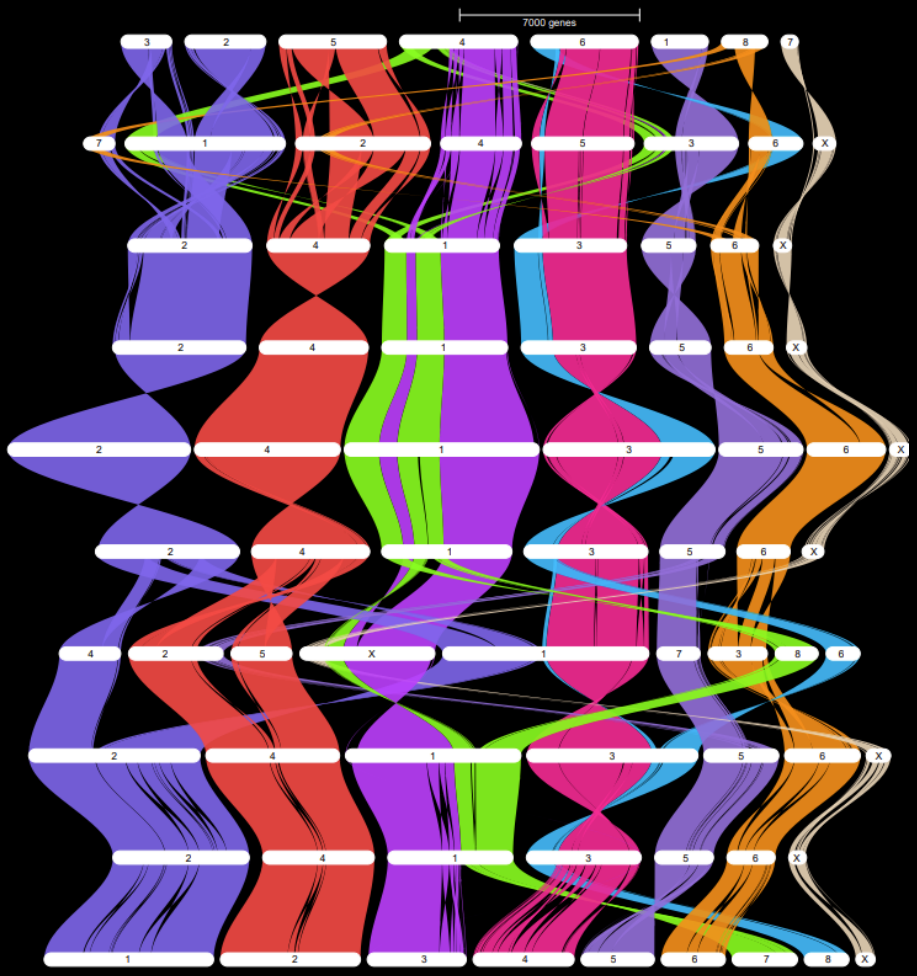
DEATH, DISEASE, AND DNA DOWN UNDER:

EXPLORING THE GENETIC INTERPLAY OF MARSUPIALS AND THEIR DISEASES



CLEOPATRA PETROHILOS

FACULTY OF SCIENCE
UNIVERSITY OF SYDNEY
2025

A thesis submitted to fulfil requirements for the degree of Doctor of Philosophy

This research was supported by an Australian Government Research Training Program (RTP) scholarship.

STATEMENT OF ORIGINALITY

This is to certify that the content of this thesis is my own work. This thesis has not been submitted for any other degree or purpose.

I certify that the intellectual content of this thesis is the product of my own work, and that all assistance received in preparing this thesis and all sources have been acknowledged.

Cleopatra Petrohilos 25/08/25

ACKNOWLEDGEMENTS

Thank you to my supervisors Emma, Carolyn and Kathy for the endless support, encouragement and inspiration. Thank you, Emma, for always regularly finding time to meet and discuss ideas, no matter how busy you are. Thank you for making things fun and exciting, for never letting me feel like a burden. You have always gone above and beyond and I would have never have made it into a PhD, let alone submitted this thesis without you. Thank you, Carolyn, for making me feel like every problem can be solved. Thank you for always encouraging me to take advantage of every opportunity that comes my way, for believing in me and most importantly always being honest with me. I always know I can trust you. Also, thank you for not minding me decorating the office with my home-made taxidermy and for putting me in the best hotels when cancelled flights leave me stranded interstate. Thank you, Kathy, for your boundless enthusiasm and positivity. I am in awe of everything you achieve. Thank you for always supporting me whatever direction I want to take and for always being my biggest champion.

