochondrial DNA." *Proc Natl Acad Sci U S A* **77**(11): 6715-6719.
- Gimeno, E. J., Massone, A. R., Marino, F. P. and Idiart, J. R. (1995). "Intermediate filament expression and lectin histochemical features of canine transmissible venereal tumour." *APMIS* **103**(9): 645-650.
- Ginel, P. J., Molleda, J. M., Novales, M., Martin, E., Margarito, J. M. and Lopez, R. (1995). "Primary Transmissible Venereal Tumor in the Nasal Cavity of a Dog." *Veterinary Record* **136**(9): 222-223.
- Gonzalez, C. M., Griffey, S. M., Naydan, D. K., Flores, E., Cepeda, R., Cattaneo, G. and Madewell, B. R. (2000). "Canine transmissible venereal tumour: a morphological and immunohistochemical study of 11 tumours in growth phase and during regression after chemotherapy." *J Comp Pathol* **122**(4): 241-248.
- Goodwin, S., McPherson, J. D. and McCombie, W. R. (2016). "Coming of age: ten years of next-generation sequencing technologies." *Nat Rev Genet* **17**(6): 333-351.
- Gradoni, L. (2013). "Epidemiological surveillance of leishmaniasis in the European Union: operational and research challenges." *Euro Surveill* **18**(30): 20539.
- Gray, M. M., Granka, J. M., Bustamante, C. D., Sutter, N. B., Boyko, A. R., Zhu, L., Ostrander, E. A. and Wayne, R. K. (2009). "Linkage disequilibrium and demographic history of wild and domestic canids." *Genetics* **181**(4): 1493-1505.

- Guedes, R. M. C., Nogueira, R. H. G., Santos, S. R. Q. and Souza, E. M. (1996). "Extragenital transmissible venereal tumor in a dog. Case report." Arq Bras Med Vet Zootec **48**(3): 369-374.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010). "New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0." Systematic Biology **59**(3): 307-321.
- Gundry, R. L., Allard, M. W., Moretti, T. R., Honeycutt, R. L., Wilson, M. R., Monson, K. L. and Foran, D. R. (2007). "Mitochondrial DNA analysis of the domestic dog: control region variation within and among breeds." J Forensic Sci **52**(3): 562-572.
- Gurel, A. (2002). "Transmissible Venereal Tumors Detected in the Extragenital Organs of Dogs." Israel J Vet Med **57**(2): 1-8.
- Habano, W., Sugai, T., Nakamura, S. I., Uesugi, N., Yoshida, T. and Sasou, S. (2000). "Microsatellite instability and mutation of mitochondrial and nuclear DNA in gastric carcinoma." Gastroenterology **118**(5): 835-841.
- Hagstrom, E., Freyer, C., Battersby, B. J., Stewart, J. B. and Larsson, N. G. (2014). "No recombination of mtDNA after heteroplasmy for 50 generations in the mouse maternal germline." Nucleic Acids Res **42**(2): 1111-1116.
- Hall, T. A. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." Nucleic Acids Symposium Series **41**: 95-98.
- Hamir, A. N. (1985). "Primary penile and nasal transmissible venereal tumours in a dog." Aust Vet J **62**(12): 430-432.
- Hamir, A. N. (1986). "Neoplasms of dogs in Papua New Guinea." Aust Vet J **63**(10): 342-343.
- Hanahan, D. and Weinberg, R. A. (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.
- Hanahan, D. and Weinberg, R. A. (2011). "Hallmarks of cancer: the next generation." Cell **144**(5): 646-674.
- Harmelin, A., Pinthus, J. H., Katzir, N., Kapon, A., Volcani, Y., Amariglio, E. N. and Rehavi, G. (2001). "Use of a murine xenograft model for canine transmissible venereal tumor." Am J Vet Res **62**(6): 907-911.
- Hatefi, Y. (1985). "The mitochondrial electron transport and oxidative phosphorylation system." Annu Rev Biochem **54**: 1015-1069.
- Hayakawa, K., Esposito, E., Wang, X., Terasaki, Y., Liu, Y., Xing, C., Ji, X. and Lo, E. H. (2016). "Transfer of mitochondria from astrocytes to neurons after stroke." Nature **535**(7613): 551-555.
- Hayes, H. M., Jr., Biggar, R. J., Pickle, L. W., Hoover, R. and Toft, J. D., 2nd (1983). "Canine transmissible venereal tumor: a model for Kaposi's sarcoma?" Am J Epidemiol **117**(1): 108-109.
- Hecht, N. B., Liem, H., Kleene, K. C., Distel, R. J. and Ho, S. M. (1984). "Maternal inheritance of the mouse mitochondrial genome is not mediated by a loss or gross alteration of the paternal mitochondrial DNA or by methylation of the oocyte mitochondrial DNA." Dev Biol **102**(2): 452-461.
- Hecht, S. S. (2003). "Tobacco carcinogens, their biomarkers and tobacco-induced cancer." Nat Rev Cancer **3**(10): 733-744.
- Higgins, D. A. (1966). "Observations on the canine transmissible venereal tumour as seen in the Bahamas." Vet Rec **79**(3): 67-71.
- Hill, J. H., Chen, Z. and Xu, H. (2014). "Selective propagation of functional mitochondrial DNA during oogenesis restricts the transmission of a deleterious mitochondrial variant." Nat Genet **46**(4): 389-392.
- Hoarau, G., Holla, S., Lescasse, R., Stam, W. T. and Olsen, J. L. (2002). "Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*." Mol Biol Evol **19**(12): 2261-2264.
- Hobday, F. (1900). Canine and Feline Surgery (Chapter XV. Operations on the genital organs), W.&A.K. Johnston.

- Hobday, F. (1905). "Observations on contagious venereal tumours in canine patients." Vet J **2**: 342-346.
- Hobday, F. (1906). Surgical diseases of the dog and cat, with chapters on anaesthetics and obstetrics (second edition of 'Canine and feline surgery') W. T. Keener & Co.
- Holmes, J. M. (1981). "Measurement of the rate of death of canine transmissible venereal tumour cells transplanted into dogs and nude mice." Res Vet Sci **30**(2): 248-250.
- Hordijk, W. and Gascuel, O. (2005). "Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood." Bioinformatics **21**(24): 4338-4347.
- Hsiao, Y. W., Liao, K. W., Chung, T. F., Liu, C. H., Hsu, C. D. and Chu, R. M. (2008). "Interactions of host IL-6 and IFN-gamma and cancer-derived TGF-beta1 on MHC molecule expression during tumor spontaneous regression." Cancer Immunol Immunother **57**(7): 1091-1104.
- Hsiao, Y. W., Liao, K. W., Hung, S. W. and Chu, R. M. (2002). "Effect of tumor infiltrating lymphocytes on the expression of MHC molecules in canine transmissible venereal tumor cells." Vet Immunol Immunopathol **87**(1-2): 19-27.
- Hsiao, Y. W., Liao, K. W., Hung, S. W. and Chu, R. M. (2004). "Tumor-infiltrating lymphocyte secretion of IL-6 antagonizes tumor-derived TGF-beta 1 and restores the lymphokine-activated killing activity." J Immunol **172**(3): 1508-1514.
- Huson, D. H. and Scornavacca, C. (2012). "Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks." Syst Biol.
- Idowu, A. L. (1984). "A retrospective evaluation of four surgical methods of treating canine transmissible venereal tumour." J Small Anim Pract **25**: 191-198.
- Idowu, L. (1977). "The chromosomes of the transmissible venereal tumour of dogs in Ibadan, Nigeria." Res Vet Sci **22**(3): 271-273.
- International Monetary Fund. (2013). "World Economic Outlook database." Retrieved 9th March 2013, from <http://www.imf.org/external/pubs/ft/weo/2011>.
- Islam, M. N., Das, S. R., Emin, M. T., Wei, M., Sun, L., Westphalen, K., Rowlands, D. J., Quadri, S. K., Bhattacharya, S. and Bhattacharya, J. (2012). "Mitochondrial transfer from bone-marrow-derived stromal cells to pulmonary alveoli protects against acute lung injury." Nat Med **18**(5): 759-765.
- Jackson, C. (1944). "The cytology of the contagious (venereal) tumour of the dog." Onderstepoort J Vet Sci Anim Ind **20**: 97-118.
- Ju, Y. S., Alexandrov, L. B., Gerstung, M., Martincorena, I., Nik-Zainal, S., Ramakrishna, M., Davies, H. R., Papaemmanuil, E., Gundem, G., Shlien, A., Bolli, N., Behjati, S., Tarpey, P. S., Nangalia, J., Massie, C. E., Butler, A. P., Teague, J. W., Vassiliou, G. S., Green, A. R., Du, M. Q., Unnikrishnan, A., Pimanda, J. E., Teh, B. T., Munshi, N., Greaves, M., Vyas, P., El-Naggar, A. K., Santarius, T., Collins, V. P., Grundy, R., Taylor, J. A., Hayes, D. N., Malkin, D., IcgC Breast Cancer Group, IcgC Chronic Myeloid Disorders Group, IcgC Prostate Cancer Group, Foster, C. S., Warren, A. Y., Whitaker, H. C., Brewer, D., Eeles, R., Cooper, C., Neal, D., Visakorpi, T., Isaacs, W. B., Bova, G. S., Flanagan, A. M., Futreal, P. A., Lynch, A. G., Chinnery, P. F., McDermott, U., Stratton, M. R. and Campbell, P. J. (2014). "Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer." Elife **3**: 10.7554/eLife.02935.
- Ju, Y. S., Tubio, J. M., Mifsud, W., Fu, B., Davies, H. R., Ramakrishna, M., Li, Y., Yates, L., Gundem, G., Tarpey, P. S., Behjati, S., Papaemmanuil, E., Martin, S., Fullam, A., Gerstung, M., IcgC Prostate Cancer Working Group, IcgC Bone Cancer Working Group, IcgC Breast Cancer Working Group, Nangalia, J., Green, A. R., Caldas, C., Borg, A., Tutt, A., Lee, M. T., van't Veer, L. J., Tan, B. K., Aparicio, S., Span, P. N., Martens, J. W., Knappskog, S., Vincent-Salomon, A., Borresen-

- Dale, A. L., Eyfjord, J. E., Flanagan, A. M., Foster, C., Neal, D. E., Cooper, C., Eeles, R., Lakhani, S. R., Desmedt, C., Thomas, G., Richardson, A. L., Purdie, C. A., Thompson, A. M., McDermott, U., Yang, F., Nik-Zainal, S., Campbell, P. J. and Stratton, M. R. (2015). "Frequent somatic transfer of mitochondrial DNA into the nuclear genome of human cancer cells." *Genome Res* **25**(6): 814-824.
- Kakpakova, E. S., Pogosiants, E. E. and Ponomarev, V. I. (1968). "[Characteristics of the karyotype of cells of transmissible dog sarcoma]." *Vopr Onkol* **14**(11): 43-50.
- Kallberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H. and Xu, J. (2012). "Template-based protein structure modeling using the RaptorX web server." *Nat Protoc* **7**(8): 1511-1522.
- Kang, E., Wu, J., Gutierrez, N. M., Koski, A., Tippner-Hedges, R., Agaronyan, K., Platero-Luengo, A., Martinez-Redondo, P., Ma, H., Lee, Y., Hayama, T., Van Dyken, C., Wang, X., Luo, S., Ahmed, R., Li, Y., Ji, D., Kayali, R., Cinnioglu, C., Olson, S., Jensen, J., Battaglia, D., Lee, D., Wu, D., Huang, T., Wolf, D. P., Temiakov, D., Belmonte, J. C., Amato, P. and Mitalipov, S. (2016). "Mitochondrial replacement in human oocytes carrying pathogenic mitochondrial DNA mutations." *Nature* **540**(7632): 270-275.
- Karlson, A. G. and Mann, F. C. (1952). "The transmissible venereal tumor of dogs: observations on forty generations of experimental transfers." *Ann N Y Acad Sci* **54**(6): 1197-1213.
- Katzir, N., Arman, E., Cohen, D., Givol, D. and Rechavi, G. (1987). "Common origin of transmissible venereal tumors (TVT) in dogs." *Oncogene* **1**(4): 445-448.
- Katzir, N., Rechavi, G., Cohen, J. B., Unger, T., Simoni, F., Segal, S., Cohen, D. and Givol, D. (1985). "Retroposon insertion into the cellular oncogene c-myc in canine transmissible venereal tumor." *Proc Natl Acad Sci U S A* **82**(4): 1054-1058.
- Kauffman, H. M., McBride, M. A., Cherikh, W. S., Spain, P. C. and Delmonico, F. L. (2002). "Transplant tumor registry: donors with central nervous system tumors." *Transplantation* **73**(4): 579-582.
- Kim, K. S., Lee, S. E., Jeong, H. W. and Ha, J. H. (1998). "The complete nucleotide sequence of the domestic dog (*Canis familiaris*) mitochondrial genome." *Mol Phylogenet Evol* **10**(2): 210-220.
- Kimeto, B. and Mugeru, G. M. (1974). "Transmissible venereal tumour of dog in Kenya." *Bull Epizoot Dis Afr* **22**(4): 327-329.
- Knudson, A. G., Jr. (1971). "Mutation and cancer: statistical study of retinoblastoma." *Proc Natl Acad Sci U S A* **68**(4): 820-823.
- Kraytsberg, Y., Schwartz, M., Brown, T. A., Ebraldise, K., Kunz, W. S., Clayton, D. A., Vissing, J. and Khrapko, K. (2004). "Recombination of human mitochondrial DNA." *Science* **304**(5673): 981.
- Kretzschmar, M., Mangen, M. J., Pinheiro, P., Jahn, B., Fevre, E. M., Longhi, S., Lai, T., Havelaar, A. H., Stein, C., Cassini, A., Kramarz, P. and CoDE consortium (2012). "New methodology for estimating the burden of infectious diseases in Europe." *PLoS Med* **9**(4): e1001205.
- Ladoukakis, E. D. and Zouros, E. (2001). "Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA." *Mol Biol Evol* **18**(7): 1168-1175.
- Larman, T. C., DePalma, S. R., Hadjipanayis, A. G., Cancer Genome Atlas Research Network, Protopopov, A., Zhang, J., Gabriel, S. B., Chin, L., Seidman, C. E., Kucherlapati, R. and Seidman, J. G. (2012). "Spectrum of somatic mitochondrial mutations in five cancers." *Proc Natl Acad Sci U S A* **109**(35): 14087-14091.
- Larson, G., Karlsson, E. K., Perri, A., Webster, M. T., Ho, S. Y., Peters, J., Stahl, P. W., Piper, P. J., Lingaas, F., Fredholm, M., Comstock, K. E., Modiano, J. F., Schelling, C., Agoulnik, A. I., Leegwater, P. A., Dobney, K., Vigne, J. D., Vila, C., Andersson,

- L. and Lindblad-Toh, K. (2012). "Rethinking dog domestication by integrating genetics, archeology, and biogeography." *Proc Natl Acad Sci U S A* **109**(23): 8878-8883.
- Larsson, N. G. and Clayton, D. A. (1995). "Molecular genetic aspects of human mitochondrial disorders." *Annu Rev Genet* **29**: 151-178.
- Lee, E. Y. and Muller, W. J. (2010). "Oncogenes and tumor suppressor genes." *Cold Spring Harb Perspect Biol* **2**(10): a003236.
- Lee, S. R. and Han, J. (2017). "Mitochondrial nucleoid: Shield and switch of the mitochondrial genome." *Oxid Med Cell Longev* **2017**: 8060949.
- Leonard, J. A., Wayne, R. K., Wheeler, J., Valadez, R., Guillen, S. and Vila, C. (2002). "Ancient DNA evidence for Old World origin of New World dogs." *Science* **298**(5598): 1613-1616.
- Levy, E., Mylonakis, M. E., Saridomichelakis, M. N., Polizopoulou, Z. S., Psychogios, V. and Koutinas, A. F. (2006). "Nasal and oral masses in a dog." *Vet Clin Pathol* **35**(1): 115-118.
- Li, H. (2011). "A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data." *Bioinformatics* **27**(21): 2987-2993.
- Li, H. and Durbin, R. (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." *Bioinformatics* **25**(14): 1754-1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009). "The Sequence Alignment/Map format and SAMtools." *Bioinformatics* **25**(16): 2078-2079.
- Liao, K. W., Lin, Z. Y., Pao, H. N., Kam, S. Y., Wang, F. I. and Chu, R. M. (2003). "Identification of canine transmissible venereal tumor cells using in situ polymerase chain reaction and the stable sequence of the long interspersed nuclear element." *J Vet Diagn Invest* **15**(5): 399-406.
- Lightowers, R. N., Chinnery, P. F., Turnbull, D. M. and Howell, N. (1997). "Mammalian mitochondrial genetics: heredity, heteroplasmy and disease." *Trends Genet* **13**(11): 450-455.
- Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., Clamp, M., Chang, J. L., Kulbokas, E. J., Zody, M. C., Mauceli, E., Xie, X., Breen, M., Wayne, R. K., Ostrander, E. A., Ponting, C. P., Galibert, F., Smith, D. R., DeJong, P. J., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C. W., Cook, A., Cuff, J., Daly, M. J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K. P., Parker, H. G., Pollinger, J. P., Searle, S. M., Sutter, N. B., Thomas, R., Webber, C., Baldwin, J., Abebe, A., Abouelleil, A., Aftuck, L., Ait-Zahra, M., Aldredge, T., Allen, N., An, P., Anderson, S., Antoine, C., Arachchi, H., Aslam, A., Ayotte, L., Bachantsang, P., Barry, A., Bayul, T., Benamara, M., Berlin, A., Bessette, D., Blitshteyn, B., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Brown, A., Cahill, P., Calixte, N., Camarata, J., Cheshatsang, Y., Chu, J., Citroen, M., Collymore, A., Cooke, P., Dawoe, T., Daza, R., Decktor, K., DeGray, S., Dhargay, N., Dooley, K., Dooley, K., Dorje, P., Dorjee, K., Dorris, L., Duffey, N., Dupes, A., Egbiremolen, O., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Ferreira, P., Fisher, S., FitzGerald, M., Foley, K., Foley, C., Franke, A., Friedrich, D., Gage, D., Garber, M., Gearin, G., Giannoukos, G., Goode, T., Goyette, A., Graham, J., Grandbois, E., Gyaltzen, K., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Healy, C., Hegarty, R., Honan, T., Horn, A., Houde, N., Hughes, L., Hunnicutt, L., Husby, M., Jester, B., Jones, C., Kamat, A., Kanga, B., Kells, C., Khazanovich, D., Kieu, A. C., Kisner, P., Kumar, M., Lance, K., Landers, T., Lara, M., Lee, W., Leger, J. P., Lennon, N., Leuper, L.,

- LeVine, S., Liu, J., Liu, X., Lokyitsang, Y., Lokyitsang, T., Lui, A., Macdonald, J., Major, J., Marabella, R., Maru, K., Matthews, C., McDonough, S., Mehta, T., Meldrim, J., Melnikov, A., Meneus, L., Mihalev, A., Mihova, T., Miller, K., Mittelman, R., Mlenga, V., Mulrain, L., Munson, G., Navidi, A., Naylor, J., Nguyen, T., Nguyen, N., Nguyen, C., Nguyen, T., Nicol, R., Norbu, N., Norbu, C., Novod, N., Nyima, T., Olandt, P., O'Neill, B., O'Neill, K., Osman, S., Oyono, L., Patti, C., Perrin, D., Phunkhang, P., Pierre, F., Priest, M., Rachupka, A., Raghuraman, S., Rameau, R., Ray, V., Raymond, C., Rege, F., Rise, C., Rogers, J., Rogov, P., Sahalie, J., Settupalli, S., Sharpe, T., Shea, T., Sheehan, M., Sherpa, N., Shi, J., Shih, D., Sloan, J., Smith, C., Sparrow, T., Stalker, J., Stange-Thomann, N., Stavropoulos, S., Stone, C., Stone, S., Sykes, S., Tchuinga, P., Tenzing, P., Tesfaye, S., Thoulutsang, D., Thoulutsang, Y., Topham, K., Topping, I., Tsamla, T., Vassiliev, H., Venkataraman, V., Vo, A., Wangchuk, T., Wangdi, T., Weiland, M., Wilkinson, J., Wilson, A., Yadav, S., Yang, S., Yang, X., Young, G., Yu, Q., Zainoun, J., Zembek, L., Zimmer, A. and Lander, E. S. (2005). "Genome sequence, comparative analysis and haplotype structure of the domestic dog." *Nature* **438**(7069): 803-819.
- Liou, C. W., Lin, T. K., Chen, J. B., Tiao, M. M., Weng, S. W., Chen, S. D., Chuang, Y. C., Chuang, J. H. and Wang, P. W. (2010). "Association between a common mitochondrial DNA D-loop polycytosine variant and alteration of mitochondrial copy number in human peripheral blood cells." *J Med Genet* **47**(11): 723-728.
- Liu, C. C., Wang, Y. S., Lin, C. Y., Chuang, T. F., Liao, K. W., Chi, K. H., Chen, M. F., Chiang, H. C. and Chu, R. M. (2008). "Transient downregulation of monocyte-derived dendritic-cell differentiation, function, and survival during tumoral progression and regression in an in vivo canine model of transmissible venereal tumor." *Cancer Immunol Immunother* **57**(4): 479-491.
- Locke, K. B., Yeh, F. J. and Hooper, P. T. (1975). "Letter: Transmissible venereal tumour in dogs in Australia." *Aust Vet J* **51**(9): 449.
- Lott, M. T., Leipzig, J. N., Derbeneva, O., Xie, H. M., Chalkia, D., Sarmady, M., Procaccio, V. and Wallace, D. C. (2013). "mtDNA variation and analysis using Mitomap and Mitomaster." *Curr Protoc Bioinformatics* **44**: 1 23 21-26.
- Lowe, S. W. and Lin, A. W. (2000). "Apoptosis in cancer." *Carcinogenesis* **21**(3): 485-495.
- Lunt, D. H. and Hyman, B. C. (1997). "Animal mitochondrial DNA recombination." *Nature* **387**(6630): 247.
- Ma, H. and O'Farrell, P. H. (2016). "Selfish drive can trump function when animal mitochondrial genomes compete." *Nat Genet* **48**(7): 798-802.
- Ma, H., Xu, H. and O'Farrell, P. H. (2014). "Transmission of mitochondrial mutations and action of purifying selection in *Drosophila melanogaster*." *Nat Genet* **46**(4): 393-397.
- Mambo, E., Gao, X., Cohen, Y., Guo, Z., Talalay, P. and Sidransky, D. (2003). "Electrophile and oxidant damage of mitochondrial DNA leading to rapid evolution of homoplasmic mutations." *Proc Natl Acad Sci U S A* **100**(4): 1838-1843.
- Manning, P. J. and Martin, P. D. (1970). "Metastasis of canine transmissible venereal tumor to the adenohipophysis." *Pathol Vet* **7**(2): 148-152.
- Marchal, T., Chabanne, L., Kaplanski, C., Rigal, D. and Magnol, J. P. (1997). "Immunophenotype of the canine transmissible venereal tumour." *Vet Immunol Immunopathol* **57**(1-2): 1-11.
- Mardis, E. R. and Wilson, R. K. (2009). "Cancer genome sequencing: a review." *Hum Mol Genet* **18**(R2): R163-168.
- Martin, D. and Rybicki, E. (2000). "RDP: detection of recombination amongst aligned sequences." *Bioinformatics* **16**(6): 562-563.

- Martin, D. P., Murrell, B., Golden, M., Khoosal, A. and Muhire, B. (2015). "RDP4: Detection and analysis of recombination patterns in virus genomes." *Virus Evolution* **1**(1).
- Martincorena, I., Raine, K. M., Gerstung, M., Dawson, K. J., Haase, K., Van Loo, P., Davies, H., Stratton, M. R. and Campbell, P. J. (2017). "Universal Patterns of Selection in Cancer and Somatic Tissues." *Cell* **171**(5): 1029-1041 e1021.
- Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D. C., Fullam, A., Alexandrov, L. B., Tubio, J. M., Stebbings, L., Menzies, A., Widaa, S., Stratton, M. R., Jones, P. H. and Campbell, P. J. (2015). "Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin." *Science* **348**(6237): 880-886.
- Marzuki, S., Noer, A. S., Lertrit, P., Thyagarajan, D., Kapsa, R., Utthanaphol, P. and Byrne, E. (1991). "Normal variants of human mitochondrial DNA and translation products: the building of a reference data base." *Hum Genet* **88**(2): 139-145.
- Mathon, N. F. and Lloyd, A. C. (2001). "Cell senescence and cancer." *Nat Rev Cancer* **1**(3): 203-213.
- Matsui, Y. (1909). "Uber transplantable-sarcomatige Neubildung des Hundes." *Gann* **4**: 123.
- McFarland, C. D., Korolev, K. S., Kryukov, G. V., Sunyaev, S. R. and Mirny, L. A. (2013). "Impact of deleterious passenger mutations on cancer progression." *Proc Natl Acad Sci U S A* **110**(8): 2910-2915.
- McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P. and Cunningham, F. (2010). "Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor." *Bioinformatics* **26**(16): 2069-2070.
- Mello Martins, M. I., Ferreira de Souza, F. and Gobello, C. (2005). "Canine Transmissible Venereal Tumor: Etiology, Pathology, Diagnosis and Treatment." *Recent Adv in Small Anim Repro, Ithaca, NY, International Veterinary Information Service*.
- Metzger, M. J., Reinisch, C., Sherry, J. and Goff, S. P. (2015). "Horizontal transmission of clonal cancer cells causes leukemia in soft-shell clams." *Cell* **161**(2): 255-263.
- Metzger, M. J., Villalba, A., Carballal, M. J., Iglesias, D., Sherry, J., Reinisch, C., Muttaray, A. F., Baldwin, S. A. and Goff, S. P. (2016). "Widespread transmission of independent cancer lineages within multiple bivalve species." *Nature* **534**(7609): 705-709.
- Michaels, G. S., Hauswirth, W. W. and Laipis, P. J. (1982). "Mitochondrial DNA copy number in bovine oocytes and somatic cells." *Dev Biol* **94**(1): 246-251.
- Mikaelian, I., Girard, C. and Ivascu, I. (1998). "Transmissible venereal tumor: A consequence of sex tourism in a dog." *Can Vet J* **39**(9): 591-591.
- Milinovich, G. J., Williams, G. M., Clements, A. C. and Hu, W. (2013). "Internet-based surveillance systems for monitoring emerging infectious diseases." *Lancet Infect Dis*.
- Miller, W. W. (1990). "Ocular metastasis of a Transmissible Venereal Tumor." *Canine Pract* **15**(3): 19-21.
- Misdorp, W. (1996). "Veterinary cancer epidemiology." *Vet Q* **18**(1): 32-36.
- Moore, A. E., Rhoads, C. P. and Southam, C. M. (1957). "Homotransplantation of human cell lines." *Science* **125**(3239): 158-160.
- Moore, P. S. and Chang, Y. (2010). "Why do viruses cause cancer? Highlights of the first century of human tumour virology." *Nat Rev Cancer* **10**(12): 878-889.
- Moraes, C. T., Kenyon, L. and Hao, H. (1999). "Mechanisms of human mitochondrial DNA maintenance: the determining role of primary sequence and length over function." *Mol Biol Cell* **10**(10): 3345-3356.
- Morange, M. (2011). "History of cancer research." *eLS. John Wiley & Sons, Ltd: Chichester*: DOI: 10.1002/9780470015902.a9780470003088.pub9780470015902.

- Moschoi, R., Imbert, V., Nebout, M., Chiche, J., Mary, D., Prebet, T., Saland, E., Castellano, R., Pouyet, L., Collette, Y., Vey, N., Chabannon, C., Recher, C., Sarry, J. E., Alcor, D., Peyron, J. F. and Griessinger, E. (2016). "Protective mitochondrial transfer from bone marrow stromal cells to acute myeloid leukemic cells during chemotherapy." *Blood* **128**(2): 253-264.
- Moulton, J. E. (1990). *Tumours of Domestic Animals*, University of California Press, Berkeley and Los Angeles.
- Mozos, E., Mendez, A., Gomez-Villamandos, J. C., Martin De Las Mulas, J. and Perez, J. (1996). "Immunohistochemical characterization of canine transmissible venereal tumor." *Vet Pathol* **33**(3): 257-263.
- Mukaratirwa, S., Chimonyo, M., Obwolo, M., Gruys, E. and Nederbragt, H. (2004). "Stromal cells and extracellular matrix components in spontaneous canine transmissible venereal tumour at different stages of growth." *Histol Histopathol* **19**(4): 1117-1123.
- Mukherjee, S. (2010). *The emperor of all maladies : A biography of cancer*, New York, Scribner.
- Murchison, E. P. (2008). "Clonally transmissible cancers in dogs and Tasmanian devils." *Oncogene* **27**(suppl 2): S19-30.
- Murchison, E. P., Wedge, D. C., Alexandrov, L. B., Fu, B., Martincorena, I., Ning, Z., Tubio, J. M., Werner, E. I., Allen, J., De Nardi, A. B., Donelan, E. M., Marino, G., Fassati, A., Campbell, P. J., Yang, F., Burt, A., Weiss, R. A. and Stratton, M. R. (2014). "Transmissible [corrected] dog cancer genome reveals the origin and history of an ancient cell lineage." *Science* **343**(6169): 437-440.
- Murgia, C., Pritchard, J. K., Kim, S. Y., Fassati, A. and Weiss, R. A. (2006). "Clonal origin and evolution of a transmissible cancer." *Cell* **126**(3): 477-487.
- Murphree, A. L. and Benedict, W. F. (1984). "Retinoblastoma: clues to human oncogenesis." *Science* **223**(4640): 1028-1033.
- Murray, M., James, Z. H. and Martin, W. B. (1969). "A study of the cytology and karyotype of the canine transmissible venereal tumour." *Res Vet Sci* **10**(6): 565-568.
- Mylonakis, M. E., Saridomichelakis, M. N., Lazaridis, V., Leontides, L. S., Kostoulas, P. and Koutinas, A. F. (2008). "A retrospective study of 61 cases of spontaneous canine epistaxis (1998 to 2001)." *J Small Anim Pract* **49**(4): 191-196.
- Nak, D., Nak, Y., Cangul, I. T. and Tuna, B. (2005). "A Clinico-pathological study on the effect of vincristine on transmissible venereal tumour in dogs." *J Vet Med Physiol Pathol Clin Med* **52**(7): 366-370.
- National Archives. (2017a). "The Dogs Act, 1871." Retrieved 12th August 2013, from <http://www.legislation.gov.uk/ukpga/Vict/34-35/56/introduction/enacted>.
- National Archives. (2017b). "The Dogs Act, 1906." Retrieved 12th August 2013, from <http://www.legislation.gov.uk/ukpga/Edw7/6/32/contents/enacted>.
- Ndiritu, C. G., Mbogwa, S. W. and Sayer, P. D. (1977). "Extragenitally located transmissible venereal tumor in dogs." *Mod Vet Pract* **58**: 940-946.
- Nei, M. and Gojobori, T. (1986). "Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions." *Mol Biol Evol* **3**(5): 418-426.
- New Zealand Government. (2013). "Import Health Standards: Cats and Dogs, CATDOG.GEN, 4th April 2013." Retrieved 20th June 2013, from <http://www.biosecurity.govt.nz/files/ihs/catdog.gen.pdf>.
- News Track India. (2011). "Latitude and Longitude of Countries." Retrieved 11th March 2013, from <http://www.newstrackindia.com/information/worldinfo/Latitude-and-Longitude-of-Countries.html>.
- Nguyen, T. H. (2004). "Mechanisms of metastasis." *Clin Dermatol* **22**(3): 209-216.

- Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L. A., Menzies, A., Martin, S., Leung, K., Chen, L., Leroy, C., Ramakrishna, M., Rance, R., Lau, K. W., Mudie, L. J., Varela, I., McBride, D. J., Bignell, G. R., Cooke, S. L., Shlien, A., Gamble, J., Whitmore, I., Maddison, M., Tarpey, P. S., Davies, H. R., Papaemmanuil, E., Stephens, P. J., McLaren, S., Butler, A. P., Teague, J. W., Jonsson, G., Garber, J. E., Silver, D., Miron, P., Fatima, A., Boyault, S., Langerod, A., Tutt, A., Martens, J. W., Aparicio, S. A., Borg, A., Salomon, A. V., Thomas, G., Borresen-Dale, A. L., Richardson, A. L., Neuberger, M. S., Futreal, P. A., Campbell, P. J., Stratton, M. R. and Breast Cancer Working Group of the International Cancer Genome Consortium (2012a). "Mutational processes molding the genomes of 21 breast cancers." *Cell* **149**(5): 979-993.
- Nik-Zainal, S., Van Loo, P., Wedge, D. C., Alexandrov, L. B., Greenman, C. D., Lau, K. W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M., Shlien, A., Cooke, S. L., Hinton, J., Menzies, A., Stebbings, L. A., Leroy, C., Jia, M., Rance, R., Mudie, L. J., Gamble, S. J., Stephens, P. J., McLaren, S., Tarpey, P. S., Papaemmanuil, E., Davies, H. R., Varela, I., McBride, D. J., Bignell, G. R., Leung, K., Butler, A. P., Teague, J. W., Martin, S., Jonsson, G., Mariani, O., Boyault, S., Miron, P., Fatima, A., Langerod, A., Aparicio, S. A., Tutt, A., Sieuwerts, A. M., Borg, A., Thomas, G., Salomon, A. V., Richardson, A. L., Borresen-Dale, A. L., Futreal, P. A., Stratton, M. R., Campbell, P. J. and Breast Cancer Working Group of the International Cancer Genome Consortium (2012b). "The life history of 21 breast cancers." *Cell* **149**(5): 994-1007.
- Nowell, P. C. (1976). "The clonal evolution of tumor cell populations." *Science* **194**(4260): 23-28.
- Nowinsky, M. A. (1876). "Zur Frage über die Impfung der krebsigen Geschwülste." *Zbl Med Wissensch* **14**: 790-791.
- Ohta, S. (2006). "Contribution of somatic mutations in the mitochondrial genome to the development of cancer and tolerance against anticancer drugs." *Oncogene* **25**(34): 4768-4776.
- Okumura, N., Ishiguro, N., Nakano, M., Matsui, A. and Sahara, M. (1996). "Intra- and interbreed genetic variations of mitochondrial DNA major non-coding regions in Japanese native dog breeds (*Canis familiaris*)." *Anim Genet* **27**(6): 397-405.
- Ortega-Pacheco, A., Segura-Correa, J. C., Jimenez-Coello, M. and Linde Forsberg, C. (2007). "Reproductive patterns and reproductive pathologies of stray bitches in the tropics." *Theriogenology* **67**(2): 382-390.
- Oshimura, M., Sasaki, M. and Makino, S. (1973). "Chromosomal banding patterns in primary and transplanted venereal tumors of the dog." *J Natl Cancer Inst* **51**(4): 1197-1203.
- Osipov, N. E. and Golubeva, V. A. (1976). "Diagnosis and treatment of transmissible sarcoma of dogs." *Veterinaria* **7**: 97-98.
- Ostrander, E. A. and Wayne, R. K. (2005). "The canine genome." *Genome Res* **15**(12): 1706-1716.
- Ostrow, S. L., Barshir, R., DeGregori, J., Yeger-Lotem, E. and Hershberg, R. (2014). "Cancer evolution is associated with pervasive positive selection on globally expressed genes." *PLoS Genet* **10**(3): e1004239.
- Pal, S. K. (2011). "Mating system of free-ranging dogs (*Canis familiaris*)." *International Journal of Zoology* **2011**: doi:10.1155/2011/314216.
- Pang, J. F., Kluetsch, C., Zou, X. J., Zhang, A. B., Luo, L. Y., Angleby, H., Ardalán, A., Ekstrom, C., Skollermo, A., Lundeborg, J., Matsumura, S., Leitner, T., Zhang, Y. P. and Savolainen, P. (2009). "mtDNA data indicate a single origin for dogs south of

- Yangtze River, less than 16,300 years ago, from numerous wolves." Mol Biol Evol **26**(12): 2849-2864.
- Park, M. S., Kim, Y., Kang, M. S., Oh, S. Y., Cho, D. Y., Shin, N. S. and Kim, D. Y. (2006). "Disseminated transmissible venereal tumor in a dog." J Vet Diagn Invest **18**(1): 130-133.
- Parker, H. G., Kim, L. V., Sutter, N. B., Carlson, S., Lorentzen, T. D., Malek, T. B., Johnson, G. S., DeFrance, H. B., Ostrander, E. A. and Kruglyak, L. (2004). "Genetic structure of the purebred domestic dog." Science **304**(5674): 1160-1164.
- Parker, H. G., Shearin, A. L. and Ostrander, E. A. (2010). "Man's best friend becomes biology's best in show: genome analyses in the domestic dog." Annu Rev Genet **44**: 309-336.
- Parker, R. C., Varmus, H. E. and Bishop, J. M. (1984). "Expression of v-src and chicken c-src in rat cells demonstrates qualitative differences between pp60v-src and pp60c-src." Cell **37**(1): 131-139.
- Parsonnet, J. (1995). "Bacterial infection as a cause of cancer." Environ Health Perspect **103**(suppl 8): 263-268.
- Pasquier, J., Guerrouahen, B. S., Al Thawadi, H., Ghiabi, P., Maleki, M., Abu-Kaoud, N., Jacob, A., Mirshahi, M., Galas, L., Rafii, S., Le Foll, F. and Rafii, A. (2013). "Preferential transfer of mitochondria from endothelial to cancer cells through tunneling nanotubes modulates chemoresistance." J Transl Med **11**: 94.
- Patel, D., Rorbach, J., Downes, K., Szukszto, M. J., Pekalski, M. L. and Minczuk, M. (2017). "Macropinocytic entry of isolated mitochondria in epidermal growth factor-activated human osteosarcoma cells." Sci Rep **7**(1): 12886.
- Pearse, A. M. and Swift, K. (2006). "Allograft theory: transmission of devil facial-tumour disease." Nature **439**(7076): 549.
- Pereira, J. S., Silva, A. B., Martins, A. L., Ferreira, A. M. and Brooks, D. E. (2000). "Immunohistochemical characterization of intraocular metastasis of a canine transmissible venereal tumor." Vet Ophthalmol **3**(1): 43-47.
- Pereira, L., Van Asch, B. and Amorim, A. (2004). "Standardisation of nomenclature for dog mtDNA D-loop: a prerequisite for launching a Canis familiaris database." Forensic Sci Int **141**(2-3): 99-108.
- Perez, J., Bautista, M. J., Carrasco, L., Gomez-Villamandos, J. C. and Mozos, E. (1994). "Primary Extragenital Occurrence of Transmissible Venereal Tumors: Three Case Reports." Canine Pract **19**(1): 7-10.
- Polyak, K., Li, Y., Zhu, H., Lengauer, C., Willson, J. K., Markowitz, S. D., Trush, M. A., Kinzler, K. W. and Vogelstein, B. (1998). "Somatic mutations of the mitochondrial genome in human colorectal tumours." Nat Genet **20**(3): 291-293.
- Posada, D. and Crandall, K. A. (2001). "Evaluation of methods for detecting recombination from DNA sequences: computer simulations." Proc Natl Acad Sci U S A **98**(24): 13757-13762.
- Powell White, C. (1902). "Contagious growths in dog." Br Med J **2**: 176-177.
- Powers, R. D. (1968). "Immunologic properties of canine transmissible venereal sarcoma." Am J Vet Res **29**(8): 1637-1645.
- Prier, J. E. and Johnson, J. H. (1964). "Malignancy in a Canine Transmissible Venereal Tumor." J Am Vet Med Assoc **145**: 1092-1094.
- Purohit, G. N. (2009). "Canine Transmissible Venereal Tumor: A Review." IJ Vet Med **6**: DOI: 10.5580/a5586a.
- Pye, R. J., Pemberton, D., Tovar, C., Tubio, J. M., Dun, K. A., Fox, S., Darby, J., Hayes, D., Knowles, G. W., Kreiss, A., Siddle, H. V., Swift, K., Lyons, A. B., Murchison, E. P. and Woods, G. M. (2016). "A second transmissible cancer in Tasmanian devils." Proc Natl Acad Sci U S A **113**(2): 374-379.

- Quail, D. F. and Joyce, J. A. (2013). "Microenvironmental regulation of tumor progression and metastasis." *Nat Med* **19**(11): 1423-1437.
- R Core Team (2013). "R: A language and environment for statistical computing." *R Foundation for Statistical Computing*: <http://www.R-project.org/>.
- Ramadinha, R. R., Teixeira, R. S., Bomfim, P. C., Mascarenhas, M. B., França, T. N., Peixoto, T. C., Costa, S. Z. R. and Peixoto, P. V. (2016). "Response of canine transmissible venereal tumor to vincristine sulfate and vinblastine sulfate chemotherapy." *Rev Bras Med Vet* **38**: 65-69.
- Rebbeck, C. A., Leroi, A. M. and Burt, A. (2011). "Mitochondrial capture by a transmissible cancer." *Science* **331**(6015): 303.
- Rebbeck, C. A., Thomas, R., Breen, M., Leroi, A. M. and Burt, A. (2009). "Origins and evolution of a transmissible cancer." *Evolution* **63**(9): 2340-2349.
- Richardson, V. B., Sapp, W. J. and Adams, E. W. (1987). "The distribution of C-bands in canine transmissible venereal tumor cells." *Cornell Vet* **77**(2): 161-167.
- Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. and Mesirov, J. P. (2011). "Integrative genomics viewer." *Nat Biotechnol* **29**(1): 24-26.
- Rooney, J. P., Ryde, I. T., Sanders, L. H., Howlett, E. H., Colton, M. D., Germ, K. E., Mayer, G. D., Greenamyre, J. T. and Meyer, J. N. (2015). "PCR based determination of mitochondrial DNA copy number in multiple species." *Methods Mol Biol* **1241**: 23-38.
- Rosenberg, A. S. and Singer, A. (1992). "Cellular basis of skin allograft rejection: an in vivo model of immune-mediated tissue destruction." *Annu Rev Immunol* **10**: 333-358.
- Rottcher, D. (1972). "Clinical features and pathology of transmissible venereal tumours in dogs in Kenya." *Tierarztl Umsch* **27**: 235-238.
- Rowley, J. D. (1973). "Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining." *Nature* **243**(5405): 290-293.
- Rust, J. H. (1949). "Transmissible lymphosarcoma in the dog." *J Am Vet Med Assoc* **114**(862): 10-14.
- Rustom, A., Saffrich, R., Markovic, I., Walther, P. and Gerdes, H. H. (2004). "Nanotubular highways for intercellular organelle transport." *Science* **303**(5660): 1007-1010.
- Said, R. A., Silva, L. F., Albuquerque, A. R. O. L., Sousa-Neta, E. M. and O., L. M. (2009). "Efficacy and side effects of Vincristine sulphate treatment on canine transmissible venereal tumour." *Proceedings of the 34th World Small Animal Veterinary Congress WSAVA, 2009, Sao Paulo.*: Retrieved 10th October 2013, <http://www.ivis.org/proceedings/wsava/2009/lecture2030/2014.pdf?LA=2011>.
- Sala-Torra, O., Hanna, C., Loken, M. R., Flowers, M. E., Maris, M., Ladne, P. A., Mason, J. R., Senitzer, D., Rodriguez, R., Forman, S. J., Deeg, H. J. and Radich, J. P. (2006). "Evidence of donor-derived hematologic malignancies after hematopoietic stem cell transplantation." *Biol Blood Marrow Transplant* **12**(5): 511-517.
- Samuels, D. C., Li, C., Li, B., Song, Z., Torstenson, E., Boyd Clay, H., Rokas, A., Thornton-Wells, T. A., Moore, J. H., Hughes, T. M., Hoffman, R. D., Haines, J. L., Murdock, D. G., Mortlock, D. P. and Williams, S. M. (2013). "Recurrent tissue-specific mtDNA mutations are common in humans." *PLoS Genet* **9**(11): e1003929.
- Sandusky, G. E., Carlton, W. W. and Wightman, K. A. (1987). "Diagnostic immunohistochemistry of canine round cell tumors." *Vet Pathol* **24**(6): 495-499.