Thank you to all past and present members of the Australasian Wildlife Genomics Group who always found time to help me, encourage me and most importantly, have a laugh with me. Thank you to all the post-docs, especially Kimberley and Luke for the endless patience and support. Thank you to all the students that I shared an office with over the years, especially my ever-supportive friends Adele, Lucy, Simon and Daniel. Your encouragement and confidence in me got me through, when I didn't have any confidence in myself.

Thank you to my friends for all the adventures that kept me sane. In no particular order, a special thank you to: the library ladies Maddie, Odette, Kerry, Julie and Bronwyn for all the coffees in the park; the Kitten Licks trivia team for all the beers, debates and victories; Murray and Kimberley for all the hikes in the mountains; Ros and Ev for all the games nights; everyone at City Strength for helping me stay strong; Erin for always being the

cleverest little bunny; and Deb for being by my side as we slayed all the dragons that crossed our path.

Thank you to all my friends and colleagues at Woollahra Library for supporting me ever since I decided to start collecting science degrees as a particularly expensive and time-consuming hobby. Thank you for (mostly) not complaining when I hijacked the tearoom freezer with dead insects during that entomology subject. Special thank you to my supervisors Fiona, Denise and Vicki for supporting me to take a year off to do honours; Kerry-Ann for still giving me shifts and Susan for being such an amazing desk buddy all those years.

Thank you to my family and especially my parents, Peter, Jono and Rita for the unconditional love and unwavering encouragement to be myself. Thank you for always supporting my increasingly odd life choices over the years but always and above all my education.

Lastly, thank you to Brendan for never being dubious, for knowing I would succeed even when I didn't. Life's a joke and we are all punchlines.

The work presented in this thesis was funded and supported by the ARC Centre of Excellence in Peptide and Protein Science (CE200100012), the ARC Linkage Project (LP180100244), the NCRIS-funded Bioplatforms Australia Threatened Species Initiative and an Australian Government Research Training Program Scholarship.

AUTHORSHIP ATTRIBUTION STATEMENT

This thesis consists of a general introduction followed by three data chapters and a general discussion with each research chapter formatted for submission to academic journals. Two chapters (Chapter 3 and 4) have been published, and Chapter 2 is under review at *Molecular Biology and Evolution*. The candidate is the principal author for each of these papers, and the inclusion of these published manuscripts follows the University of Sydney's Thesis and Examination of Higher Degrees by Research Policy 2015 guidelines for a thesis with publications. In the published manuscripts, table and figure numbering has been updated to reflect the chapter.

Chapter 3 of this thesis includes the publication: Petrohilos, C., Peel, E., Silver, L. W., Belov, K., & Hogg, C. J. (2025). AMPed up immunity: 418 whole genomes reveal intraspecific diversity of koala antimicrobial peptides. *Immunogenetics*, 77(1), 11.

Cleopatra Petrohilos, the candidate, designed the study with guidance from Emma Peel and Luke W. Silver. Cleopatra Petrohilos and Luke W. Silver performed all data analysis with assistance from Emma Peel. Carolyn J. Hogg and Katherine Belov sourced funding and undertook project management and supervision. Cleopatra Petrohilos wrote the main manuscript text with feedback and revisions provided by all authors.

Chapter 4 of this thesis includes the publication: Petrohilos, C., Peel, E., Batley, K. C., Fox, S., Hogg, C. J., & Belov, K. (2025). No Evidence for Distinct Transcriptomic Subgroups of Devil Facial Tumor Disease (DFTD). *Evolutionary Applications*, *18*(4), e70091.

Carolyn J. Hogg and Katherine Belov conceptualised and sourced funding for the study. Cleopatra Petrohilos, the candidate, Emma Peel, Kimberley C. Batley, Carolyn J. Hogg and Katherine Belov designed the study. Samantha Fox either collected, or coordinated

collection, of the samples and provided access to individual metadata. Cleopatra Petrohilos and Kimberley C. Batley performed RNA extractions. Cleopatra Petrohilos performed analyses and wrote the main manuscript, with feedback and revisions provided by all authors.