- Sastry, G. A., Narayana, J. V., Rao, P. R. and Christopher, J. (1965). "A case of metastatic venereal tumour in a bitch." Indian Vet J **42**(9): 658-660.
- Sato, M. and Sato, K. (2013). "Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA." Biochim Biophys Acta **1833**(8): 1979-1984.
- Savolainen, P., Arvestad, L. and Lundeberg, J. (2000). "A novel method for forensic DNA investigations: repeat-type sequence analysis of tandemly repeated mtDNA in domestic dogs." J Forensic Sci **45**(5): 990-999.
- Savolainen, P., Rosen, B., Holmberg, A., Leitner, T., Uhlen, M. and Lundeberg, J. (1997). "Sequence analysis of domestic dog mitochondrial DNA for forensic use." J Forensic Sci **42**(4): 593-600.
- Savolainen, P., Zhang, Y. P., Luo, J., Lundeberg, J. and Leitner, T. (2002). "Genetic evidence for an East Asian origin of domestic dogs." Science **298**(5598): 1610-1613.
- Schon, E. A., DiMauro, S. and Hirano, M. (2012). "Human mitochondrial DNA: roles of inherited and somatic mutations." Nat Rev Genet **13**(12): 878-890.
- Seligmann, C. G. (1906). "On the occurrence of new growths among the natives of British New Guinea." Royal Society of Medicine, Pathological Section.
- Shannon, L. M., Boyko, R. H., Castelhamo, M., Corey, E., Hayward, J. J., McLean, C., White, M. E., Abi Said, M., Anita, B. A., Bondjengo, N. I., Calero, J., Galov, A., Hedimbi, M., Imam, B., Khalap, R., Lally, D., Masta, A., Oliveira, K. C., Perez, L., Randall, J., Tam, N. M., Trujillo-Cornejo, F. J., Valeriano, C., Sutter, N. B., Todhunter, R. J., Bustamante, C. D. and Boyko, A. R. (2015). "Genetic structure in village dogs reveals a Central Asian domestication origin." Proc Natl Acad Sci U S A **112**(44): 13639-13644.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D. and Higgins, D. G. (2011). "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." Mol Syst Biol **7**: 539.
- Singh, J., Rana, J. S., Sood, N., Pangawkar, G. R. and Gupta, P. P. (1996). "Clinico-pathological studies on the effect of different anti-neoplastic chemotherapy regimens on transmissible venereal tumours in dogs." Vet Res Commun **20**(1): 71-81.
- Skinner, M. E., Uzilov, A. V., Stein, L. D., Mungall, C. J. and Holmes, I. H. (2009). "JBrowse: a next-generation genome browser." Genome Res **19**(9): 1630-1638.
- Slater, G. S. and Birney, E. (2005). "Automated generation of heuristics for biological sequence comparison." BMC Bioinformatics **6**: 31.
- Smith, G. B. and Washbourn, J. W. (1897). "Pathological society of London. Meeting held on 6th April. 1897." The Lancet **10**: 1025-1026.
- Smith, G. B. and Washbourn, J. W. (1898a). "Infective sarcoma in dogs." Br Med J (Clin Res Ed) **2**: 1346-1347.
- Smith, G. B. and Washbourn, J. W. (1898b). "Infective venereal tumours in dogs." J Pathol Bacteriol **5**: 99-110.
- Smith, J. M. (1992). "Analyzing the mosaic structure of genes." J Mol Evol **34**(2): 126-129.
- Sobral, R. A. (1998). "Occurrence of canine transmissible venereal tumor in dogs from the Jaboticabal region, Brazil." Ars Veterinaria **14**(1): 1-10.
- Sofuni, T. and Makino, S. (1963). "A supplementary study on the chromosomes of venereal tumors of the dog." Gann **54**: 149-154.
- Spees, J. L., Olson, S. D., Whitney, M. J. and Prockop, D. J. (2006). "Mitochondrial transfer between cells can rescue aerobic respiration." Proc Natl Acad Sci U S A **103**(5): 1283-1288.
- Stamatakis, A. (2014). "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies." Bioinformatics **30**(9): 1312-1313.

- Stewart, J. B., Alaei-Mahabadi, B., Sabarinathan, R., Samuelsson, T., Gorodkin, J., Gustafsson, C. M. and Larsson, E. (2015). "Simultaneous DNA and RNA mapping of somatic mitochondrial mutations across diverse human cancers." PLoS Genet **11**(6): e1005333.
- Stewart, J. B. and Chinnery, P. F. (2015). "The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease." Nat Rev Genet **16**(9): 530-542.
- Sticker, A. (1902). "Ueber den Krebs der Thiere." Arch Klin Chir **65**: 1023-1087.
- Sticker, A. (1904). "Transplantables Lyphosarkom des Hundes." Z Krebsforsch **1**: 413-444.
- Sticker, A. (1905). "Transplantables Lymphosarkoma des Hundes." Berl Tierarztl Wschr **21**: 353-358.
- Sticker, A. (1906a). "Transplantables Rundezellensarkom des Hundes Ein Beitrag zer Lehre der Krebsübertragbarkeit." Z Krebsforsch **4**: 227-314.
- Sticker, A. (1906b). "Übertragung von Tumoren bei Hunden durch den Geschlechtsakt." Berl Tier Woschr **50**: 894-995.
- Sticker, A. (1907). "Endemischer Krebs." Z Krebsforsch **5**(2): 215-224.
- Stimmelmayer, R. (2010). "Transmissible Venereal Tumor: A Special Canine Case of Contagious Cancer Cells." Vet Times.
- Stookey, J. L. (1969). "Transmissible venereal tumors of dogs." Natl Cancer Inst Monogr **32**: 315-320.
- Strakova, A. (2013). "The global distribution and epidemiological features of Canine Transmissible Venereal Tumour." Part II undergraduate project report, Univesity of Cambridge.
- Strakova, A. and Murchison, E. P. (2014). "The changing global distribution and prevalence of canine transmissible venereal tumour." BMC Vet Res **10**(1): 168.
- Strakova, A. and Murchison, E. P. (2015). "The cancer which survived: insights from the genome of an 11000 year-old cancer." Curr Opin Genet Dev **30**: 49-55.
- Strakova, A., Ni Leathlobhair, M., Wang, G. D., Yin, T. T., Airikkala-Otter, I., Allen, J. L., Allum, K. M., Banske-Issa, L., Bisson, J. L., Castillo Domracheva, A., de Castro, K. F., Corrigan, A. M., Cran, H. R., Crawford, J. T., Cutter, S. M., Delgadillo Keenan, L., Donelan, E. M., Faramade, I. A., Flores Reynoso, E., Fotopoulou, E., Fruean, S. N., Gallardo-Arrieta, F., Glebova, O., Hafelin Manrique, R. F., Henriques, J. J., Ignatenko, N., Koenig, D., Lanza-Perea, M., Lobetti, R., Lopez Quintana, A. M., Losfelt, T., Marino, G., Martincorena, I., Martinez Castaneda, S., Martinez-Lopez, M. F., Meyer, M., Nakanwagi, B., De Nardi, A. B., Neunzig, W., Nixon, S. J., Onsare, M. M., Ortega-Pacheco, A., Peleteiro, M. C., Pye, R. J., Reece, J. F., Rojas Gutierrez, J., Sadia, H., Schmeling, S. K., Shamanova, O., Ssuna, R. K., Steenland-Smit, A. E., Svitich, A., Thoya Ngoka, I., Vitalaru, B. A., de Vos, A. P., de Vos, J. P., Walkinton, O., Wedge, D. C., Wehrle-Martinez, A. S., van der Wel, M. G., Widdowson, S. A. and Murchison, E. P. (2016). "Mitochondrial genetic diversity, selection and recombination in a canine transmissible cancer." Elife **5**: 10.7554/eLife.14552.
- Stratton, M. R. (2011). "Exploring the genomes of cancer cells: progress and promise." Science **331**(6024): 1553-1558.
- Stratton, M. R., Campbell, P. J. and Futreal, P. A. (2009). "The cancer genome." Nature **458**(7239): 719-724.
- Strauss, D. C. and Thomas, J. M. (2010). "Transmission of donor melanoma by organ transplantation." Lancet Oncol **11**: 790-796.
- Stubbs, E. L. and Furth, J. (1934). "Experimental studies on venereal sarcoma of the dog." Am J Pathol **10**(2): 275-286.
- Tan, A. S., Baty, J. W., Dong, L. F., Bezawork-Geleta, A., Endaya, B., Goodwin, J., Bajzikova, M., Kovarova, J., Peterka, M., Yan, B., Pesdar, E. A., Sobol, M., Filimonenko, A., Stuart, S., Vondrusova, M., Kluckova, K., Sachaphibulkij, K., Rohlena, J., Hozak,

- P., Truksa, J., Eccles, D., Haupt, L. M., Griffiths, L. R., Neuzil, J. and Berridge, M. V. (2015). "Mitochondrial genome acquisition restores respiratory function and tumorigenic potential of cancer cells without mitochondrial DNA." Cell Metab **21**(1): 81-94.
- Tarafder, A. and Samad, M. A. (2010). "Prevalence of clinical diseases of pet dogs and risk perception of zoonotic infection by dog owners in Bangladesh." Bangl J Vet Med **8**(2): 163-174.
- Taylor Bennett, B., Taylor, Y. and Epstein, R. (1975). "Segregation of the clinical course of transmissible venereal tumor with DL-A haplotypes in canine families." Transplant Proc **7**(4): 503-505.
- Taylor, R. W. and Turnbull, D. M. (2005). "Mitochondrial DNA mutations in human disease." Nat Rev Genet **6**(5): 389-402.
- Thacher, C. and Bradley, R. L. (1983). "Vulvar and vaginal tumors in the dog: a retrospective study." J Am Vet Med Assoc **183**(6): 690-692.
- Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V. J., Sawyer, S. K., Greenfield, D. L., Germonpre, M. B., Sablin, M. V., Lopez-Giraldez, F., Domingo-Roura, X., Napierala, H., Uerpmann, H. P., Loponte, D. M., Acosta, A. A., Giemisch, L., Schmitz, R. W., Worthington, B., Buikstra, J. E., Druzhkova, A., Graphodatsky, A. S., Ovodov, N. D., Wahlberg, N., Freedman, A. H., Schweizer, R. M., Koepfli, K. P., Leonard, J. A., Meyer, M., Krause, J., Paabo, S., Green, R. E. and Wayne, R. K. (2013). "Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs." Science **342**(6160): 871-874.
- ThermoFisher Scientific. (2008). "PicoGreen manual." Retrieved 31st July 2015, from <http://tools.thermofisher.com/content/sfs/manuals/mp07581.pdf>.
- Thomas, R., Rebbeck, C., Leroy, A. M., Burt, A. and Breen, M. (2009). "Extensive conservation of genomic imbalances in canine transmissible venereal tumors (CTVT) detected by microarray-based CGH analysis." Chromosome Res **17**(7): 927-934.
- Thomas, R., Scott, A., Langford, C. F., Fosmire, S. P., Jubala, C. M., Lorentzen, T. D., Hitte, C., Karlsson, E. K., Kirkness, E., Ostrander, E. A., Galibert, F., Lindblad-Toh, K., Modiano, J. F. and Breen, M. (2005). "Construction of a 2-Mb resolution BAC microarray for CGH analysis of canine tumors." Genome Res **15**(12): 1831-1837.
- Thorburn, M. J., Gwynn, R. V., Ragbeer, M. S. and Lee, B. I. (1968). "Pathological and cytogenetic observations on the naturally occurring canine venereal tumour in Jamaica (Sticker's tumour)." Br J Cancer **22**(4): 720-727.
- Thorvaldsdottir, H., Robinson, J. T. and Mesirov, J. P. (2013). "Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration." Brief Bioinform **14**(2): 178-192.
- Thyagarajan, B., Padua, R. A. and Campbell, C. (1996). "Mammalian mitochondria possess homologous DNA recombination activity." J Biol Chem **271**(44): 27536-27543.
- Tolar, J. and Neglia, J. P. (2003). "Transplacental and other routes of cancer transmission between individuals." J Pediatr Hematol Oncol **25**(6): 430-434.
- Torralba, D., Baixauli, F. and Sanchez-Madrid, F. (2016). "Mitochondria know no boundaries: mechanisms and functions of intercellular mitochondrial transfer." Front Cell Dev Biol **4**: 107.
- Tuppen, H. A., Blakely, E. L., Turnbull, D. M. and Taylor, R. W. (2010). "Mitochondrial DNA mutations and human disease." Biochim Biophys Acta **1797**(2): 113-128.
- Ujvari, B., Dowton, M. and Madsen, T. (2007). "Mitochondrial DNA recombination in a free-ranging Australian lizard." Biol Lett **3**(2): 189-192.

- Varughese, E. E., Singla, V. K., Ratnakaran, U. and Gandotra, V. K. (2012). "Successful management of metastatic transmissible venereal tumour to skin of mammary region." Reprod Domest Anim **47**(suppl 6): 366-369.
- Vazquez-Mota, N., Simon-Martinez, J., Cordova-Alarcon, E., Lagunes, L. and Fajardo, R. (2008). "The T963C mutation of TP53 gene does not participate in the clonal origin of canine TVT." Vet Res Commun **32**(2): 187-191.
- Vennervald, B. J. and Polman, K. (2009). "Helminths and malignancy." Parasite Immunol **31**(11): 686-696.
- Vermooten, M. I. (1987). "Canine transmissible venereal tumor (TVT): a review." J S Afr Vet Assoc **58**(3): 147-150.
- Verscheure, S., Backeljau, T. and Desmyter, S. (2015). "In silico discovery of a nearly complete mitochondrial genome Numt in the dog (*Canis lupus familiaris*) nuclear genome." Genetica **143**(4): 453-458.
- Vila, C., Savolainen, P., Maldonado, J. E., Amorim, I. R., Rice, J. E., Honeycutt, R. L., Crandall, K. A., Lundeberg, J. and Wayne, R. K. (1997). "Multiple and ancient origins of the domestic dog." Science **276**(5319): 1687-1689.
- Vila, C., Seddon, J. and Ellegren, H. (2005). "Genes of domestic mammals augmented by backcrossing with wild ancestors." Trends Genet **21**(4): 214-218.
- von Bergmann, E. (1895). "Gelungene Carcinomübertragung beim Hunde." Centralblatt für Chirurgie **27**: XXIV. Kongress.
- Vonholdt, B. M., Pollinger, J. P., Lohmueller, K. E., Han, E., Parker, H. G., Quignon, P., Degenhardt, J. D., Boyko, A. R., Earl, D. A., Auton, A., Reynolds, A., Bryc, K., Brisbin, A., Knowles, J. C., Mosher, D. S., Spady, T. C., Elkahoulou, A., Geffen, E., Pilot, M., Jedrzejewski, W., Greco, C., Randi, E., Bannasch, D., Wilton, A., Shearman, J., Musiani, M., Cargill, M., Jones, P. G., Qian, Z., Huang, W., Ding, Z. L., Zhang, Y. P., Bustamante, C. D., Ostrander, E. A., Novembre, J. and Wayne, R. K. (2010). "Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication." Nature **464**(7290): 898-902.
- Wade, H. (1908). "An experimental investigation of infective sarcoma of the dog, with a consideration of its relationship to cancer." J Pathol Bacteriol **12**: 384-425.
- Walberg, M. W. and Clayton, D. A. (1981). "Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA." Nucleic Acids Res **9**(20): 5411-5421.
- Wallace, D. C. (2012). "Mitochondria and cancer." Nat Rev Cancer **12**(10): 685-698.
- Wang, G. D., Zhai, W., Yang, H. C., Wang, L., Zhong, L., Liu, Y. H., Fan, R. X., Yin, T. T., Zhu, C. L., Poyarkov, A. D., Irwin, D. M., Hytonen, M. K., Lohi, H., Wu, C. I., Savolainen, P. and Zhang, Y. P. (2016). "Out of southern East Asia: the natural history of domestic dogs across the world." Cell Res **26**(1): 21-33.
- Wang, Y. and Bogenhagen, D. F. (2006). "Human mitochondrial DNA nucleoids are linked to protein folding machinery and metabolic enzymes at the mitochondrial inner membrane." J Biol Chem **281**(35): 25791-25802.
- Warburg, O. (1956a). "On respiratory impairment in cancer cells." Science **124**(3215): 269-270.
- Warburg, O. (1956b). "On the origin of cancer cells." Science **123**(3191): 309-314.
- Watson, J. D. and Crick, F. H. (1953). "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid." Nature **171**(4356): 737-738.
- Wayne, R. K. (1986a). "Cranial morphology of domestic and wild canids: the influence of development on morphological change." Evolution **40**(2): 243-261.
- Wayne, R. K. (1986b). "Limb morphology of domestic and wild canids: the influence of development on morphologic change." J Morphol **187**(3): 301-319.
- Wayne, R. K. (1993). "Molecular evolution of the dog family." Trends Genet **9**(6): 218-224.

- Wayne, R. K., Nash, W. G. and O'Brien, S. J. (1987a). "Chromosomal evolution of the Canidae. I. Species with high diploid numbers." *Cytogenet Cell Genet* **44**(2-3): 123-133.
- Wayne, R. K., Nash, W. G. and O'Brien, S. J. (1987b). "Chromosomal evolution of the Canidae. II. Divergence from the primitive carnivore karyotype." *Cytogenet Cell Genet* **44**(2-3): 134-141.
- Wayne, R. K., Van Valkenburgh, B., Kat, P. W., Fuller, T. K., Johnson, W. E. and O'Brien, S. J. (1989). "Genetic and morphological divergence among sympatric canids." *J Hered* **80**(6): 447-454.
- Weber, W. T., Nowell, P. C. and Hare, W. C. (1965). "Chromosome studies of a transplanted and a primary canine venereal sarcoma." *J Natl Cancer Inst* **35**(3): 537-547.
- Wehr, V. (1888). "Demonstration der durch Impfung von Hund auf Hund erzeugten Carcinomknotchen." *Zentralbl Chir* **15**: 8-9.
- Wehr, V. (1889). "Weiter Mittheilungen über die positiven Ergebnisse der Carcinom-Ueberimpfungen von Hund auf Hund." *Archiv Klin Chir* **39**(226-228): 226.
- Weir, E. C., Pond, M. J., Duncan, J. R. and Polzin, D. J. (1978). "Extragenital occurrence of transmissible venereal tumor in the dog: literature review and case reports." *J Am Anim Hosp Assoc* **14**: 532-536.
- Westermann, B. (2010). "Mitochondrial fusion and fission in cell life and death." *Nat Rev Mol Cell Biol* **11**(12): 872-884.
- Wheeler, D. A. and Wang, L. (2013). "From human genome to cancer genome: the first decade." *Genome Res* **23**(7): 1054-1062.
- Wiesner, R. J., Ruegg, J. C. and Morano, I. (1992). "Counting target molecules by exponential polymerase chain reaction: copy number of mitochondrial DNA in rat tissues." *Biochem Biophys Res Commun* **183**(2): 553-559.
- World Atlas. (2012). "Countries listed by continent." Retrieved 10th October 2012, from <http://www.worldatlas.com/cntycont.htm#.UlbgNL5wbmQ>.
- World Bank. (2013). "Country and Lending Groups, 2013." Retrieved 12th March 2013, from <http://data.worldbank.org/about/country-classifications/country-and-lending-groups#Low>.
- Wright, D. H., Peel, S., Cooper, E. H. and Hughes, D. T. (1970). "Transmissible venereal sarcoma of dogs. A histochemical and chromosomal analysis of tumours in Uganda." *Rev Eur Etud Clin Biol* **15**(2): 155-160.
- Yang, T. J. (1987). "Metastatic transmissible venereal sarcoma in a dog." *J Am Vet Med Assoc* **190**(5): 555-556.
- Yang, T. J., Chandler, J. P. and Dunne-Anway, S. (1987). "Growth stage dependent expression of MHC antigens on the canine transmissible venereal sarcoma." *Br J Cancer* **55**(2): 131-134.
- Yang, T. J. and Jones, J. B. (1973). "Canine transmissible venereal sarcoma: transplantation studies in neonatal and adult dogs." *J Natl Cancer Inst* **51**(6): 1915-1918.
- Yuan, Y., Ju, Y. S., Kim, Y., Li, J., Wang, Y., Yang, Y., Martincorena, I., Creighton, C., Weinstein, J. N., Xu, Y., Han, L., Kim, H.-L., Nakagawa, H., Park, K., Campbell, P. J. and Liang, H. (2017). "Comprehensive Molecular Characterization of Mitochondrial Genomes in Human Cancers." [bioRxiv 161356: https://doi.org/10.1101/161356](https://doi.org/10.1101/161356).
- Zhang, J., Liu, H., Luo, S., Lu, Z., Chavez-Badiola, A., Liu, Z., Yang, M., Merhi, Z., Silber, S. J., Munne, S., Konstantinidis, M., Wells, D., Tan, J. J. and Huang, T. (2017). "Live birth derived from oocyte spindle transfer to prevent mitochondrial disease." *Reprod Biomed Online* **34**(4): 361-368.
- Zong, W. X., Rabinowitz, J. D. and White, E. (2016). "Mitochondria and Cancer." *Mol Cell* **61**(5): 667-676.
- Zsurka, G., Kraytsberg, Y., Kudina, T., Kornblum, C., Elger, C. E., Khrapko, K. and Kunz, W. S. (2005). "Recombination of mitochondrial DNA in skeletal muscle of

individuals with multiple mitochondrial DNA heteroplasmy." Nat Genet
37(8): 873-877.

APPENDIX

APPENDIX: CHAPTER 2

Additional file 2-1 Questionnaire on canine transmissible venereal tumour worldwide distribution.

PROJECT ON CANINE TRANSMISSIBLE VENEREAL TUMOUR WORLDWIDE DISTRIBUTION
Cancer Genome Project, Department of Veterinary Medicine, University of Cambridge

CTVT (also known as Sticker's Sarcoma, Infective Venereal Tumour, TVT, Venereal Sarcoma) is a common disease of dogs and is found worldwide. Primary tumours occur on the external genitalia of both sexes and the disease is usually transmitted during mating.



The aim of this research project is to collect data on worldwide prevalence of CTVT, which will help us to identify CTVT risk factors and therefore new possibilities for CTVT prevention. In addition to completing the questionnaire, we would like to invite you to participate in research collaboration by collecting samples and have the opportunity to win a trip to Cambridge. The collection of samples will be used for a genetics study of transmissible cancers and the evolutionary processes which have permitted their emergence and spread worldwide.

All the data will be kept confidential. Your contribution to this project is much appreciated!

YOUR INFORMATION

Name:

Email address (to keep you updated with project):

Country:

Town/village:

Type of organisation:

- Private veterinary clinic
- Veterinary school
- Pathology laboratory
- Animal welfare organisation
- Other, please specify:

Number of dogs in your area:

- Fewer dogs than humans
- One dog per human
- 2-5 dogs per human
- > 5 dogs per human

Most dogs in your area are:

- Spayed/Neutered
- Entire

If you have extra information, please, comment:

Is there a noticeable population of free roaming dogs in your area?

- Yes
- No

If you have extra information, please, comment:

CTVT QUESTIONNAIRE

1) Have you ever seen a case of CTVT?

- Yes
- No

2) What is your expert opinion on prevalence (% of affected dogs) of CTVT in your region?

- None
- Less than 0.5%
- 0.5-1%
- 1 - 3%
- 3 - 5%
- 5 - 10%
- 10 - 20%
- More than 20%

If you have more precise data, please comment:

3) How many CTVT case per year do you see?

4) Most of my CTVT cases are from

- Urban area
- Rural area
- Semirural area
- Unknown

5) Have you observed CTVT in other regions of your country or the world?

6) Is it noticeably absent from any region in your country?

7) How often does metastasis (lesions in more than one place) occur in the cases you have observed?

- 0-5%
- 5 -10%
- 10 -15%
- 15 -20%
- More than 20%

What other parts of the body are affected (e.g. skin, lymph nodes,...)?

8) Have you ever observed spontaneous regression of tumours (without treatment)?

- Yes
- No

If yes, please provide details:

9) Treatment used (tick all that apply)

- Vincristine
- Doxorubicine
- Radiotherapy
- Surgery
- No treatment
- Other

If other, please specify:

10) What proportion of tumours goes into complete remission after using the selected treatment?

- 0-20%
- 20-40%
- 40-60%
- 60-80%
- 80-100%

11) Gender of the majority of affected dogs

- Male
- Female
- Equally represented

Comment on percentage of predominant sex:

12) Most of dogs with CTVT are:

- Otherwise healthy
- Diseased/carrying parasites

If diseased, what is the most frequent disease/parasite in dogs with CTVT?

- Injured
- Thin, emaciated

What is the most common disease/parasite in your practice?

13) Comment on any unusual cases of CTVT that you have seen:

14) Have you ever come across CTVT in other species (e.g. fox, wolf, coyote, jackal)?

15) Any other comments:

16) Have you got any PHOTOGRAPHS you could send me?(My email address: ctvt@sanger.ac.uk)

17) Is there a national veterinary organisation in your country?

Yes

No

If yes, what is its name/contact?

18) WOULD YOU LIKE TO TAKE PART IN RESEARCH COLLABORATION? (All materials required for sampling as well as postage are paid for by us!)

No

Yes

If yes, please be sure to provide your name and email address at the top of the questionnaire!

Additional file 2-2 Contemporary and historical reports providing a numerical value for CTVT prevalence.

Year	Location	Prevalence (%)	Details	Author	Title	Reference
1938	France (Toulouse)	1.8%	103 CTVT cases in 5643 dogs	R. Lasserre	Recherches sur le Cancer des animaux domestiques	Rev Med Vet 90, 425-451, 1938
1968	Jamaica (Kingston)	0.8%	8 CTVT cases in 1000 dogs	M. J. Thorburn	Pathological and cytogenetic observations on the naturally occurring Canine Venereal Tumour in Jamaica (Sticker's tumour)	Brit J Cancer 22, 4, 720-727, 1968
1969	Nigeria (Lagos)	11.9%	38 CTVT cases out of 318 dogs presented at the clinic	G. O. Esuruoso	Observations in An experimental Veterinary Clinic in the Ikeja Airport Area of Lagos	Nigerian Vet J 1, 7-15, 1972
1972	Kenya (Kabete)	1.0%	Reported prevalence in the population	D. Rottcher	Clinical features and pathology of transmissible venereal tumours in dogs in Kenya	Tierarztl Umschau 27, 235-238, 1972
1974	Kenya (Kabete)	3.7%	55 CTVT cases in 1498 post mortem and biopsy cases	A. Kimeto	Transmissible Venereal Tumour of Dog in Kenya	B Anim Health Prod Afr 22, 327-329, 1974
1985	Papua New Guinea (Boroko)	16.0%	28 CTVT cases in 170 necropsied dogs (between 1982-83)	A.N. Hamir	Primary penile and nasal transmissible venereal tumours in a dog	Aust Vet J, 62, 12, 430-432, 1985
1986	Papua New Guinea (Boroko)	19.8%	53 CTVT cases 268 post mortem examinations	A. N. Hamir	Neoplasms of dogs in Papua New Guinea	Aust Vet J 63, 10, 342-343, 1986
1998	Brazil (Jaboticabal)	1.9%	400 cases seen in a 12 year period	R. A. Sobral	Occurrence of canine transmissible venereal tumor in dogs from the Jaboticabal region, Brazil	Ars Veterinaria 14, 1, 1-10, 1998
2003	Mexico (Merida - Yucatan)	2.6%	8 CTVT cases in 303 dogs	A. Ortega-Pacheco	Prevalence of transmissible venereal tumor of stray dogs in Merida, Yucatan, Mexico	Rev Biomed 14, 83-87, 2003
2004	Brazil (Botucatu)	9.9%	576 CTVT cases out of 5798 cytological exams	A. S. Amaral	Cytological diagnosis of transmissible venereal tumor in the Botucatu region, Brazil (descriptive study: 1994-2003)	RPCV 99, 167-171, 2004

2004	Peru (Lima)	8.4%	78 CTVT cases in 925 dogs examined	N. Mendoza	Frequency of the Transmissible Venereal Tumour in Dogs: Caseload in the Pathology Laboratory of the National University of San Marcos (Period 1998-2004)	Rev Inv Vet Peru 21, 1, 42-47, 2010
2006	Mexico (Merida - Yucatan)	5.4%	318 post mortem examinations of male stray dogs	A. Ortega-Pacheco	Pathological Conditions of the Reproductive Organs of Male Stray Dogs in the Tropics: Prevalence, Risk Factors, Morphological Findings and Testosterone Concentrations	Reprod Dom Anim 41, 429-437, 2006
2007	Mexico (Merida - Yucatan)	15.3%	300 post mortem examinations of female stray dogs	A. Ortega-Pacheco	Reproductive patterns and reproductive pathologies of stray bitches in the tropics	Theriogenology 67, 382-390, 2007
2009	India (Nagpur)	1.3%	73 CTVT cases in 5877 dogs examined	L. A. Khan	Incidence of Venereal Granuloma and its Medicinal Treatment in stray Dogs of Nagpur City	Vet World 2, 1, 13-14, 2009
2010	Mexico (Mexico City)	17.5%	717 dogs examined	J. C. Cruz	Canine Transmissible Venereal Tumor in the Metropolitan Area of Mexico City	Revista Cientifica 20, 4, 362-366, 2010
2010	Bangladesh (Dhaka)	0.5%	20 CTVT cases recorded out of 3670 sick pet dogs presented at the hospital	M. Tarafder	Prevalence of clinical diseases of pet dogs and risk perception of zoonotic infection by dog owners in Bangladesh	Bangl J Vet Med 8, 2, 163-174, 2010
2011	India (Chennai)	2.7% in males	18 CTVT cases in 668 dogs examined	T. Sathiamoorthy	Prevalence of reproductive disorders in the stray dogs of Chennai City	JIVA 9, 2, 62-63, 2011
2011	India (Chennai)	3.9% in females	24 CTVT cases in 615 dogs examined	T. Sathiamoorthy	Prevalence of reproductive disorders in the stray dogs of Chennai City	JIVA 9, 2, 62-63, 2011
2013	Ecuador (Otavalo)	2.3%	5 CTVT cases out of 216 sterilized dogs	A. Strakova and E. P. Murchison	Personal experience	

Additional file 2-3 Summary of CTVT prevalence data by country.

Country	Higher confidence average CTVT prevalence (3 or more responses)	Lower confidence average CTVT prevalence (1-2 responses)	Number of responses
Afghanistan	N/A	No data	1
Albania	N/A	N/A	0
Algeria	N/A	1-3%	1
American Samoa	N/A	3-5%	2
Andorra	N/A	N/A	0
Angola	N/A	N/A	0
Antigua and Barbuda	N/A	N/A	0
Argentina	5-10%	N/A	3
Armenia	N/A	1-3%	1
Australia	Less than 0.5%	N/A	18
Austria	N/A	None	2
Azerbaijan	N/A	N/A	0
Bahamas	N/A	0.5-1%	2
Bahrain	N/A	N/A	0
Bangladesh	1-3%	N/A	4
Barbados	N/A	0.5-1%	1
Belarus	0.5-1%	N/A	9
Belgium	N/A	None	2
Belize	10-20%	N/A	6
Benin	N/A	N/A	0
Bhutan	N/A	N/A	0
Bolivia	N/A	N/A	0
Bosnia and Herzegovina	N/A	N/A	0
Botswana	5-10%	N/A	4
Brazil	1-3%	N/A	10
Brunei	N/A	N/A	0
Bulgaria	N/A	0.5-1%	1
Burkina	N/A	N/A	0
Burma (Myanmar)	N/A	N/A	0
Burundi	N/A	N/A	0
Cambodia	N/A	5-10%	1
Cameroon	N/A	N/A	0
Canada	None	N/A	3
Cape Verde	N/A	N/A	0
Central African Republic	N/A	N/A	0
Chad	N/A	N/A	0

Chile	3-5%	N/A	11
China	Less than 0.5%	N/A	20
Colombia	N/A	1-3%	2
Comoros	N/A	N/A	0
Congo	N/A	N/A	0
Congo, Democratic Republic of	N/A	N/A	0
Costa Rica	3-5%	N/A	3
Croatia	N/A	N/A	0
Cuba	3-5%	N/A	3
Cyprus	N/A	No data	1
Czech Republic	None	N/A	4
Denmark	N/A	None	1
Djibouti	N/A	N/A	0
Dominica	N/A	3-5%	1
Dominican Republic	1-3%	N/A	3
East Timor (Timor-Leste)	N/A	3-5%	1
Ecuador	3-5%	N/A	4
Egypt	N/A	1-3%	1
El Salvador	N/A	N/A	0
Equatorial Guinea	N/A	N/A	0
Eritrea	N/A	N/A	0
Estonia	Less than 0.5%	N/A	3
Ethiopia	N/A	N/A	0
Fiji	N/A	3-5%	2
Finland	None	N/A	3
France	Less than 0.5%	N/A	4
Gabon	N/A	N/A	0
Gambia	N/A	10-20%	1
Georgia	N/A	N/A	0
Germany	N/A	None	1
Ghana	5-10%	N/A	3
Greece	1-3%	N/A	4
Grenada	N/A	10-20%	2
Guatemala	N/A	N/A	0
Guinea	N/A	N/A	0
Guinea-Bissau	N/A	N/A	0
Guyana	N/A	3-5%	1
Haiti	N/A	N/A	0
Honduras	N/A	5-10%	1
Hungary	N/A	Less than 0.5%	2

Iceland	N/A	N/A	0
India	1-3%	N/A	7
Indonesia	N/A	N/A	0
Iran	N/A	1-3%	1
Iraq	N/A	N/A	0
Ireland	N/A	N/A	0
Israel	Less than 0.5%	N/A	41
Italy	Less than 0.5%	N/A	23
Ivory Coast	N/A	10-20%	2
Jamaica	N/A	N/A	0
Japan	Less than 0.5%	N/A	3
Jordan	N/A	N/A	0
Kazakhstan	N/A	Less than 0.5%	1
Kenya	3-5%	N/A	7
Kiribati	N/A	N/A	0
Korea, North	N/A	N/A	0
Korea, South	N/A	N/A	0
Kuwait	N/A	None	2
Kyrgyzstan	N/A	0.5-1%	1
Laos	N/A	N/A	0
Latvia	N/A	N/A	0
Lebanon	N/A	N/A	0
Lesotho	N/A	More than 20%	1
Liberia	N/A	1-3%	1
Libya	N/A	1-3%	1
Liechtenstein	N/A	N/A	0
Lithuania	N/A	N/A	0
Luxembourg	N/A	N/A	0
Macedonia	5-10%	N/A	3
Madagascar	N/A	N/A	0
Malawi	3-5%	N/A	3
Malaysia	3-5%	N/A	5
Maldives	N/A	N/A	0
Mali	1-3%	N/A	3
Malta	N/A	None	1
Marshall Islands	N/A	N/A	0
Mauritania	N/A	N/A	0
Mauritius	3-5%	N/A	3
Mexico	5-10%	N/A	5
Micronesia	N/A	3-5%	1
Moldova	N/A	N/A	0

Monaco	N/A	N/A	0
Mongolia	N/A	N/A	0
Montenegro	N/A	N/A	0
Morocco	N/A	More than 20%	1
Mozambique	5-10%	N/A	6
Namibia	N/A	1-3%	1
Nauru	N/A	N/A	0
Nepal	N/A	N/A	0
Netherlands	None	N/A	3
New Zealand	None	N/A	3
Nicaragua	5-10%	N/A	3
Niger	N/A	N/A	0
Nigeria	1-3%	N/A	8
Norway	None	N/A	3
Oman	N/A	N/A	0
Pakistan	1-3%	N/A	3
Palau	N/A	N/A	0
Panama	1-3%	N/A	3
Papua New Guinea	N/A	N/A	0
Paraguay	3-5%	N/A	3
Peru	3-5%	N/A	4
Philippines	1-3%	N/A	4
Poland	N/A	N/A	0
Portugal	Less than 0.5%	N/A	4
Puerto Rico	N/A	0.5-1%	1
Qatar	N/A	N/A	0
Reunion	Less than 0.5%	N/A	4
Romania	5-10%	N/A	3
Russian Federation	1-3%	N/A	102
Rwanda	N/A	N/A	0
Saint Kitts and Nevis	N/A	1-3%	2
Saint Lucia	N/A	N/A	0
Saint Vincent and the Grenadines	N/A	1-3%	1
Samoa	5-10%	N/A	3
San Marino	N/A	N/A	0
Sao Tome and Principe	N/A	N/A	0
Saudi Arabia	N/A	N/A	0
Senegal	N/A	0.5-1%	1
Serbia	N/A	N/A	0
Seychelles	N/A	N/A	0
Sierra Leone	N/A	N/A	0

Singapore	N/A	Less than 0.5%	2
Slovakia	N/A	N/A	0
Slovenia	N/A	None	2
Solomon Islands	N/A	0.5-1%	2
Somalia	N/A	N/A	0
South Africa	0.5-1%	N/A	23
South Sudan	N/A	N/A	0
Spain	Less than 0.5%	N/A	14
Sri Lanka	N/A	3-5%	2
Sudan	N/A	N/A	0
Suriname	0.5-1%	N/A	4
Swaziland	N/A	1-3%	2
Sweden	None	N/A	5
Switzerland	None	N/A	4
Syria	N/A	N/A	0
Taiwan	N/A	0.5-1%	1
Tajikistan	N/A	N/A	0
Tanzania	Less than 0.5%	N/A	3
Thailand	5-10%	N/A	6
Togo	N/A	N/A	0
Tonga	N/A	N/A	0
Trinidad and Tobago	N/A	Less than 0.5%	1
Tunisia	N/A	0.5-1%	1
Turkey	3-5%	N/A	3
Turkmenistan	N/A	N/A	0
Tuvalu	N/A	N/A	0
Uganda	1-3%	N/A	4
Ukraine	1-3%	N/A	91
United Arab Emirates	N/A	N/A	0
United Kingdom	None	N/A	15
United States	Less than 0.5%	N/A	35
Uruguay	N/A	1-3%	1
Uzbekistan	N/A	N/A	0
Vanuatu	Less than 0.5%	N/A	3
Vatican City	N/A	N/A	0
Venezuela, Bolivarian Republic of	N/A	5-10%	2
Viet Nam	1-3%	N/A	3
Yemen	N/A	N/A	0
Zambia	5-10%	N/A	3
Zimbabwe	N/A	0.5-1%	2

Additional file 2-4 Contemporary and historical reports of naturally occurring CTVT cases in the published literature. Reports are ordered by date. Multiple reports from the same publication are indicated on separate lines. Only reports referring to primary naturally occurring cases are included, while articles reporting only cases of experimental transplantation are excluded.

Year	Location	First Author	Title	Reference
1810	UK (London)	D. P. Blaine	A domestic treatise on the diseases of horses and dogs	Published 1810, 4th edition, 161-2, 1810
1876	Russia (St.Petersburg)	M. A. Nowinsky	Zur Frage über die Impfung der krebsigen Geschwulste	Zbl Med Wissench 14, 790-791, 1876
1888	Germany (Lemberg)	Wehr	Demonstration der durch Impfung von Hund auf Hund erzeugten Carcinomknoten	Zbl Chir 15 (suppl. to no. 24), 8-9, 1888
1889	Germany (Lemberg)	Wehr	Weiter Mittheilungen über die positiven Ergebnisse der Carcinom-Ueberimpfungen von Hund auf Hund	Arch Klin Chir 39, 226-228, 1889
1894	France (Paris)	S. Duplay	Tumeurs expérimentales chez les animaux	Atti dell'XI Congresso Medico Internazionale, Roma, 2, 103-104, 1894
1895	Germany (Berlin, Göttingen)	E. von Bergmann	Gelungene Carcinomübertragung beim Hunde	Zbl Chir 27, XXIV. Kongress, 1895
1895	Poland (Wrocław)	E. von Bergmann	Gelungene Carcinomübertragung beim Hunde	Zbl Chir 27, XXIV. Kongress, 1895
1897	UK (London)	G. B. Smith	Infective venereal tumours in dogs	Transactions Pathological Society London, 48, 310-323, 1897
1897	UK (London)	Pathological society of London	Meeting held on 6th April 1897	The Lancet, April 10, 1025-1026, 1897
1898	UK (London)	G. B. Smith	Infective venereal tumours in dogs	J Pathol Bacteriol 5, 99-110, 1898
1900	UK (London)	F. Hobday	Operations on the genital organs	In Canine and Feline Surgery, W.&A.K. Johnston, 1900
1902	UK (Leeds)	C. Powell White	Contagious growths in dogs	Br Med J July 19, 2, 176-177, 1902
1904	Germany (Frankfurt)	A. Sticker	Transplantables Lyphosarkom des Hundes	Z Krebsforsch 1, 413-444, 1904
1904	Papua New Guinea (Port Moresby)	C. G. Seligmann	On the occurrence of new growths among the natives of British New Guinea	Third Scientific Report Invest Imp Cancer Res Fund, 26-40, 1908
1905	UK (London)	F. Hobday	Observations on Contagious Venereal Tumours in Canine Patients	Vet J 2 (new series), 342-346, 1905
1905	UK (London)	E.F. Bashford	Comparison between the transmission of an infective granuloma of the dog and carcinoma of the mouse	Scientific report, Imp Cancer Res Fund 2, 33-37, 1905

1906	Canada (Montreal) - imported from the UK	C. French	Surgical diseases and surgery of the dog	Published by Washington D.C. pp. 254-256, 286-287 and 365-367
1906	UK (London)	F. Hobday	Tumours on the Penis and Prepuce, Malignant Tumours of the Vagina	Surgical diseases of the dog and cat and anaesthetics, second edition, Bailliere, Tindall and Cox, 8 Henrietta Street, 1906
1906	Germany (Berlin)	A. Sticker	Übertragung von Tumoren bei Hunden durch den Geschlechtsakt	Berl Tier Wochschr 50, 894-995, 1906
1906	US (New York City)	S.P. Beebe	A study of the so called infectious lymphosarcoma of dogs	J Med Res 15, 209-227, 1906
1907	US (New York City)	S.P. Beebe	The growth of lymphosarcoma in dogs- summary of recent observations	J Am Med Assoc 49, 18, 1492-1493, 1907
1907	Germany (Berlin)	P. Bergell (A. Sticker)	Ueber Pathogenese und über den spezifischen Abbau der Krebsgeschwulste	Deut Med Wochenschr 2, 38, 1521-1522, 1907
1907	France (Paris)	M.A. Borrel	Lymphosarcome du chien	Sem Med (Paris) 27, 94-95, 1907
1907	France	M. Cadéac	Infection sarcomateuse aigue du chien	J Med Vet Zootech 58, 266-269, 1907
1907	UK (Edinburgh, London)	H. Wade	An experimental investigation of infective sarcoma of the dog, with a consideration of its relationship to cancer	J Pathol 12, 384-425, 1907
1907	Germany (Berlin)	A. Sticker	Endemischer Krebs	Z Krebsforsch 5, 2, 215-224, 1907
1908	Sri Lanka	C. G. Seligmann	On the occurrence of new growths among the natives of British New Guinea	Third Scientific Report Invest Imp Cancer Res Fund, 26-40, 1908
1909	Japan	Y. Matsui	Über transplantable-sarcomatige Neubildung des Hundes	Gann 4, 123, 1909
1921	US (New York City)	W. L. Williams	the Venereal Tumors of the Dog. Venereal Granulomata. Lympho-sarcoma.	In The Diseases of the Genital Organs of Domestic Animals, published by Ithaca, N.Y., 1921
1927	Japan	S. Matsuba	Studies on venereal tumor in dog	J Jap Soc Vet Sc 6, 167, 1927
1929	US (Minnesota, Rochester)	W. H. Feldman	So called infectious sarcoma of the dog in an unusual anatomic situation	Am J Path 5, 183-195, 1929
1931	Germany (Berlin)	H. Auler	Über Tumoren des Hundes	Z Krebsforsch 35, 1-11, 1931
1932	US (Minnesota, Rochester)	W. H. Feldman	Transmissible Lymphosarcoma of Dogs	In Neoplasms of Domesticated Animals, Chp 20, 343-56, Philadelphia, WB Saunders Co., 1932