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Cleopatra Petrohilos 25/08/25

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Katherine Belov 22/8/25

ARTIFICIAL INTELLIGENCE

ChatGPT 3.5 (https://chatgpt.com/) was used for assistance in some coding and debugging, with all scripts were carefully evaluated before use.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PRESENTATIONS

Petrohilos, C., Patchett, A., Hogg, C.J., Belov, K. and Peel, E. *Tasmanian devil cathelicidins exhibit anticancer activity against Devil Facial Tumour Disease (DFTD) cells.* Evolutionary Biology and Ecology of Cancer Summer School, Cambridge, UK, 13th -17th June 2022; poster presentation.

Petrohilos, C., Peel, E., Batley, K.C., Hogg, C.J., and Belov, K. *The Influence of Translocations on Devil Facial Tumour Disease Evolution*. Faculty of Science, School of Life and Environmental Sciences HDR Showcase, Sydney, November 11th 2022; oral presentation.

Petrohilos, C., Peel, E., Batley, K.C., Fox, S., Hogg, C.J., and Belov, K. *Attack of the Clones: No Evidence for Distinct Phenotypic Subgroups of Devil Facial Tumour Disease (DFTD).* Faculty of Science, School of Life and Environmental Sciences HDR Showcase, Sydney, November 6th 2023; poster presentation.

Petrohilos, C., Peel, E., Batley, K.C., Fox, S., Hogg, C.J., and Belov, K. *Attack of the Clones: No Evidence for Distinct Phenotypic Subgroups of Devil Facial Tumour Disease (DFTD).* Australasian Evolutionary Society Annual Conference, Adelaide, 13th – 15th December 2023; oral presentation.

Petrohilos, C., Peel, E., Silver, L.W., Belov, K. and Hogg, C.J. *AMPed Up Immunity: 418 Whole Genomes Reveal Intraspecific Diversity of Koala Antimicrobial Peptides.* CIPPS EMCR Retreat, Kingscliff, 7th – 10th May, 2024; poster presentation.

Petrohilos, C., Peel, E., Silver, L.W., Belov, K. and Hogg, C.J. *AMPed Up Immunity: 418 Whole Genomes Reveal Intraspecific Diversity of Koala Antimicrobial Peptides.* Genetics Society of Australasia Annual Conference, Sydney 30th June – 3rd July, 2024; oral presentation.

Petrohilos, C., Peel, E., Silver, L.W., Hogg, C.J. and Belov, K. *When Cells Rebel: a comparative genomics investigation into marsupial cancer susceptibility*. CIPPS Annual Symposium, Sydney, 25th – 28th November, 2024; oral presentation.

Petrohilos, C., Peel, E. Silver, L.W., Hogg, C.J. and Belov, K. *When Cells Rebel: a comparative genomics investigation into marsupial cancer susceptibility*. Australasian Evolutionary Society Annual Conference, Perth, 4th – 6th December, 2024; oral presentation.

Petrohilos, C., Peel, E., Hogg, C.J. and Belov, K. *Death, Disease and DNA Down Under: exploring the genetic interplay of marsupials and their diseases*. Faculty of Science, School of Life and Environmental Sciences HDR Showcase, Sydney, June 6th, 2025; oral presentation.

Petrohilos, C., Peel, E., Hogg, C.J. and Belov, K. *When Cells Rebel: investigating marsupial cancer susceptibility*. Australian Mammal Society Conference, Toowoomba, June 23rd – 25th, 2025; oral presentation.

Petrohilos, C., Peel, E., Hogg, C.J. and Belov, K. *When Cells Rebel: using comparative genomics to investigate cancer susceptibility in marsupials*. Society of Molecular Biology and Evolution Conference, Beijing, China, July 20th – 24th, 2025; poster presentation.

Petrohilos, C., Peel, E., Hogg, C.J. and Belov, K. *Death, Disease and DNA Down Under: exploring the genetic interplay of marsupials and their diseases*. CIPPS Annual Symposium, Gold Coast, 1st – 4th September, 2025; oral presentation.