1932	UK	W. H. Feldman	Transmissible Lymphosarcoma of Dogs	In Neoplasms of Domesticated Animals, Chp 20, 343-56, Philadelphia, WB Saunders Co., 1932
1932	China (Beijing)	A. I. H. Wong	Radium treatment in the so-called infectious sarcoma of dogs	Chin Med J-Peking 46, 377-382, 1932
1933	US (Tennessee, Nashville)	W. A. DeMonbreun	An experimental investigation concerning the nature of contagious lymphosarcoma of dogs	Am J Cancer 21, 295-321, 1933
1933	US (Pennsylvania, Philadelphia)	E. L. Stubbs	Experimental studies on venereal sarcoma of the dog	Am J Pathol 10, 275-286, 1933
1934	France (Lyon, Collonges, Mureils)	R. Bauduin	Les Papillomes du vagin chez la chienne	Thesis presented at La Faculte de Medicine et de Pharmacie de Lyon, 30th January 1934, Editors: BOSC Freres, M&L RIOU, 1934
1938	France (Toulouse)	R. Lasserre	Recherches sur le Cancer des animaux domestiques	Rev Med Vet 90, 425-451, 1938
1938	Italy (Bologna)	F. Cella	Spleno-reticolo-sarcoma e sarcoma di Sticker in un cane	La Nuova Veterinaria 304-309, 1938
1939	Italy (Messina)	P. Ajello	Osservazioni sul tumore genitale contagioso del cane	Thesis presented at Facolta do medicina veterinaria della R. Universita di Messina, 1939
1939	New Guinea	P. Ajello	Osservazioni sul tumore genitale contagioso del cane	Thesis presented at Facolta do medicina veterinaria della R. Universita di Messina, 1939
1939	France (Paris)	L. Bory	Le lympho-sarcome infectieux des Chiens	Clinique et Pathologie comparee, Paris, Masson, 23-25, 1939
1939	Italy (Bologna)	F. Cella	Sopra un secondo caso di granuloma venereo del cane (c.d. sarcoma di sticker) osservato nell'Emilia - metastasi splenica	La Nuova Veterinaria 12, 217-220, 1939
1944	South Africa (Pretoria)	C. Jackson	The Cytology of the Contagious (Venereal) Tumour of the Dog	Onderstepoort J Vet Sci and Anim Indus 20, 97-118, 1944
1946	France	G. Lesbouyries	Néoplasies vénériennes du Chien et de la Chienne	Recueil Med Vet 122, 1, 7-19, 1946
1947	Italy (Sardinia)	A. Carta	Indagini sperimentali sul cosi detto sarcoma di sticker	Profilassi 20, 22-23, 1947
1947	Ireland (Waterford)	L. N. Gleeson	An Account of Infectious Venereal Granuloma in the Greyhound	Vet Rec 59, 411-413, 1947
1949	Puerto Rico (San Juan)	J. H. Rust	Transmissible Lymphosarcoma in the Dog	JAVMA 114, 10-14, 1949

1949	France (Toulouse)	F. Gouaud	Contribution a l'etude du sarcome de Sticker: therapeutique par les rayons X	Thesis presented at the Veterinary School in Toulouse in 1949
1949	France (Toulouse)	P. Lacour	Contribution a l'etude du sarcome de sticker: etude clinique, cytologique et therapeutique basee sur 66 observations originales: cenception	Thesis presented at the Veterinary School in Toulouse in 1949
1949	France (Toulouse)	M. Neveu	Contribution a l'etude du 'Sarcome de Sticker' essai de therapeutique, deductions histo-physiologiques	Thesis presented at the Veterinary School in Toulouse, 1949
1950	France (Lyon)	P.Collet	Essais de traitement des tumeurs vénériennes du chien par la podophylline (sarcome de Sticker)	Bull Soc Sci Vet Lyon, 52, 115-122, 1950
1950	France (Toulouse)	I. Nanta	Les tumeurs veneriennes du chien. Cadre nosologique.	Rev Vet 101, 298-320, 356-366, 1950
1950	France (Toulouse)	P. Genty	Contribution a l'etude du Sarcome de Sticker. Epidemiologie de l'affection	Thesis presented at the Veterinary School in Toulouse, 1950
1950	France (Toulouse)	J. P. Salacroup	Contribution a l'etude du sarcome de sticker: consideration etiologique de l'affection	Thesis presented at the Veterinary School in Toulouse, 1950
1951	US (New York)	F. Bloom	The Transmissible Venereal Tumour of the Dog. Studies Indicating That the Tumor cells are Mature End Cells of Reticulo-endothelial Origin	Am J Path 27, 119-139, 1951
1952	Mali (previously French Sudan) (Bamako)	Z. Derbal	Traitement chirurgical des tumeurs vénériennes de la Chienne	Rec Med Vet 128, 26-29, 1952
1952	US (Minnesota, Rochester)	A. G. Karlson	The Transmissible Venereal Tumor of Dogs: Observations on Forty Generations of Experimental Transfers	Ann NY Acad Sci 54, 1197-1213, 1952
1953	Senegal (Dakar)	R. Jean	Action de sérum de Bogomoletz sur un sarcome de Sticker, avec métastases malignes mammaires, chez une chienne	Rev Vet Milit, 8, 93-96, 1953
1953	France (Toulouse)	J. Cadeillan	Contribution a l'etude des tumeurs veneriennes d u chien	Thesis presented at the Veterinary School in Toulouse in 1953
1954	UK (London)	E. Cotchin	Neoplasia in the dog	Vet Rec, December 25th, Twelfth Congress paper, 879-885, 1954
1954	US (New York City and Philadelphia)	F. Bloom	Pathology of the dog and cat, the genito-urinary system, with clinical considerations	Commonwealth Bureau of Animal Health, 275-280, 1954

1954	France (Toulouse)	H. Parot	Contribution a l'etude des tumeurs veneriennes du chien (formes atypiques)	Thesis presented at the Veterinary School in Toulouse, 1954
1956	Italy (Bologna)	L. Bignozzi	L'attuale valutazione istopatologica del tumore di Sticker	Veterinaria Milano, Universita di Bologna 5, 96-101, 1956
1957	Poland (Wroclaw)	J. Utzig	Wplyw trojterpenow zswartych w zagwi brzozoweg Pliporus betulinus na guzy Stickera	Med Weter 8, 481-484, 1957
1958	Japan (Tokyo)	M. Hataya	Effects of x-ray irradiation on the transmissible venereal tumor of the dog	Gann 49, 307-318, 1958
1958	Japan	Y. Shirasu	Studies on the Transmissible Venereal Tumor of the Dog. Serial Transplantation into the Cheek Pouch of Hamsters.	Gann 49 (suppl.), 205-206, 1958
1958	Japan (Hokkaido)	S. Takayama	Existence of a stem-cell lineage in an infectious venereal tumor of the dog,	Jpn J Genet 33, 56-64, 1958
1959	US (Missouri, St.Louis)	P. V. Belkin	Extragenital Venereal Granuloma in the Abdominal Organs of a Dog	JAVMA 135, 575-576, 1959
1960	Italy (Messina)	A. Bonaduce	Alcune Ricerche sul Tumore di Sticker	Zooprofilass 15, 937-957, 1960
1960	Italy (Messina)	P. Ajello	Trasmissione del Tumore di Sticker con Materiale Acelurare	La Nuova Veterinaria 36, 179-183, 1960
1960	Italy (Perugia)	G. Dozza	Anticorpi Emoagglutinanti in Cani Portatori del Cosiddetto Tumore di Sticker	Atti Soc Ital Sc Vet, 14, p 531-535, 1960
1960	Brazil (Belo Horizonte)	L. Lazzer	Tratamiento quirurgico de tumor veneréo en perro	Rev Mil Vet 13, 140, 1960
1960	France (Maisons-Alfort)	F. Legneau	Essai de transmission d'une néoplasie vénérienne du chien (sarcome de Sticker)	Bull Societe Francaise Dermatologie Syph 67, 542-546, 1960
1961	Italy (Messina)	P. Ajello	Sulla Presenza di Inclusioni Nelle Cellule del Tumore di Sticker	Bull Societa Italiana Biolo Sperim 37, 247-249, 1961
1961	Japan (Hokkaido - Sapporo, Otaru))	S. Takayama	Cytological Studies of Tumors - A Study of Chromosomes in Venereal Tumors of the Dog	Z Krebsforsch 64, 253-261, 1961
1962	South Africa (Cape Town)	G. de Kock	Pathological studies on neoplasms of dogs in South Africa	Onderstepoort J Vet res 29, 1, 35-54, 1962
1963	US (Washington D.C.)	C. N. Barron	Intraocular tumours in Animals. V.Transmissible Venereal Tumor of Dogs	Am J Vet Res 24, 1263-1270, 1963
1963	France (Paris)	J-P. Carteaude	Contribution a l'étude cytologique de la tumeur vénérienne du chien (sarcome de Sticker)	Réunion de Paris, Séance du 14 November 1963, Societe de dermatologie et de Syphiligraphie, 1963

1963	Japan (Hyogo, Sapporo, Otaru, Osaka)	S. Makino	Some epidemiologic aspects of venereal tumors of dogs as revealed by chromosome and DNA studies	Ann NY Acad Sci 108, 1106-1122, 1963
1963	Japan (Sapporo)	T. Sofuni	A supplementary study on the chromosomes of venereal tumors of the dog	Gann 54, 149-154, 1963
1964	US (Pennsylvan ia, Philadelphia)	J. E. Prier	Malignancy in a Canine Transmissible Venereal Tumor	JAVMA 145, 11, 1092- 1094, 1964
1965	India	G. A. Sastry	A case of metastatic venereal tumor in a bitch	Ind Vet J 42, 658-659, 1965
1965	US (Pennsylvan ia, Philadelphia)	W. T. Weber	Chromosome Studies of a Transplanted and a Primary Canine Venereal Sarcoma	J Natl Cancer I 35, 3, 537-547, 1965
1966	US (Alabama, Tuskegee Institute)	E. W. Adams	Canine Venereal Tumor - serum protein electrophoresis, transaminase, and lactic dehydrogenase activity	Cornell Vet 57, 572-578, 1966
1966	Bahamas	D. A. Higgins	Observations on the Canine Transmissible Venereal Tumour as seen in the Bahamas	Vet Rec 79, 67-71, 1966
1966	US (Alabama, Tuskegee Institute)	E. W. Adams	Growth and Maintenance of the Canine Venereal Tumor in Continuous Culture	Cancer Research 28, 753-757, 1966
1966	France (Paris)	G. Barski	Cytogenetic Study of Sticker Venereal Sarcoma in European Dogs	J Natl Cancer I 37, 6, 787-797, 1966
1966	US (Pennsylvan ia, Philadelphia)	J. M. McKenna	Some Immunological Aspects of Canine Neoplasms	Cancer Res 26, 137-142, 1966
1967	US (Pennsylvan ia, Philadelphia)	R. S. Brodey	Neoplasms of the canine Uterus, Vagina and Vulva: A Clinicopathologic Survey of 90 cases	JAVMA 151, 1294-1307, 1967
1967	France (Toulouse)	C. Lombard	Considérations sur la nature et recherches sur l'ultrastructure du sarcome de Sticker du chien	B Cancer 54, 3, 357-365, 1967
1967	Japan (Hokkaido)	T. Koike	Successively transplanted canine transmissible sarcoma	Gann 70, 115-118, 1979
1967	Uganda (Kampala)	O. Bwangamoi	Tumours of Domestic Animals in Uganda	Vet Rec 81, 525, 1967
1968	Russia (Moscow)	E. S. Kakpakova	Peculiarities of the karyotype of the transmissible sarcoma cells in the dog	V Opr Onkol, 14, 43-50, 1968

1968	Jamaica (Kingston)	M. J. Thorburn	Pathological and cytogenetic observations on the naturally occurring Canine Venereal Tumour in Jamaica (Sticker's tumour)	Brit J Cancer 22, 4, 720-727, 1968
1968	US (Alabama, Decatur)	R. D. Powers	Immunologic Properties of Canine Transmissible Venereal Sarcoma	Am J Vet Res 29, 8, 1637-1645, 1968
1968	India (Trichur)	P. J. Philip	Treatment of Venereal Sarcoma in a Bitch by Vulvo-Vagino-Ovario-Hysterectomy with Perineal Urethrostomy	Ind Vet J 45, 874-877, 1968
1969	Chile (Santiago)	W. Drommer	Vergleichende licht -und elektronenmikroskopische Untersuchungen am ubertragbaren venerischen Sarkom und Histiozytom des Hundes	Path Vet 6, 273-286, 1969
1969	Germany (Hannover)	W. Drommer	Vergleichende licht -und elektronenmikroskopische Untersuchungen am ubertragbaren venerischen Sarkom und Histiozytom des Hundes	Path Vet 6, 273-287, 1969
1969	Kenya (Nairobi)	M. Murray	A Study of the Cytology and Karyotype of the Canine Transmissible Venereal Tumour	Res Vet Sci 10, 565-568, 1969
1969	UK (London)	O.F. Jackson	Transmissible Venereal Tumour in Dogs	Vet Rec 84, 125, 1969
1969	Ireland (Dublin)	T. D. Grimes	Transmissible Venereal Tumour in Dogs	Vet Rec 84, 124, 1969
1969	UK (Liverpool)	J. C. Howell	Transmissible Venereal Tumour of Dogs	Vet Rec 84, 418-819, 1969
1969	Zimbabwe	J.B.Tutt	Transmissible Venereal Tumour in a Boxer Bitch	Vet Rec 84, 13, 1969
1970	US (Alabama, Tuskegee Institute)	E. W. Adams	A Canine Venereal Tumor with Metastasis to the Brain	Path Vet 7, 498-502, 1970
1970	US (North Carolina, Winston- Salem)	P. J. Manning	Metastasis of Canine Transmissible Venereal Tumor to the Adenohypophysis	Path Vet 7, 148-152, 1970
1970	Hungary (Budapest)	M. Sellyei	Neue Angaben zur chromosomalen Struktur des Sticker-Sarkoms	Z Krebsforsch 74, 7-14, 1970
1970	Uganda (Kampala)	D. H. Wright	Transmissible Venereal Sarcoma of Dogs. A Histochemical and Chromosomal Analysis of Tumours in Uganda.	Rev Europ Etudes Clin Biol 15, 155-160, 1970
1970	Malaysia	D. H. Wright	Transmissible Venereal Sarcoma of Dogs. A Histochemical and Chromosomal Analysis of Tumours in Uganda.	Rev Europ Etudes Clin Biol 15, 155-161, 1970

1971	US (Illinois, Chicago)	R. B. Epstein	Histocompatibility Typing and Course of Canine Venereal Tumors Transplanted into Unmodified Random Dogs	Cancer Res, 34, 788-794, 1974
1972	Japan (Hokkaido, Ryukyu Islands)	C. G. McLeod	Transmissible Venereal Tumor with Metastases in Three Dogs	JAVMA 161, 2, 199-200, 1972
1972	Malaysia	D. Cohen	Thymidine labelling studies in a Transmissible Venereal Tumour of the dog	Br J Cancer, 26, 413-419, 1972
1972	Japan (Sapporo)	M. Oshimura	Chromosomal Banding Patterns in Primary and Transplanted Venereal Tumors of the Dog	J Natl Cancer I 51, 4, 1197-1203, 1973, 1972
1972	Nigeria (Lagos)	G. O. Esuruoso	Observations in An experimental Veterinary Clinic in the Ikeja Airport Area of Lagos	Nigerian Vet J 1, 7-15, 1972
1972	Kenya	D. Rottcher	Clinical features and pathology of transmissible venereal tumours in dogs in Kenya	Tierarztl Umschau 27, 235-238, 1972
1973	France (Toulouse)	P. Cabanie	Étude Ultrastructurale du Sarcome de Sticker du Chien a Différents Stades de son Évolution	Revue Med Vet, 124, 10, 1239-1253, 1973
1973	Mexico (Mexico City)	P. Hernández-Jáuregui	Ultrastructural and Histochemical Pattern of Regressing Canine Venereal Lymphoma After Cyclophosphamide Treatment	J Natl Cancer Inst, 51, 1187-1196, 1973
1973	US (Tennessee, Knoxville)	T. J. Yang	Canine Transmissible Venereal Sarcoma: Transplantation Studies in Neonatal and Adult Dogs	J Natl Cancer I 51, 1915-1918, 1973
1973	Malaysia	D. Cohen	The Biological Behavior of the Transmissible Venereal Tumor in Immunosupressed Dogs	European J Cancer 9, 253-258, 1973
1973	Nigeria (Ibadan)	O. O. Oduye	Metastatic transmissible venereal tumor in dogs	J Small Anim Pract 14, 625-637, 1973
1974	Kenya (Kabete)	A. Kimeto	Transmissible Venereal Tumour of Dog in Kenya	B Anim Health Prod Afr 22, 327-329, 1974
1974	US (Missouri, Columbia)	J. J. Broadhurst	Neoplasms of the reproductive system	Published in R.W. Kirk (ed.), Current Veterinary Therapy, W.B. Saunders, Philadelphia, 928-937, 1974
1974	Malaysia	R. B. Epstein	Histocompatibility Typing and Course of Canine Venereal Tumors Transplanted into Unmodified Random Dogs	Cancer Res 34, 788-793, 1974
1974	Italy (Perugia)	M. Battistacci	Ricerche ultrastrutturali sul Sarcoma di Sticker	Nuova Vet 50, 226-236, 1974

1975	US (Texas)	J.M. Cockrill	Ultrastructural Characteristics of Canine Transmissible Venereal Tumor at Various Stages of Growth and Regression	Am J Vet Res, 36, 5, 577-681, 1975
1975	Australia (Alice Springs)	K. B. Locke	Transmissible Venereal Tumour in Dogs in Australia	Aust Vet J 51, 449, 1975
1975	Sri Lanka	S.G. Wettimuny	Canine neoplasms in Sri Lanka	Ceylon Vet J 23, 1-7, 1975
1975	Italy (Sassari)	P. Muzzetto	Il sarcoma di Sticker nota II. (trattamento chirurgico nella cagna)	Clin Vet 98, 124-128, 1975
1976	Russia (Moscow)	N. E. Osipov	Diagnosis and treatment of transmissible sarcoma of dogs	Veterinaria (Moscow) 7, 97-98, 1976
1976	Italy (Sicily)	C. Murgia	Clonal Origin and Evolution of a Transmissible Cancer	Cell 126, 3, 477-487, 2006
1976	Brazil (Botucatu)	A. C. Alexandrino	Tumor venéreo transmissível em caes na regio de Botucatu	Arq Esc Vet UFMG 28, 1, 101-104, 1976
1976	Croatia (Zagreb)	K. Cermak	Preosivi venericni tumor pasa	Prax Vet 24, 279-287, 1976
1977	Iran (Shiraz)	B. Ivoghli	Canine Transmissible Venereal Tumor in Iran	Vet Pathol 14, 289-290, 1977
1977	India (Izzat Nagar)	G. C. Mohanty	Growth and Morphological Characteristics of Canine Venereal Tumor Cells <i>in vitro</i>	Vet Pathol 14, 420-425, 1977
1977	Kenya (Kabete)	C. G. Ndiritu	Extragenitally located transmissible venereal tumor in dogs	Mod Vet Pract 940-946, 1977
1977	US (Tennessee, Knoxville or Connecticut, Storrs)	J. R. Kennedy	Canine transmissible Venereal Sarcoma: Electron Microscopic Changes With Time After Transplantation	Br J Cancer 36, 375-385, 1977
1977	Nigeria (Ibadan)	L. Idowu	The chromosomes of the transmissible venereal tumour of dogs in Ibadan, Nigeria	Res Vet Sci 22, 271-273, 1977
1977	India (Jabalpur)	S. K. Pandey	Canine transmissible venereal sarcoma: clinical trial with autogenous formalized vaccine	Indian Vet J 54, 852-853, 1977
1977	Poland (Warsaw)	A. Wasecki	Zostosowonie preporotu Vinblastin w leczeniu guzów Sticker	Med Weter 33, 142-143, 1977
1977	France (Maisons-Alfort)	X. Woimant	Neoplasie venerienne du chien et de la chienne (sarcome de Sticker)	Rec Med Vet 153, 331-338, 1977
1978	Israel	D. Cohen	The transmissible venereal tumor of the dog - a naturally occurring allograft? A review.	Isr J Med Sci 14, 1, 14-19, 1978
1978	Israel (Be'er Sheva)	D. Cohen	Two Cases of the Transmissible Venereal Tumor of the Dog in Israel	Refuah Vet 35, 12-13, 1978

1978	Kenya (Kabete)	J. A. Spence	Metastasis of a transmissible venereal tumour to the pituitary	J Small Anim Pract 19,175-184, 1978
1978	US (Georgia, Athens)	E. C. Weir	Extragenital Occurrence of Transmissible Venereal Tumor in the Dog: Literature Review and Case Reports	J Am Anim Hosp Assoc 14, 532-536, 1978
1979	US (Georgia, Athens)	J. R. Duncan	Cytology of Canine Cutaneous Round Cell Tumors	Vet Pathol 16, 673-679, 1979
1980	Italy (Messina, Sicilia, Calabria, Torino)	P. Ajello	Il tumore di Sticker	Annali Facolta Medicina Veterinaria Mess 17, 289-339, 1980
1980	Austria (Vienna)	P. Ajello	Il tumore di Sticker	Annali Facolta Medicina Veterinaria Mess 17, 289-339, 1980
1980	US (Georgia, Athens)	N. O. Brown	Chemotherapeutic Management of Transmissible Venereal Tumours in 30 dogs	JAVMA 176, 983-986, 1980
1980	South Africa (Onderstepoort)	I. B. J. van Rensburg	Extragenital malignant transmissible venereal tumour in a bitch	J S Afr Vet Assoc 51, 3, 199-201, 1980
1980	Nigeria (Nsukka)	J. O. Omamegbe	The surgical treatment of transmissible venereal tumour (TVT) in male dogs - An Assesment of three surgical methods of treatment	Nigerian Vet J 9:-12, 1980
1980	France (Toulouse)	G. Madiot	Traitement medico-chirurgical du sarcome de Sticker	Thesis presented at the Veterinary School in Toulouse, 1980
1981	US (Connecticut, Storrs)	J.P. Chandler	Canine Transmissible Venereal sarcoma: distribution of T and B lymphocytes in blood, draining lymph nodes and tumours at different stages of growth	Br J Cancer 44, 514-520, 1981
1982	US (Georgia, Athens)	C. A. Calvert	Vincristine for treatment of transmissible venereal tumor in the dog: clinical reports	JAVMA 181, 2, 163-164, 1982
1982	India (Andhra Pradesh)	Ch. Choudary	Certain Canine Neoplasms Encountered in Andhra Pradesh	Indian Vet J 59, 100-102, 1982
1982	US (Pennsylvania, Philadelphia)	D. E. Thrall	Orthovoltage radiotherapy of canine transmissible venereal tumors	Vet Radiol 23, 217-219, 1982
1982	Nigeria	E. I. Amber	Canine transmissible venereal tumor: Evaluation of surgical excision of primary and metastatic lesions in Zaria - Nigeria.	J Am Anim Hosp Assoc 18, 350-352, 1982
1983	US (Florida, Gainesville)	R. M. Bright	Transmissible venereal tumor of the soft palate in a dog	JAVMA 183, 8, 893-895, 1983

1983	US (New York)	C. Thacher	Vulvar and vaginal tumors in the dog: A retrospective study	JAVMA 183, 6, 690-692, 1983
1983	US	H. M. Hayes	Canine Transmissible Venereal Tumor: A Model for Kaposi's Sarcoma?	Am J Epidemiol 117, 1, 108-109, 1983
1983	Senegal (Dakar)	R. Parent	Presence of the canine transmissible venereal tumor in the nasal cavity of dogs in the area od Dakar (Senegal)	Can Vet J 24, 287-288, 1983
1984	US Virgin Islands	D. L. Hill	Canine Transmissible Venereal Sarcoma: Tumor Cell and Infiltrating Leukocyte Ultrastructure at Different Growth Stages	Vet Pathol 21, 39-45, 1984
1984	Nigeria (Ibadan)	A. L. Idowu	A retrospective evaluation of four surgical methods of treating canine transmissible venereal tumour	J Small Anim Pract 25, 193-198
1985	Papua New Guinea (Boroko)	A. N. Hamir	Primary penile and nasal transmissible venereal tumours in a dog	Aust Vet J, 62, 12, 430-432, 1985
1985	Israel (Be'er Sheva)	D. Cohen	The canine transmissible venereal tumor: a unique result of tumor progression	Adv Cancer Res 43, 75-111, 1985
1985	Spain	C. Murgia	Clonal Origin and Evolution of a Transmissible Cancer	Cell 126, 3, 477-487, 2006
1986	Nigeria (Zaria)	E.I. Amber	Oronasal transmissible venereal tumor in a dog	Mod Vet Pract, 67, 154, 1986
1986	Papua New Guinea (Boroko)	A. N. Hamir	Neoplasms of dogs in Papua New Guinea	Aust Vet J 63, 10, 342-343, 1986
1986	India (Ranchi)	L. L. Dass	Malignant Transmissible Venereal Tumor	Canine Pract 13, 3, 15-18, 1986
1987	US (Connecticut, Storrs and Illinois, Champaign)	N. Katzir	Common origin of transmissible venereal tumors (TVT) in dogs	Oncogene 1, 445-448, 1987
1987	Israel (Be'er Sheva)	N. Katzir	Common origin of transmissible venereal tumors (TVT) in dogs	Oncogene 1, 445-449, 1987
1987	Nigeria (Nsukka)	B. M. Anene	Common Diseases of Dogs in Nigeria	Zariya Veterinarian 2, 1, 46-55, 1987
1987	Brazil (Jaboticabal)	C.L.M Daleck	Avaliacao de Diferentes Métodos Diagnosticos do Tumor Venéreo Transmissível (T.V.T.) em Caes	ARS Veterinaria 3(2), 187-194, 1987
1987	South Africa (Medunsa)	M. I. Vermooten	Canine Transmissible Venereal Tumour (TVT): A Review	J S Afr Vet Assoc 58, 3, 147-150, 1987
1987	US (Indiana, West Lafayette)	G. E. Sandusky	Diagnostic Immunohistochemistry of Canine Round Cell Tumors	Vet Pathol 24, 495-499, 1987
1987	US Virgin Islands	T. J. Yang	Metastatic transmissible venereal sarcoma in a dog	JAVMA 190, 5, 555-556, 1987

1987	US (Connecticut, Storrs)	T. J. Yang	Growth stage dependent expression of MHC antigens on the canine transmissible venereal sarcoma	Br J Cancer 55, 131-134, 1987
1987	Brazil (Jaboticabal)	A. A. Camacho	Estudo sobre a eficiencia da vincristine no tratamento de caes com tumor venereo transmissivel	ARS Veterinaria 3(1), 37-42, 1987
1988	Greece (Thessaloniki)	C. Boscov	Canine transmissible venereal tumor: clinical observations and treatment	Anim Familiaris 3, 10-15, 1988
1989	US (Oklahoma, Stillwater)	R. L. Cowell	Cytology of cutaneous lesions	Vet Clin N Am-Small 19, 4, 769-794, 1989
1989	India (Ranchi)	L. L. Dass	Surgical treatment of canine transmissible venereal tumour - a retrospective study	Indian Vet J 66, 255-258, 1989
1989	India (Jabalpur)	S. K. Pandey	Incidence, treatment approach and metastasis of canine-transmissible venereal sarcoma	Indian J Anim Sci 59, 5, 510-513, 1989
1989	Germany (Munich)	C. Laging	Beobachtungen zum Ubertragbaren Venerischen Tumor (Sticker) beim Hund	Tierarztl Prax 17, 85-87, 1989
1990	Nigeria (Zaria)	E. I. Amber	Single-Drug Chemotherapy of Canine Transmissible Venereal Tumor With Cyclophosphamide, Methotrexate, or Vincristine	J Vet Intern Med 4, 144-147, 1990
1990	Tanzania (Morongo)	E. K. Batamuzi	Canine transmissible tumor in Morongo, Tanzania	Prev Vet J 21, 152-154, 1990
1990	US (Alabama, Auburn)	W. W. Miller	Ocular metastasis of a Transmissible Venereal Tumor	Canine Pract 15, 3, 19-21, 1990
1990	US	J. E. Moulton	Tumours of Domestic Animals	University of California Press, Berkeley and Los Angeles 10, 498-502, 1990
1990	India (Parbhani)	V. S. Panchbhai	Use of autogenous vaccine in transmissible canine venereal tumour	Indian Vet J 67, 983-984, 1990
1991	US (Connecticut, Storrs and Illinois, Champaign)	E.N. Amariglio	Identity of rearranged LINE/c-MYC junction sequences specific for the canine transmissible venereal tumor	Proc. Natl. Acad. Sci. USA 88, 8136-8139, 1991
1991	Israel (Tel-Aviv, Jerusalem)	E.N. Amariglio	Identity of rearranged LINE/c-MYC junction sequences specific for the canine transmissible venereal tumor	Proc. Natl. Acad. Sci. USA 88, 8136-8140, 1991
1991	Tanzania (Morogoro)	E.K. Batamuzi	Anal and perianal transmissible venereal tumour in a bitch	Vet Rec, 129, 556, 1991
1991	St. Kitts and Nevis	D. Kroger	An Unusual Presentation of Canine Transmissible Venereal Tumor vol.	Canine Pract 16, 6, 17-21, 1991

1991	India (Calcutta)	U. Das	Clinical report on the efficacy of chemotherapy in canine transmissible venereal sarcoma	Indian Vet J 68, 249-252, 1991
1991	India (Calcutta)	A. K. Das	A clinical report on the efficacy of vincristine on canine transmissible venereal sarcoma	Indian Vet J 68, 575-576, 1991
1992	Tanzania (Morogoro)	E. K. Batamuzi	Risk factors associated with canine transmissible venereal tumour in Tanzania	Prev Vet Med 13, 13-17, 1992
1993	Tanzania (Morogoro)	E.K. Batamuzi	Role of exfoliative cytology in the diagnosis of canine transmissible venereal tumour	J Small Anim Pract 34, 399-401, 1993
1993	India (Punjab)	V. K. Gandotra	Occurrence of canine transmissible venereal tumor and evaluation of two treatments	Indian Vet J 70, 854- 857, 1993
1993	Mexico (Mexico City)	S. E. Morales	The prevalence of transmissible venereal tumor in dogs in Mexico City between 1985-1993	Vet Mexico 26, 3, 273-275, 1993
1994	India (Chennai)	S. Ayyappan	Metastatic transmissible venereal tumour in a dog, a case report	Indian Vet J, 71, 265-266, 1994
1994	Spain (Cordoba)	J. Pérez	Primary Extragenital Occurrence of Transmissible Venereal Tumors: Three Case Reports	Canine Pract 19, 1, 7-10, 1994
1994	Germany (Giessen)	N. Kirchhof	Spinal metastasis of a canine transmissible tumor	Kleintierpraxis 39, 11, 797-800, 1994
1994	India (Mohanpur)	T. B. Sen	Treatment of canine transmissible venereal tumour with cyclophosphamide – a case report	Indian Vet J 71, 1120-1122, 1994
1995	Argentina (La Plata)	E. J. Gimeno	Intermediate filament expression and lectin histochemical features of canine transmissible venereal tumour	APMIS 103, 645-650, 1995
1995	Spain (Cordoba)	P.J. Ginel	Primary transmissible venereal tumour in the nasal cavity of a dog	Vet Rec 136, 222-223, 1995
1995	Italy (Sardinia)	C. Murgia	Clonal Origin and Evolution of a Transmissible Cancer	Cell 126, 3, 477-487, 2006
1995	Israel (Jerusalem)	A. Harmelin	Correlation of Ag-NOR protein Measurements with Prognosis in Canine Transmissible Venereal Tumour	J Comp Path 112, 429-433, 1995
1995	Zambia (Lusaka)	L. N. Chiti	Anal Transmissible Venereal Tumour in a Dog	Zambian J Vet Sci 1, 1, 21-22, 1995
1995	India (Durg)	S. K. Maiti	Therapeutic management of transmissible venereal tumor with vincristine in a dog – a case report	Indian Vet J 72, 614-615, 1995

1996	Brazil (Belo Horizonte)	R. M. C. Guedes	Extragenital transmissible venereal tumor in a dog. Case report	Arq Bras Med Vet Zootec 48, 3, 369-374, 1996
1996	India (Ludhiana)	J. Singh	Clinico-Pathological Studies on the Effect of Different Anti-Neoplastic Chemotherapy Regimens on Transmissible Venereal Tumours in Dogs	Vet Res Commun 20, 71-81, 1996
1996	Spain (Cordoba)	E. Mozos	Immunohistochemical Characterization of Canine Transmissible Venereal Tumor	Vet Pathol 33, 257-263, 1996
1997	Brazil	C. Murgia	Clonal Origin and Evolution of a Transmissible Cancer	Cell 126, 3, 477-487, 2006
1997	India (Ludhiana)	J. Singh	Effect of geriforte supplementation on blood chemistry in transmissible venereal tumor affected dogs treated with vincristine sulphate	Indian Vet J 74, 420-421, 1997
1998	US (Texas, College Station)	K.S. Rogers	Transmissible venereal tumor: A Retrospective Study of 29 Cases	J Am Anim Hosp Assoc 34, 463-470, 1998
1998	Greece (Thessaloniki)	C. M. Boscós	Ocular involvement of transmissible venereal tumor in a dog	Vet Ophthalmol 1, 167-170, 1998
1998	Taiwan (Taichung)	S. C. Chang	A clinical study of primary extragenital transmissible venereal tumors in dogs	J Chin Soc Vet Sci 24, 257-263, 1998
1998	Brazil (Jaboticabal)	R. A. Sobral	Occurrence of canine transmissible venereal tumor in dogs from the Jaboticabal region, Brazil	Ars Veterinaria 14, 1, 1-10, 1998
1998	Canada (Quebec) - imported from Venezuela	I. Mikaelian	Transmissible venereal tumor: A consequence of sex tourism in a dog	Can Vet J 39, 591, 1998
1998	Spain (Cordoba)	J. Pérez	Immunohistochemical study of the local inflammatory infiltrate in spontaneous canine transmissible venereal tumour at different stages of growth	Vet Immunol Immunop 64, 133-147, 1998
1998	Greece (Thessaloniki)	P. Saratsis	Semen quality during Vincristine treatment in dogs with transmissible venereal tumor	Theriogenology 53, 1185-1192, 1998
1999	Greece (Thessaloniki)	C. M. Boscós	Cutaneous Involvement of TVT in Dogs: A Report of Two Cases	Canine Pract 24, 4, 6-11, 1999
1999	US (Georgia)	C. Murgia	Clonal Origin and Evolution of a Transmissible Cancer	Cell 126, 3, 477-487, 2006
2000	India	U. Das	Review of canine transmissible venereal sarcoma.	Vet Res Commun 24: 545-556.

2000	Turkey (Ankara)	N. Erunal- Maral	Use of Exfoliative Cytology for Diagnosis of Transmissible Venereal Tumour and Controlling the Recovery Period of the Bitch	Dtsch Tierarzt Wschr 107, 5, 175 -180, 2000
2000	Portugal (Lisboa)	A. J. A. Ferreira	Brain and ocular metastases from a transmissible venereal tumour in a dog	J Small Anim Pract 41, 165-168, 2000
2000	Chile (Santiago)	C. M. Gonzales	Canine Transmissible Venereal Tumour: A Morphological and Immunohistochemical Study of 11 Tumours in Growth Phase and during Regression after Chemotherapy	J Comp Path 122, 241- 248, 2000
2000	Brazil (Rio de Janeiro)	J. S. Pereira	Immunohistochemical characterization of intraocular metastasis of a canine transmissible venereal tumor	Vet Ophthalmol 3, 43-47, 2000
2001	Brazil (Jaboticabal)	G. N. Rodrigues	Intraocular transmissible venereal tumor in a dog	Ciencia Rural 31, 1, 141- 143, 2001
2001	Brazil (Alfenas)	M. S. Varaschin	Clinical manifestations and pathological forms of canine venereal tumors in Alfenas, Minas Gerais	Clin Vet 32, 32-38, 2001
2001	Italy (Catania)	C. Murgia	Clonal Origin and Evolution of a Transmissible Cancer	Cell 126, 3, 477-487, 2006
2001	Greece (Thessaloni ki)	L. G. Papazoglou	Primary Intranasal Transmissible Venereal Tumour in the Dog: A Retrospective study of Six Spontaneous Cases	J Vet Med 48, 391-400, 2001
2001	Israel (Rehovot)	A. Harmelin	Use of a murine xenograft model for canine transmissible venereal tumor	AJVR 62, 6, 907-910, 2001
2001	Pakistan (Faisalabad)	M. Athar	Clinico-therapeutic studies on canine transmissible venereal tumour	Pakistan Vet J 21, 1, 39- 43, 2001
2002	Italy (Naples)	F. Albanese	Case report: Primary cutaneous extragenital canine transmissible venereal tumour with <i>Leishmania</i> - laden neoplastic cells: a further suggestion of histiocytic origin?	Vet Dermatol 13, 243- 246, 2002
2002	Brazil (Sao Paulo)	C.V.S. Brandao	Transmissible venereal tumour in dogs: a retrospective study of 127 cases (1998-2000)	Rev Educ Contin 5, 25- 31, 2002
2002	South Korea (Daejeon)	Y.K. Choi	Sequence Analysis of Canine LINE-1 Elements and p53 Gene in Canine Transmissible Venereal Tumor	J Vet Sci 3(4), 285-292, 2002
2002	Turkey (Istanbul)	A. Gurel	Transmissible Venereal Tumors Detected in the Extragenital Organs of Dogs	Israel J Vet Med 57, 2, 1- 8, 2002

2002	Argentina (La Plata)	C. Gobello	Effects of vincristine treatment on semen quality in a dog with a transmissible venereal tumour	J Small Anim Pract 43, 416-417, 2002
2002	Taiwan (Taipei)	Y.-W. Hsiao	Effect of tumor infiltrating lymphocytes on the expression of MHC molecules in canine transmissible venereal tumor cells	Vet Immunol Immunop 87, 19-27, 2002
2003	India (Kolkata)	C. Murgia	Clonal Origin and Evolution of a Transmissible Cancer	Cell 126, 3, 477-487, 2006
2003	Kenya (Nairobi)	C. Murgia	Clonal Origin and Evolution of a Transmissible Cancer	Cell 126, 3, 477-487, 2006
2003	Brazil (Botucatu)	S. Bassani-Silva	Tumor Venéreo Transmissível - Revisão	Revista Petfood & health & care 2, 77-82, 2003
2003	Taiwan (Taipei)	K.-W. Liao	Identification of canine transmissible venereal tumor cells using in situ polymerase chain reaction and the stable sequence of the long interspersed nuclear element	J Vet Diagn Invest 15, 399-406, 2003
2003	Mexico (Merida - Yucatan)	A. Ortega-Pacheco	Prevalence of transmissible venereal tumor of stray dogs in Merida, Yucatan, Mexico	Rev Biomed 14, 83-87, 2003
2003	Italy (Messina)	G. Catone	Canine Transmissible Venereal Tumour Parasitised by <i>Leishmania infantum</i>	Vet Res Commun 27, 549-553, 2003
2004	Zimbabwe (Harare)	S. Mukaratirwa	Stromal cells and extracellular matrix components in spontaneous canine transmissible venereal tumour at different stages of growth	Histol Histopathol 19, 1117-1123, 2004
2004	Zimbabwe (Harare)	S. Mukaratirwa	Canine transmissible venereal tumour: cytogenetic origin, immunophenotype, and immunobiology	Vet Quart 25, 101-111, 2004
2004	Mexico (Merida - Yucatan)	C. A. Rebbeck	Origins and Evolution of a Transmissible Cancer	Evolution 63, 2340-2349, 2009
2004	South Africa (Cape Town)	C. A. Rebbeck	Origins and Evolution of a Transmissible Cancer	Evolution 63, 2340-2349, 2009
2004	Thailand (Nakhon Pathom)	C. A. Rebbeck	Origins and Evolution of a Transmissible Cancer	Evolution 63, 2340-2349, 2009
2004	Kenya (Nairobi)	C. A. Rebbeck	Origins and Evolution of a Transmissible Cancer	Evolution 63, 2340-2349, 2009
2004	Greece (Thessaloniki)	C. A. Rebbeck	Origins and Evolution of a Transmissible Cancer	Evolution 63, 2340-2349, 2009
2004	Malaysia (Selangor)	C. A. Rebbeck	Origins and Evolution of a Transmissible Cancer	Evolution 63, 2340-2349, 2009
2004	Greece (Thessaloniki)	C. M. Boscos	Canine TVT - Clinical findings, Diagnosis and Treatment	Proceedings of the 29th World Congress of World Small Animal Association, 2004

2004	Brazil (Botucatu)	A. S. Amaral	Cytological diagnosis of transmissible venereal tumor in the Botucatu region, Brazil (descriptive study: 1994-2003)	RPCV 99, 167-171, 2004
2004	Taiwan (Taipei)	Y.-W. Hsiao	Tumor-infiltrating Lymphocyte Secretion of IL-6 Antagonizes Tumor-Derived TGFbeta-1 and Restores the Lymphokine-Activated Killing Ability	J Immunol 172, 1508-1514, 2004
2004	Nigeria (Ibadan)	M. A. Tella	Complete Regression of Transmissible Venereal Tumor (TVT) in Nigerian Mongrel Dogs with Vincristine Sulphate Chemotherapy	Afr J Biomed Res 7, 133-138, 2004
2004	Peru (Lima)	N. Mendoza	Frequency of the Transmissible Venereal Tumour in Dogs: Caseload in the Pathology Laboratory of the National University of San Marcos (Period 1998-2004)	Rev Inv Vet Peru 21, 1, 42-47, 2010
2005	Italy (Messina)	C. Murgia	Clonal Origin and Evolution of a Transmissible Cancer	Cell 126, 3, 477-487, 2006
2005	Brazil (Botucatu)	M. I. Mello Martins	Canine Transmissible Venereal Tumor: Etiology, Pathology, Diagnosis and Treatment	In: Concannon PW, England G, Ithaca: NY, 2005
2005	Argentina (La Plata)	M. I. Mello Martins	Canine Transmissible Venereal Tumor: Etiology, Pathology, Diagnosis and Treatment	In: Concannon PW, England G, Ithaca: NY, 2005
2005	Turkey (Bursa)	D. Nak	A Clinico-pathological Study on the Effect of Vincristine on Transmissible Venereal Tumour in Dogs	J Vet Med A 52, 366-370, 2005
2006	Italy (Naples)	F. Albanese	Case report: Extragenital transmissible venereal tumour associated with circulating neoplastic cells in an immunologically compromised dog	Vet Comp Oncol 4, 1, 57-62, 2006
2006	Brazil (Sao Paulo)	C. P. de Brito	Immunohistochemical determination of estrogen receptor-alpha in vaginal and tumor tissues of healthy and TVT-affected bitches and their relation to serum concentrations of estradiol-17beta and progesterone	Theriogenology, 2006, 2006
2006	Mexico (Merida - Yucatan)	A. Ortega-Pacheco	Pathological Conditions of the Reproductive Organs of Male Stray Dogs in the Tropics: Prevalence, Risk Factors, Morphological Findings and Testosterone Concentrations	Reprod Dom Anim 41, 429-437, 2006