Petrohilos, C., Peel, E., Hogg, C.J. and Belov, K. *When Cells Rebel: using comparative genomics to investigate cancer susceptibility in marsupials*. CIPPS Annual Symposium, Gold Coast, 1st – 4th September, 2025; poster presentation.

AWARDS

Research Training Program stipend, 2022 – 2025

Bursary to attend Evolutionary Biology and Ecology of Cancer Summer School in Cambridge, UK

School of Life and Environmental Sciences HDR Showcase, 1st place for oral presentation, 2022

Australasian Evolutionary Society Travel Grant, 2023

Australasian Evolutionary Society Conference, top 5 student talks, 2023

CIPPS EMCR Retreat, people's choice for poster presentation, 2024

CIPPS Annual Symposium, 1st place for oral presentation, 2024

Australasian Evolutionary Society Travel Grant, 2024

Australian Mammal Society Student Travel Award, 2025

Australian Mammal Society, John Seebeck Travel Award for the best student talk or poster presented at the 2025 Australian Mammal Society Meeting who also received a travel award, 2025

School of Life and Environmental Sciences HDR Showcase, 1st place for oral presentation, 2025

CIPPS Annual Symposium, 2nd place for oral presentation, 2025

ABSTRACT

Wildlife diseases present a major threat to biodiversity worldwide. When exacerbated by factors such as climate change, habitat loss and genetic bottlenecks, they can result in species extirpation or even extinction. This has major flow-on effects to ecosystem function such as trophic cascades. Many wildlife diseases can also lead to spillover events in livestock and humans.

Marsupials are a clade of mammals that are affected by two major wildlife diseases, chlamydiosis and devil facial tumour disease (DFTD). Chlamydiosis is a bacterial infection that has had a devastating effect on koala (*Phascolarctos cinereus*) populations, with symptoms that include blindness and infertility. DFTD is a contagious cancer that affects the Tasmanian devil (*Sarcophilus harrisii*). It is almost always fatal and has led to population crashes of up to 80% across the state. In addition, other members of the Dasyuridae family to which devils belong, such as quolls (*Dasyurus* spp) and the kowari (*Dasyuroides byrnei*), have been reported to have a particularly high susceptibility to non-contagious cancer.

Genetics and genomics can provide valuable insight into how both host and disease co-evolve. This is particularly vital for conservation management of threatened species such as the koala and the devil. Recent advances in sequencing and computational algorithms have enabled researchers to explore immunogenetics at a higher resolution than ever before. Previous studies were often limited to studying microsatellites or a small number of genes, whereas we can now investigate multiple gene families across the entire genome, both between and within species.

A reduction in sequencing costs has led to an exponential increase in multi-omics data. This includes reference genomes for a plethora of non-model organisms, whole genome sequencing for multiple individuals within a population and transcriptomic data for diseases. Here I aimed to investigate the genetic interplay of marsupials and their diseases

using this increase in genomic dataset availability at multiple scales: immunogenetic variation between host species, immunogenetic variation between host individuals, and finally immunogenetic variation among the disease itself and its phenotypic consequences. As Dasyuridae are known to be more susceptible to cancers, I use this marsupial family to investigate how immunogenetic diversity varies across marsupial species between the Dasyuruomorphia and other marsupial orders (Peramelemorphia, Diprotodontia, Microbiotheria and Didelphimorphia). Using a species with well characterised disease and known immunogenetic response, the koala, I investigate how immunogenetic diversity can vary between individuals of the same species. Finally, investigating wildlife and their diseases requires an understanding of whether genetic diversity of the disease translates into phenotypic variation.

To investigate how immunogenetic diversity varies across marsupial species, I generated reference genomes for the kowari and eastern barred bandicoot (*Perameles gunnii*). I annotated cancer related genes in these and nine other marsupial genomes then identified gene families that had undergone statistically significantly rapid expansions or contractions. I identified a novel lineage of marsupial Ras genes, a well-known family of oncogenes. This lineage exhibited two order-specific expansions: one in the dasyurid species and one in the Peramelemorphia species. As these genes were almost exclusively expressed in gonad transcriptomes, I hypothesise that they have a reproductive role and that the expansions in these two orders are due to their unique reproductive biology (supernumerary young and chorioallantoic placenta, respectively). I also hypothesise that, similar to other members of the Ras gene family, these genes may result in cancer when mutated.