2006	Zimbabwe (Harare)	S. Mukaratirwa	Canine Transmissible Venereal Tumour: Assesment of Mast Cell Numbers as Indicators of the Growth Phase	Vet Res Commun 30, 613-621, 2006
2006	South Korea (Seoul)	M.-S. Park	Disseminated transmissible venereal tumor in a dog	J Vet Diagn Invest 18, 130-133, 2006
2006	Kenya (Nairobi)	T. O. Abuom	Transmissible Venereal Tumor with Subcutaneous and Bone Metastasis in a Dog	Kenya Vet 30, 1, 11-13, 2006
2006	Greece (Thessaloniki)	E. Levy	Nasal and oral masses in a dog	Vet Clin Pathol 35, 115-118, 2006
2006	Portugal (Vila Verde)	R. Marcos & M. Santos	Cutaneous transmissible venereal tumor without genital involvement in a prepubertal female dog	Vet Clin Pathol 35,106-109, 2006
2007	Brazil (Mossoro - RN)	J.S. Batista	Tumor venereo transmissivel canino com localizacao intra-ocular e metastase no baco	Acta Vet Brasilica 1, 1, 45-48, 2007
2007	Brazil (Botucatu)	S. Bassani-Silva	Propolis effect <i>in vitro</i> on canine Transmissible Venereal Tumor cells	RPCV 102, 261-265, 2007
2007	Mexico (Merida - Yucatan)	A. Ortega-Pacheco	Reproductive patterns and reproductive pathologies of stray bitches in the tropics	Theriogenology 67, 382-390 , 2007
2007	Brazil (Botucatu)	A. S. Amaral	Cytomorphological characterization of transmissible canine venereal tumor	RPVC 102, 253-260, 2007
2007	Brazil (Garca)	K. C. Florentino	Tumor venéreo transmissível cutaneo canino = Relato de Caso	Revista Cientifica Eletronica de Medicina Veterinaria - ISSN, 1679-7353, 2007
2007	Nigeria (Zaria)	A. Z. Hassan	Canine Transmissble Venereal Tumor (CTVT) presented to the ABUVTH, Zaria: A 15 Year Review	Presented at the 44th Nigerian Veterinary Medical Association Annual Congress, 22nd-24th October 2007
2007	India (Bareilly)	B. C. Nair	A study on spontaneous canine neoplasms in Bareilly, U.P.	Indian J Vet Pathol 31, 2, 166-168, 2007
2007	Taiwan (Taipei)	C.-C. Liu	Transient downregulation of monocyte-derived dendritic-cell dfferentiation, function, and survival during tumoral progression and regression in an in vivo canine model of transmissible venereal tumor	Cancer Immunol Immunother 57, 479-491, 2007
2008	Italy (Messina)	G. Marino	Alternative Vincristine Treatment in Dogs With Transmissible Venereal Tumour	Veterinaria 22, 4, 13-18, 2008
2008	India (Tamilnadu, Namakkal)	R. Thangathurai	Cytological diagnosis and its histological correlation in canine transmissible venereal tumour	Vet Archiv 78, 5, 369-376, 2008

2008	Turkey (Ankara)	A. Bastan	Uterine and Ovarian Metastasis of Transmissible Venereal Tumor in a Bitch	Turk J vet Anim Sci 32, 1, 65-66, 2008
2008	Greece (Thessaloniki)	M. E. Mylonakis	A retrospective study of 61 cases of spontaneous canine epistaxis (1998-2001)	J Small Anim Pract 49, 191-196, 2008
2008	Brazil (Belo Horizonte)	F. G. A. Santos	Apoptosis in the transplanted canine transmissible venereal tumor during growth and regression phases	Arq Bras Med Vet Zootec 60, 3, 607-612, 2008
2008	Mexico (Toluca)	N. Vázquez-Mota	The T963C mutation of TP53 gene does not participate in the clonal origin of canine TVT	Vet Res Commun 32, 187-191, 2008
2009	India (Nagpur)	L. A. Khan	Incidence of Venereal Granuloma and its Medicinal Treatment in stray Dogs of Nagpur City	Vet World 2, 1, 13-14, 2009
2009	Nigeria (Makurdi)	I.A. Kisani	A case of transmissible venereal tumor in a castrated dog in Benue state, Nigeria	J Anim Plant Sci 2009 5, 527-530, 2009
2009	Mexico (Toluca)	A. Sánchez-Servín	TP53 Polymorphisms allow for genetic sub-grouping of the canine transmissible venereal tumor	J Vet Sci 10, 353-355, 2009
2009	Brazil (Ilheus)	R. A. Said	Efficacy and side effects of Vincristine sulphate treatment on canine transmissible venereal tumour	Proceedings of the 34th World Small Animal Veterinary Congress WSAVA, 2009, Sao Paulo
2009	Zimbabwe (Harare)	S. Mukaratirwa	Combination therapy using intratumoral bacillus Calmette-Guerin (BCG) and vincristine in dogs with transmissible venereal tumours: therapeutic efficacy and histological changes.	J S Afr Vet Assoc 80,2, 92-96, 2009
2010	Brazil (Sao Paulo)	K. C. Scarpelli	Predictive factors for the regression of canine transmissible venereal tumor during Vincristine therapy	Vet J 183, 3, 362-363, 2010
2010	Brazil (Botucatu)	L. F. J. Gaspar	Spontaneous canine transmissible venereal tumor: cell morphology and influence on P-glycoprotein expression	Turk J Vet Anim Sci 34, 5, 447-454, 2010
2010	Mexico (Mexico City)	J. C. Cruz	Canine Transmissible Venereal Tumor in the Metropolitan Area of Mexico City	Rev Cientif 20, 4, 362-366, 2010
2010	Grenada	R. M. Kabuusu	Risk factors and characteristics of canine transmissible venereal tumours in Grenada, West Indies.	Vet Comp Oncol 8, 1, 50-55, 2010
2010	Bangladesh (Dhaka)	M. Tarafder	Prevalence of clinical diseases of pet dogs and risk perception of zoonotic infection by dog owners in Bangladesh	Bangl J Vet Med 8, 2, 163-174, 2010

2011	Brazil (Sao Paolo)	D. Stockmann	Detection of the tumour suppressor gene TP53 and expression of p53, Bcl-2 and p63 proteins in canine transmissible venereal tumour	Vet Comp Oncol 9, 4, 251-259, 2011
2011	Mexico	L. G. Bautista-Gomez	Analysis of canine transmissible venereal tumor genotypes using the D-loop region of mitochondrial DNA	Genes Genet Syst 86, 351-355, 2011
2011	India (Chennai)	T. Sathiamoorthy	Prevalence of reproductive disorders in the stray dogs of Chennai City	JIVA 9, 2, 62-63, 2011
2011	India (Tamil Nadu)	M. Selvaraju	Transmissible Venereal Tumor in a German Shepherd Bitch	Indian Pet J (online) 3, 2, ISSN 2230-7613, 2011
2011	Grenada	A. Chikweto	Neoplastic and Nonneoplastic Cutaneous Tumors of Dogs in Grenada, West Indies	ISRN Vet Sci Article ID 416435
2012	Italy (Messina)	G. Marino	Clinicopathological study of canine transmissible venereal tumour in leishmaniotic dogs	J Small Anim Pract 53, 323-327, 2012
2012	India (Ludhiana)	K. Gupta	Pathological and immunohistochemical studies on rare cases of primary extra-genital transmissible venereal tumours in the mammary gland	Vet Med 57, 4, 198-206, 2012
2012	India (Ludhiana)	E. E. Varughese	Successful Management of Metastatic Transmissible Venereal Tumour to Skin of Mammary Region	Reprod Dom Anim 47 (Suppl. 6), 366-369, 2012
2012	Brazil (Aracatuba)	J. T. Trevizan	Disseminated Transmissible Venereal Tumour Associated With Leishmaniasis in a Dog	Reprod Dom Anim 47 (Suppl. 6), 356-358, 2012
2012	Egypt (Ismailia)	A. M. Ibrahim	Pathology in practice. Transmissible venereal tumor located on the bulbus glandis and body of the penis.	JAVMA 241, 6, 707-709, 2012
2012	Turkey (Bursa)	G. R. Ozalp	Vincristine modulates the expression of Ki67 and apoptosis in naturally occurring canine transmissible venereal tumor (TVT)	Biotech Histochem 87, 5, 325-330, 2012
2013	Taiwan (Taichung)	Y.-C. Chen	Expression of MAGE-A restricted to testis and ovary or to various cancers in dogs	Vet Immunol Immunopathol 153, 26-34, 2013
2013	Paraguay (Asuncion)	K. Kegler	Vaginal Canine Transmissible Venereal Tumour Associated with Intra-tumoural <i>Leishmania spp.</i> Amastigotes in an Asymptomatic Female Dog	J Comp Pathol 149, 156-161, 2013
2013	Grenada	A. Chikweto	Genital and Extragenital Canine Transmissible Venereal Tumor in Dogs in Grenada, West Indies	OJVM 3, 111-114, 2013

2013	Pakistan (Faisalabad)	A. Tariq	Vincristine sulfate: An Effective Drug against Transvenereal Tumors	IJMVR 3, 11, 62-64, 2013
2014	Grenada	J. Milo	A case of ocular canine transmissible venereal tumor	Can Vet J. 55 (2),106, 2014
2014	Switzerland	M. Amrein	Vom ewigen Leben eines Hundes	NZZ am Sonntag, Wissen 16 March 2014, 58, 2014
2014	Pakistan (Lahore)	F. Awan	Comparison of Different Therapeutic Protocols in the Management of Canine Transmissible Venereal Tumour: Review of 30 Cases	Global Vet 12, 4, 499-503, 2014
2014	Brazil (Seropedica)	M. B. Mascarenhas	Immunohistochemical study of genital and extragenital forms of canine transmissible venereal tumor in Brazil	Pesq Vet Bras 34, 3, 250-254
2014	Brazil (Curitiba)	D. M. Da Silva	Treatment of canine transmissible venereal tumor using L-asparaginase, prednisone, and surgery in a clinical chemotherapy-resistant case	Turk J Vet Anim Sci 38, 220-223, 2014
2014	Brazil (Porto Alegre)	D. G. Gerardi	Expression of P-glycoprotein, multidrug resistance-associated protein, glutathione-S-transferase pi and p53 in canine transmissible venereal tumor	Pesq Vet Brasil 34, 1, 71-78, 2014
2014	Bangladesh (Cittagong)	M. S. Islam	Progressive Type of Canine Transmissible venereal Tumor (CTVT) in a male Stray Dog: a Case Report	Res J Vet Pract 2, 4, 70-72, 2014
2014	Iran	J. Javanbakht	Canine transmissible venereal tumor and seminoma: a cytopathology and chemotherapy study of tumors in the growth phase and during regression after chemotherapy	Tumor Biol, DOI 10.1007/s13277-014-1723-5, 2014

APPENDIX: CHAPTER 3

**Department of Veterinary Medicine
University of Cambridge**
Madingley Road Cambridge CB3 0ES



Telephone: +44 (0)1223 760418

Canine transmissible venereal tumour (CTVT) project

Patient name:

Animal ID:

In order to understand the genetic changes that cause canine transmissible venereal tumour (CTVT), we would like to enroll your dog into our study.

This project has been approved by the Department's Ethics and Welfare Committee [Project reference CR174], Department of Veterinary Medicine, University of Cambridge, Cambridge UK, CB3 0ES

I have read and understood the Owner Information Sheet, and consent for my dog to be enrolled in the study. I understand that I can withdraw my dog from the study at any time.

Owner name (capital letters): _____

Signature: _____

Date: _____

Canine transmissible venereal tumour (CTVT) project

Owner information sheet

About the project

The canine transmissible venereal tumour (CTVT) is a transmissible cancer that affects dogs. This disease is spread between dogs by the transfer of living cancer cells, usually during mating. CTVT causes the appearance of tumours associated with the external genitalia of male and female dogs. CTVT tumours can usually be successfully treated with chemotherapy.

The CTVT project aims to understand the genetic changes that have occurred within CTVT tumours in dogs around the world. By analysing the genetic changes in CTVT our goal is to understand the changes that initially caused CTVT, as well as the factors that are continuing to influence its evolution. We hope in future to use this information to develop better treatments or prevention strategies for CTVT. In addition, this study will help us to understand how cancer evolves more generally, including in humans.

How will we collect the samples?

Your veterinarian will take a small piece of your dog's tumour tissue and put it in a tube with chemicals for storage of DNA. In addition, your veterinarian will take a small piece of normal tissue and will store this in a tube.

What will happen to the samples?

The samples will be shipped to Cambridge, UK, where we will extract DNA from them. The DNA will be sequenced using a genetic analyser. The genetic information will then be analysed to find the mutations which make each dog's tumour unique. This information will be compared with CTVT tumours collected in different countries around the world.

After extraction, the samples will be archived at the Department of Veterinary Medicine, University of Cambridge. In future, the samples may be used for further studies of CTVT.

Where can I find more information?

To find more information about this study, please visit our website www.tcg.cam.ac.uk. Please do not hesitate to contact study coordinators Elizabeth Murchison (epm27@cam.ac.uk), Andrea Strakova (as2112@cam.ac.uk) or Tracy Wang (jw401@cam.ac.uk) at any time.

Thank you for your participation in this study!



Canine Transmissible Venereal Tumour (CTVT)

Project
Department of Veterinary Medicine
University of Cambridge

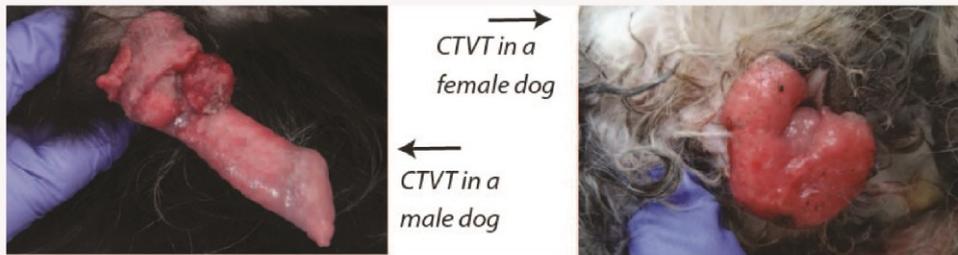
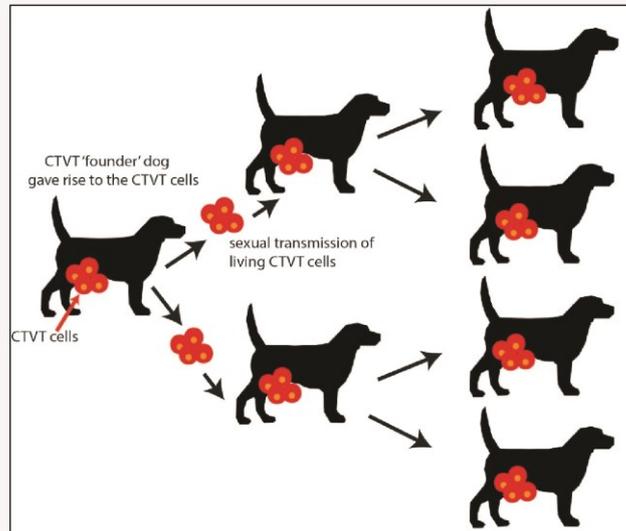


For any questions about the project or sample collection contact
Andrea Strakova (as2112@cam.ac.uk), Tracy Wang (jw401@cam.ac.uk) or Dr Elizabeth
Murchison (epm27@cam.ac.uk)

THANK YOU VERY MUCH FOR YOUR CONTRIBUTION, YOUR SAMPLES WILL BE
EXTREMELY USEFUL AND INTERESTING FOR OUR STUDY!

WHAT IS CTVT?

Canine Transmissible Venereal Tumour (CTVT), also called TVT, Sticker's sarcoma, is a transmissible cancer which arose around 11,000 years ago and has been transmitted by living cancer cells during coitus between individual dogs.



AIMS OF THE CTVT PROJECT

1. Understand the origins of CTVT and its spread through the dog population
2. Map the genetic diversity of CTVT around the world
3. Understand the evolution of CTVT
4. Develop new methods for CTVT prevention and treatment
5. Use knowledge from our studies of this unique cancer to advance understanding of human cancer evolution

CTVT INFECTED CASES

The three most important samples that we are collecting from each CTVT infected animal are:

- 1) Tumour in RNAlater (label e.g. #1 TVT, tumour, RNA)
- 2) Tumour in formalin (label e.g. #1 TVT, tumour, formalin)
- 3) Host tissue in RNAlater - gonads (better) or blood (mix 2-3ml EDTA blood + 1ml RNAlater) (label e.g. #1 TVT, ovary/testis, RNA)
- 4) Photo, if possible

Please, fill in a [data collection sheet](#) with each case.

Additional samples, if possible, but not necessary:

- 5) Serum sample (around 2-3ml of serum if possible)
- 6) Cytology smear
- 7) Follow up samples during treatment – collect tumour samples each week when animal receives Vincristine (no need for a host sample)
 - tumour in RNAlater
 - tumour in formalin
 - serum

CONTROL CASES

These are samples from the same population of dogs, but these dogs are not infected with CTVT. Collect 1-2 control samples for each CTVT case.

- 1) Host tissue in RNAlater – gonads during spay/neuter surgery (label #1 Control, ovary/testis, RNA)
- 2) Photo, if possible
- 3) Serum – not necessary, but great if possible!

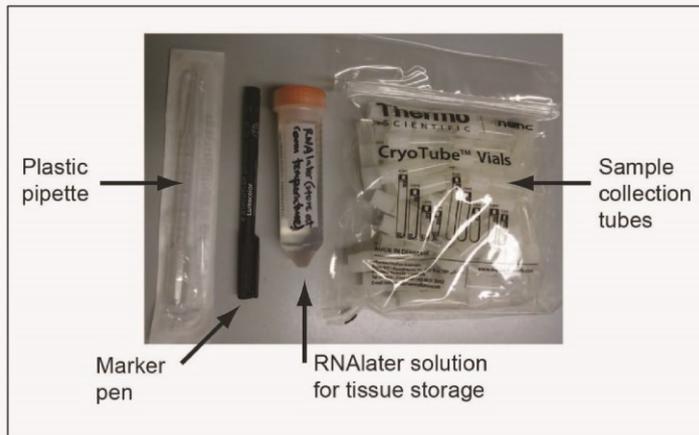
Please, fill in a [data collection form](#) with each case (even for control cases).

CTVT PROJECT SAMPLE COLLECTION AND SHIPMENT



SAMPLE COLLECTION KIT

Sample collection kit contains (photo below): RNAlater solution, 2ml cryo tubes, pipettes, marker pen, sample collection leaflet. To request a sample collection kit, contact Andrea or Tracy, and we will post it to you!



SHIPMENT OF SAMPLES

Samples can be shipped at room temperature. We will pay for all the shipping costs associated. Ship samples by **FedEx** and use our account number for the Transmissible Cancer Group: **156982300**

We have a **UK import permit** - contact **Andrea** (as2112@cam.ac.uk) or **Tracy** (jw401@cam.ac.uk) when you have samples ready for shipment to request a copy of the permit.

Shipping address:
Andrea Strakova/Tracy Wang/Dr Elizabeth Murchison
Department of Veterinary Medicine, University of
Cambridge
Maddingley Road, Cambridge, CB3 0ES
United Kingdom



Ideally, collect 5 items for each CTVT infected case:

- 1) **Tumour in RNAlater** (label e.g. #1 TVT, tumour, RNA)
Cut a small piece of the tumour (0.5cm³) and place it into small cryovial tube with 2ml of RNAlater
- 2) **Tumour in formalin** (label e.g. #1 TVT, tumour, formalin)
Cut a small piece of the tumour (0.5cm³) and place it into small cryovial tube with 2ml of formalin (not provided in the kit)
- 3) **Host tissue in RNAlater** - gonads (better), skin or blood (mix 2-3ml EDTA blood + 1ml RNAlater) (label e.g. #1 TVT, ovary/testis, RNA)
Place a small piece of the host sample (0.5cm³) into small cryovial tube with 2ml of RNAlater. Only one of the following samples is sufficient: ovary/testis (optimal), uterus or skin. (If euthanasia, you can collect internal organs). Least favourable, but still possible, is ~2-3 ml of blood in EDTA (purple tube) topped up with 1 -1.5 ml of RNAlater. Mix well.
- 4) **Photograph**, if possible
Photograph of the tumour and the dog.
- 5) Completed **data collection sheet**
As a minimum, please, record the tumour site and size, sex, age and health status of the dog.

Ideally, collect 3 items for controls = CTVT uninfected cases:

- 1) **Host tissue in RNAlater** - gonads during a spay/neuter surgery (label e.g. #1 Control, ovary/testis, RNA)
- 2) **Photograph** of the dog, if possible
- 3) Completed **data collection sheet**

Additional file 3-3 CTVT clinical data collection sheet.

Canine Transmissible Venereal Tumour (TVT) Research Project

Please fill out a sheet for each case, providing as much detail as possible. Thanks for your participation!

YOUR INFORMATION

Your name: Date:

Organization: Email:

Town/Region of sampling: Country:

DOG INFORMATION

Dog name or ID: Breed: Age:

Sex: Male Ownership Status: Stray Pregnant: Yes Puppies: Yes
 Female Owned No No

Reproductive status: Spayed/neutered When was the dog spayed/neutered?
 Entire

SAMPLE INFORMATION

TVT STATUS?:

TVT INFECTED

↓ Fill in this section only

TVT UNINFECTED (CONTROL)

↓ Fill in this section only

Tumour in RNAlater collected: Yes No

Tumour in formalin collected: Yes No

Host tissue in RNAlater collected: Yes No

Tissue: Gonad
 Skin
 Liver
 Blood (EDTA or heparin)
 Other

Photograph taken (send by email): Yes No

Host tissue in RNAlater collected: Yes No

Tissue: Gonad
 Skin
 Liver
 Blood (EDTA or heparin)
 Other

Photograph taken (send by email): Yes No

Tumour location:

Tumour size:

Tumour description:

Tumour ulceration: Yes No

Tumour discharge: Yes No

Comments (health status, treatment and metastasis details):

Please, include any other notes or comments on the other side of the sheet.

Additional file 3-4 Sample collection leaflet – printed version included in each sample collection kit.



Canine Transmissible Venereal Tumour (CTVT)

Project

Department of Veterinary Medicine
University of Cambridge



Storage of samples

- Samples collected in **RNAlater** - store in *fridge* or in the *freezer* (if storing more than 3 weeks)
- Samples collected in **formalin** - store at room temperature or *fridge* - DO NOT STORE IN THE FREEZER

Shipment of samples

Samples can be shipped at room temperature. We will pay for all the shipping costs associated.

Ship samples by **FedEx** and use our account number for the Department of Veterinary Medicine: **156982300**

We have a **UK import permit** - contact **Andrea or Tracy** (as2112@cam.ac.uk or jw401@cam.ac.uk) when you have samples ready for shipment to request a copy of the permit.