To investigate within species immunogenetic diversity, I use 418 koala whole genomes to characterise antimicrobial peptide (AMP) diversity at the nucleotide, amino acid and copy number level. This represents the first comprehensive analysis of AMP diversity in a mammalian wildlife species and reveals non-synonymous single nucleotide polymorphisms that are predicted to change peptide activity. Although nucleotide diversity was higher in

northern regions of the species' range, copy number variants (CNVs) were more common in southern populations. This included duplications of PhciCATH5, a cathelicidin with known activity against chlamydia. Although chlamydia affects koalas across their entire geographic range, clinical symptoms of the disease are greater in the north. I hypothesise that chlamydia imposes a selective pressure resulting in duplications of PhciCATH5 and recommend future studies use phenotypic metadata to assess the functional impacts of this CNV.

To treat and/or manage disease in the landscape, one needs to understand if there is genetic variation within the disease itself and if this translates to phenotypic variation. Tasmanian devils and devil facial tumour disease provides a unique opportunity to investigate this question. I generate 35 DFTD transcriptomes using tumour biopsies selected from the disease's four major genotypic clades that have been previously identified. By assessing the purity of the samples and performing unsupervised clustering I was able to determine whether these different clades have significantly different gene expression profiles. However, different algorithms (hierarchical and k-means clustering) yield conflicting results, and there is also low support for either method individually. This is the first study to take into account both tumour purity and genotypic clade when assessing differences between DFTD biopsies. More importantly, I show that all four major clades largely have a similar gene expression profile. These results provide further insight into this unique disease and has implications for therapeutic development as they suggest a single vaccine or treatment approach has potential to impact a large cross section of tumours because they all function the same way.

Overall, this thesis contributes to our understanding of the genetic interplay between marsupials and their diseases. By using a combination of reference genomes and resequenced whole genomes, I was able to characterise both inter-and intraspecific diversity in lesser studied marsupial gene families involved in disease such as oncogenes, tumour suppressors and AMPs. By using transcriptomic data and machine learning, I was

able to show that DFTD genotypic diversity does not necessarily result in phenotypic diversity. Collectively, these results can be used to inform management decisions such as translocations and lay the groundwork for future studies in marsupial disease and immunogenetics. Importantly, this work demonstrates how genomic approaches can be applied at multiple scales – from individual variation to species-level comparisons – providing a framework that is broadly applicable across wildlife disease systems

ABBREVIATIONS

24-OHC – 24S-hydroxycholesterol

AMP – Antimicrobial peptide

BAM – Binary alignment map

BLAST – Basic local alignment search tool

BUSCO – Benchmarking Universal Single-Copy Orthologs

CAFE – Computational Analysis of gene Family Evolution

CNV – Copy number variant

COSMIC – Catalogue of Somatic Mutations in Cancer

CPU – Central processing unit

CTVT – Canine Transmissible Venereal Tumour

DFT1 – Devil Facial Tumour 1

DFT2 – Devil Facial Tumour 2

DFTD – Devil Facial Tumour Disease

DNA – Deoxyribonucleic acid

EBs – Elementary bodies

eWHIS – Wildlife Health Information System

FEL – Fixed effects likelihood

FKPM – Fragments Per Kilobase Million

FUBAR – A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection

Gb – Gigabase

HMM – Hidden markov model

IG – Immunoglobulin

LR – Likelihood ratio

LXR-β – Liver-X nuclear receptor-β

Mb – Megabase

Mbp – Megabase pair

MDS – Multidimensional scaling

MEME – Mixed effects model evolution

MHC – Major histocompatibility complex

MMLQ – Median minimum base quality for bases around variant

NCBI – National Center for Biotechnology Information

NGS – Next generation sequencing

NK – Natural killer

nsSNPs – Non-synonymous single nucleotide polymorphisms

NSW – New South Wales

PCR – Polymerase Chain Reaction

QLD – Queensland

RAM – Random-access memory

RBs – Reticular bodies

RDA – Redundancy analysis

RIN – RNA integrity score

RNA – Ribonucleic acid

Sb – Strand bias

SNP – Single nucleotide polymorphism

SNV – Single nucleotide variants

TB – Terabyte

TCR – T-cell receptor

TLR – Toll like receptor

TMM - Trimmed mean of M values

TPM – Transcripts per million

t-SNE – t-Distributed Stochastic Neighbor Embedding

VAF – Variant allele fraction

VCF – Variant Call Format

WGS – Whole genome sequencing

εὕρηκα!
Archimedes

The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' but 'That's funny…'
Isaac Asimov