Shipping address:

Andrea Strakova, Tracy Wang and Dr Elizabeth Murchison
Department of Veterinary Medicine, University of Cambridge
Madingley Road, Cambridge, CB3 0ES, United Kingdom

Sample collection kit contains: RNAlater solution, 2ml cryo tubes, pipettes, marker pen, sample collection leaflet.

To **request a sample collection kit**, contact **Andrea or Tracy** and we will post it to you!



For any questions about the project or sample collection contact **Andrea Strakova** (as2112@cam.ac.uk), **Tracy Wang** (jw401@cam.ac.uk) or **Dr Elizabeth Murchison** (epm27@cam.ac.uk).

THANK YOU VERY MUCH FOR YOUR CONTRIBUTION, YOUR SAMPLES WILL BE EXTREMELY USEFUL AND INTERESTING FOR OUR STUDY!

What is CTVT?

Canine Transmissible Venereal Tumour (CTVT), also called **TVT, Sticker's sarcoma**, is a transmissible cancer which arose around 11,000 years ago and has been transmitted by living cancer cells during coitus between individual dogs.



Aims of the CTVT project

- 1) Understand the origins of CTVT and its spread through the dog population
- 2) Map the genetic diversity of CTVT around the world
- 3) Understand the evolution of CTVT
- 4) Develop new methods for CTVT prevention and treatment
- 5) Use knowledge from our studies of this unique cancer to advance understanding of human cancer evolution

Sample collection protocol

Ideally, collect 5 items for each **CTVT infected case**:

- 1) **Tumour in RNAlater** (label e.g. #1 TVT, tumour, RNA)
Cut a small piece of the tumour (0.5cm³) and place it into small cryovial tube with 2ml of RNAlater
- 2) **Tumour in formalin** (label e.g. #1 TVT, tumour, formalin)
Cut a small piece of the tumour (0.5cm³) and place it into small cryovial tube with 2ml of formalin (not provided in the kit)

- 3) **Host tissue in RNAlater** - gonads (better), skin or blood (mix 2-3ml EDTA blood + 1ml RNAlater) (label e.g. #1 TVT, ovary/testis, RNA)

Place a small piece of the host sample (0.5cm³) into small cryovial tube with 2ml of RNAlater. Only one of the following samples is sufficient: ovary/testis (optimal), uterus or skin. (If euthanasia, can collect internal organs). Least favourable, but still possible, is ~2-3 ml of blood in EDTA (purple tube) topped up with 1-1.5 ml of RNAlater. Mix well.

- 4) **Photograph**, if possible

Photograph of the tumour and the dog.

- 5) Completed **data collection sheet**

As a minimum, please, record the tumour site and size, sex, age and health status of the dog.

Ideally, collect 3 items for controls = **CTVT uninfected cases**:

- 1) **Host tissue in RNAlater** - gonads during a spay/neuter surgery (label e.g. #1 Control, ovary/testis, RNA)
- 2) **Photograph** of the dog, if possible
- 3) Completed **data collection sheet**

APPENDIX: CHAPTER 4

Additional file 4-1 Confirmation of CTVT diagnosis.

Confirmation of CTVT diagnosis.

Quantitative PCR (qPCR) was performed for *LINE-MYC*, a CTVT-specific rearrangement (Katzir et al., 1985; Katzir et al., 1987). Each reaction was performed in triplicate and a standard curve was used to detect relative DNA input at each locus. 'Normalised input' represents the relative *LINE-MYC* input detected in each sample normalised to *ACTB* ('Confirmation of canine transmissible venereal tumour (CTVT) diagnosis', Materials and methods). In general, we consider normalised input >0.05 as indicative of presence of *LINE-MYC*. Sufficient DNA was not available for samples 1380T and 1381T; diagnosis in these cases was performed with histopathology. *2T was grown as a xenograft.

File available at: <https://doi.org/10.7554/eLife.14552.017>

Additional file 4-2 Sequencing coverage and tumour cell fraction.

Sequencing coverage and tumour cell fraction.

(A) Average per-base coverage for whole genome (CanFam3.1) and for mtDNA genome (CanFam3.1; NC_002008). (B) List of 11 CTVT hosts with low coverage mtDNA regions. (C) Estimated tumour cell fraction for 449 CTVT tumours; tumour cell fraction was estimated by calculating the average VAF for variant substitutions present in tumour but not in matched host for each tumour.

File available at: <https://doi.org/10.7554/eLife.14552.016>

Additional file 4-3 Single point substitution variant lists.

Single point substitution variant lists.

(A) Total number of substitution variants (n = 1005) identified in 449 CTVT tumours. (B) CTVT tumour somatic substitutions list (n = 928), including the average VAF value normalised for host contamination (see Materials and methods 'Classification of tumour substitutions'). Back mutations are not included on the list. (C) CTVT tumour conservative somatic substitutions list (n = 835), including the average VAF value normalised for host contamination (see Materials and methods 'Classification of tumour substitutions'). Back mutations are not included on the list. (D) Germline clade defining substitutions lists. Substitutions present in the pool of host substitutions and also shared between all samples within a clade (see Materials and methods 'Classification of tumour substitutions'). (E) Potential somatic substitutions lists. Substitutions not present in the pool of host substitutions, but shared between all samples within a clade (see Materials and methods 'Classification on tumour substitutions'). (F) Total number of substitution variants (n = 1152) identified in 338 CTVT host samples and 252 publicly available dog mitochondrial genomes (see Supplementary file 8).

Available at: <https://doi.org/10.7554/eLife.14552.018>

Additional file 4-4 Sample information.

Sample information.

Summary of information available for 449 CTVT tumours and 338 hosts sequenced in this study. Includes data on location, year of collection, CTVT mtDNA clade, tumour and host mtDNA haplotypes, breed, age and sex.

<https://doi.org/10.7554/eLife.14552.015>

Additional file 4-5 Annotation of single point substitutions.

Annotation of single point substitutions.

Annotation of individual point substitution mutations in (A) 449 CTVT tumours (see list Supplementary file 4A, excluding back mutations) and (B) 338 CTVT hosts (see list Supplementary file 4F). Annotation was performed using Variant Effect Predictor (McLaren et al., 2010). In cases where a single substitution affects two different genes, the two annotations are shown on different lines.

<https://doi.org/10.7554/eLife.14552.020>

Additional file 4-6 Publicly available mitochondrial dog genomes used in the study.

Publicly available mitochondrial dog genomes used in the study.

Summary of Genbank accession numbers and metadata for 252 publicly available dog mitochondrial genomes included in this study (see Materials and methods 'Extracting substitution variants from publicly available dog sequences' and Supplementary file 4F).

<https://doi.org/10.7554/eLife.14552.022>

Additional file 4-7 Summary of small insertions and deletions (indels).

Summary of small insertions and deletions (indels).

(A) Total number of insertions and deletions identified in tumours (n = 27), including the average VAF value normalised for host contamination (see Materials and methods) (B) Total number of homoplasmic insertions and deletions in CTVT hosts (n = 7), including the average VAF value (see Materials and methods).

<https://doi.org/10.7554/eLife.14552.019>

Additional file 4-8 Annotation of insertions and deletions (indels).

Annotation of insertions and deletions (indels).

Annotation of individual indels (A) unique to CTVT tumours and (B) homoplasmic in CTVT hosts. Annotation was performed using Variant Effect Predictor (McLaren et al., 2010).

<https://doi.org/10.7554/eLife.14552.021>

Additional file 4-9 Maximum likelihood phylogenetic trees for CTVT clades 1 to 5.

Maximum likelihood phylogenetic trees for CTVT clades 1 to 5.

Maximum likelihood phylogenetic trees for CTVT mtDNA in (A) clade 1 (n = 170) (B) clade 2 (n = 252) (C) clade 3 (n = 22) (D) clade 4 (n = 3) and (E) clade 5 (n = 2), rooted with haplotypes CTVT1 to CTVT5 respectively, which contain clade-defining germline and potential somatic substitutions specific to each clade (Figure 1—figure supplement 4). Bootstrap values were calculated from 100 bootstrap replicates and are shown where bootstrap values ≥ 60 . Scale bars indicate base substitutions per site. Clade 5 contains only two tumours, which are identical both to each other and to the CTVT5 haplotype; thus the tree for this clade was created separately and does not have a scale bar.

<https://doi.org/10.7554/eLife.14552.005>

Additional file 4-10 Ancient mtDNA recombination in CTVT clade 1.

Ancient mtDNA recombination in CTVT clade 1.

Maximum likelihood cladograms constructed using clade 1 mtDNA positions (A) 1-5429bp and (B) 5430-16176bp (see Materials and methods). Trees were constructed with 153 clade 1 CTVT mtDNAs rooted with the CTVT1 haplotype, which contains clade 1 clade-defining germline and potential somatic substitutions (Materials and methods, Figure 1—figure supplement 4). Bootstrap values were calculated from 100 bootstrap replicates and are shown where bootstrap values ≥ 60 .

<https://doi.org/10.7554/eLife.14552.014>

Additional file 4-11 CTVT tumour and host mtDNA haplotype lists.

CTVT tumour and host mtDNA haplotype lists.

<https://doi.org/10.7554/eLife.14552.025>

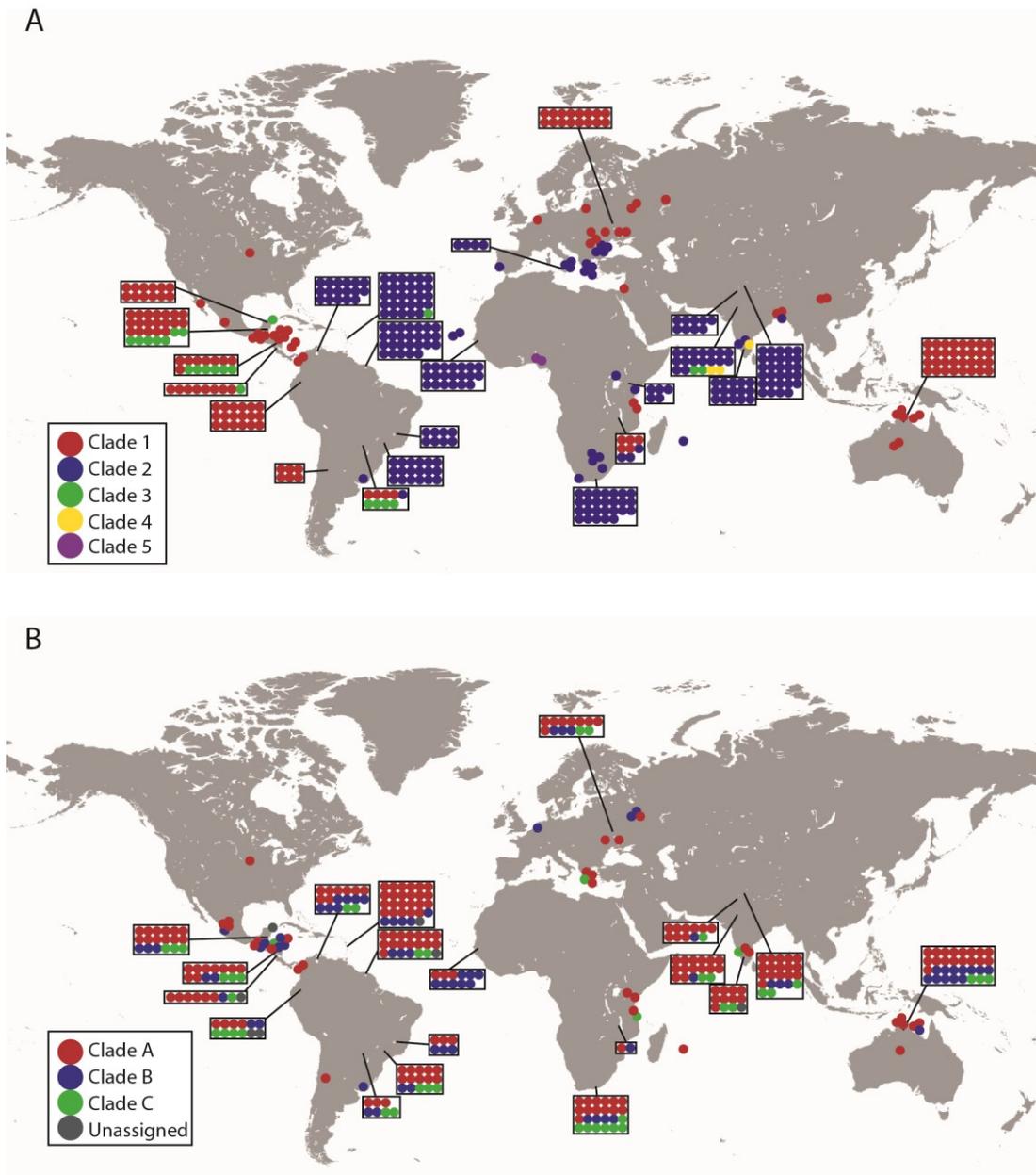
Additional file 4-12 Maximum likelihood phylogenetic tree of CTVT mtDNA.

Maximum likelihood phylogenetic tree of CTVT mtDNA.

Maximum likelihood phylogenetic tree constructed using 449 complete CTVT mitochondrial genomes and 590 complete dog mitochondrial genomes. All sequences are labelled with sample identifier, country, breed and haplotype name. The sample identifier for CTVT hosts is the sample name (Supplementary file 1), the sample identifier for the publicly available dogs is the accession number. Scale bar indicates base substitutions per site.

<https://doi.org/10.7554/eLife.14552.004>

Additional file 4-13 Geographical locations and mtDNA clades for CTVT tumours and hosts. Each dot represents the location of (A) a CTVT tumour, coloured by CTVT mtDNA clade (B) a CTVT host, coloured by dog mtDNA clade.



Additional file 4-14 Summary of back mutations.

Summary of back mutations.

List of back mutations attributable and non-attributable to a putative recombination event.

<https://doi.org/10.7554/eLife.14552.024>

Additional file 4-15 Substitutions with corresponding VAF (before normalisation) for each of 449 CTVT tumours and 338 CTVT hosts.

Substitutions with corresponding VAF (before normalisation) for each of 449 CTVT tumours and 338 CTVT hosts.

<https://doi.org/10.7554/eLife.14552.026>

Additional file 4-16 Indels with corresponding VAF (before normalisation) for each of 438 CTVT tumours and 334 CTVT hosts.

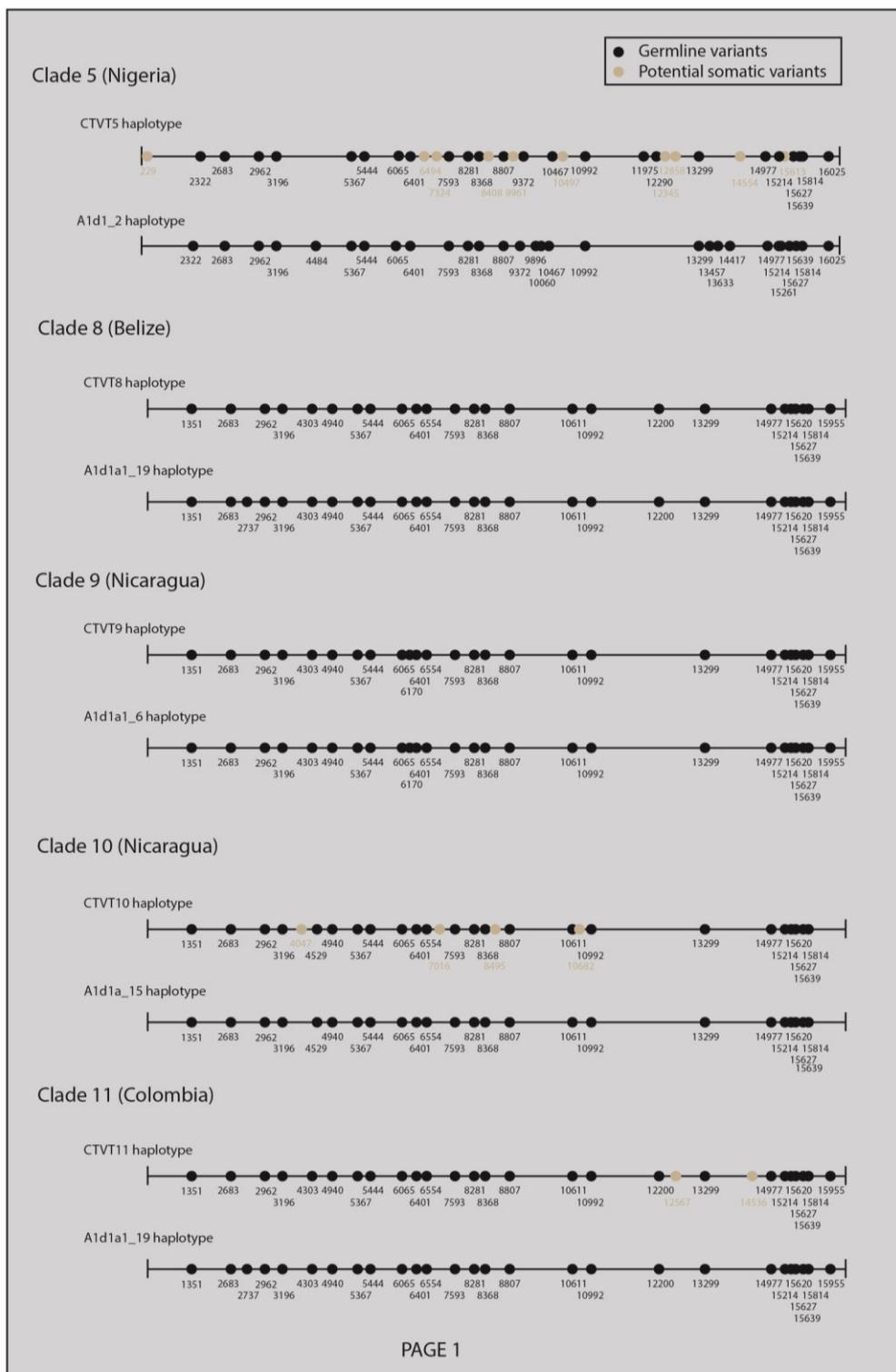
Indels with corresponding VAF (before normalisation) for each of 438 CTVT tumours and 334 CTVT hosts.

Samples with very high coverage of the mitochondrial genome were excluded from the indel analysis (see Chapter 4, section 4.2.6.1).

<https://doi.org/10.7554/eLife.14552.024>

APPENDIX: CHAPTER 5

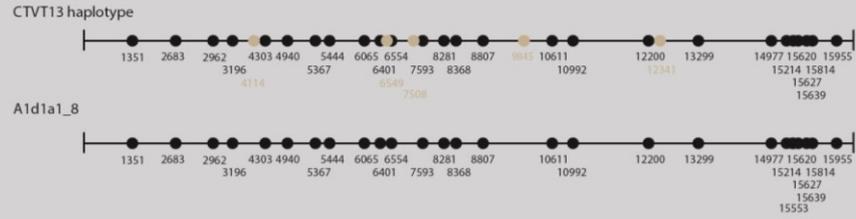
Additional file 5-1 Reconstructed donor haplotypes for CTVT mtDNA clades 5, 8-18. MtDNA is represented linearly and genome coordinates are labelled. Diagrams represent the likely donor haplotype for each of the CTVT mtDNA clades 5, 8-18. The coordinates for each substitution variant position are shown, and substitutions are colour-coded either as “germline” (i.e. they are present in all tumours within a clade and are found in the most closely related dog mtDNA haplotype, which is represented below each of the clade diagrams or they are found in the most closely related dog mtDNA haplotype only); or “potential somatic” (i.e. they are present in all tumours within a clade but are not found in the most closely related dog mtDNA haplotype).



Clade 12 (Colombia)



Clade 13 (Chile)



Clade 14 (The Gambia)



Clade 15 (Grenada)



Clade 16 (India)



Clade 17 (Paraguay)

CTVT17 haplotype



A1d1a_1 haplotype



Clade 18 (Nicaragua)

CTVT18 haplotype



A1d1a_1 haplotype



APPENDIX: PUBLICATIONS ASSOCIATED WITH THIS THESIS

The following publications associated with this Thesis are further enclosed at the end of this Thesis:

Additional file 6-1 Publications associated with this Thesis.

Strakova A and Murchison EP. The cancer which survived: Insights from the genome of an 11,000 year-old cancer. Curr. Opin. Genet. Dev. 30, 49-55 (2015).

Strakova A, Murchison EP. The changing global distribution and prevalence of canine transmissible venereal tumour. BMC Vet Res. 10, 168 (2014).

Castro FK, Strakova A, Tinucci Costa M, Murchison EP. Evaluation of a genetic assay for canine transmissible venereal tumour diagnosis in Brazil. Vet Comp Oncol. 15: 615–618 (2016).

Strakova A, Ni Leathlobhair M, Wang G-D, Yin T-T, Airikkala-Otter I, Allen JL, Allum KM, Bansse-Issa L, Bisson JL, Castillo Domracheva A, de Castro KF, Corrigan AM, Cran HR, Crawford JT, Cutter SM, Delgadillo Keenan L, Donelan EM, Faramade IA, Flores Reynoso E, Fotopoulou E, Fruean SN, Gallardo-Arrieta F, Glebova O, Häfelin Manrique RF, Henriques JJGP, Ignatenko N, Koenig D, Lanza-PereaM, Lobetti R, Lopez Quintana AM, Losfelt T, Marino G, Martincorena I, Martínez Castañeda S, Martínez-López MF, Meyer M, NakanwagiB, De Nardi AB, Neunzig W, NixonSJ, Onsare MM, Ortega-PachecoA, Peleteiro MC, PyeRJ, ReeceJF, Rojas Gutierrez J, Sadia H, Schmeling SK, Shamanova O, Ssuna RK, Steenland-Smit AE, Svitich A, Thoya Ngoka I, Vițălaru BA, de Vos AP, de Vos JP, Walkinton O, Wedge DC, Wehrle-Martinez AS, van der Wel MG, Widdowson SAE and Murchison EP. Mitochondrial genetic diversity, selection and recombination in a canine transmissible cancer. eLife. (2016); DOI: 10.7554/eLife.14552.