CHAPTER ONE:

GENERAL INTRODUCTION

CHAPTER 1. GENERAL INTRODUCTION

1.1 WILDLIFE DISEASES

Wildlife diseases are increasingly being recognised as a major threat to biodiversity worldwide (Langwig et al., 2015; Machalaba et al., 2020). In recent decades, there has been a rise in such diseases, for example white-nose syndrome in bats (Hoyt et al., 2021), chytridiomycosis in amphibians (Scheele et al., 2019), canine distemper virus in black footed ferrets (Thorne & Williams, 1988) and cetacean morbillivirus in dolphins (Van Bressem et al., 2014). Not only do such outbreaks have devastating effects on individual populations, they can also have broader ramifications on ecosystem function through trophic cascades (Holdo et al., 2009; Schultz et al., 2016). This is of particular concern for species that are already vulnerable due to factors such as habitat fragmentation and climate change (Russell et al., 2020). Wildlife diseases also have the potential to impact agriculture and public health through inter-species spillover. The majority of emerging infectious diseases to affect human populations have zoonotic origins in wildlife. These diseases include severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Ebola virus, Zika virus, mpox (historically called monkey pox) and human immunodeficiency virus (HIV) (Holmes, 2022; Jones et al., 2008; Rahman et al., 2020).

Recent advances in genetics and genomics have led to a diverse range of tools for studying wildlife diseases. Genomic and transcriptomic methods can also be used to identify novel targets for vaccine development (Blanchong et al., 2016; Doolan et al., 2014). Diagnostic tests through PCR can rapidly identify pathogens, allowing for accurate surveillance and monitoring of diseases (Galli et al., 2006; Lorch et al., 2010). Molecular techniques are also used to construct transmission networks to understand how and when pathogens are spread (VanderWaal et al., 2014). Genome wide association studies have been used to identify genes and genetic variants that render certain individuals more susceptible to disease (Batley et al., 2019; Queirós et al., 2018). Characterising immunogenetic diversity

at both the individual and species level is important as this can act as a proxy for disease susceptibility. This knowledge can then directly inform conservation practices such as planning translocations, which have the potential to introduce novel pathogens to a population (Cunningham, 1996) or alter the dynamics of a disease (Aiello et al., 2014).

This is particularly important in Australia, which has one of the highest rates of mammalian extinction ever recorded (Roycroft et al., 2021). Although this is due to many factors including introduced predators, habitat loss and bushfires (Legge et al., 2023), one major threat to Australian biodiversity is wildlife diseases (Ward et al., 2021).

1.2 MARSUPIALIA

Marsupialia are one of the three major clades of mammals (monotreme, marsupial and eutherian) (Killian et al., 2001). Metatherian mammals diverged from eutherian mammals between the late Jurassic and early Cretaceous period, 110 – 160 million years ago (Bi et al., 2018). Although originating in North America, marsupials spread through South America and Antarctica, reaching Australasia by the early Cenozoic period (Goin, 2023). Australasia's isolation and lack of eutherian mammals enabled significant diversification (Baker et al., 2023; Cáceres & Dickman, 2023). Marsupials have now diversified to occupy nearly all the same ecological and morphological niches as eutherian mammals on other continents, with the exceptions of aquatic habitats and powered flight (Martin & Weisbecker, 2023; Potter et al., 2023).

Marsupials are characterized by their unique reproductive biology. They have the shortest gestation period of any mammals, ranging from 9.5 days in stripe-faced dunnart (*Sminthopsis macroura*) (Selwood & Woolley, 1991) to 52 days in Ningbing Pseudantechinus *(Pseudantechinus ningbing)* (Woolley, 1988). In contrast, eutherian gestation periods range from 19.5 days in the mouse (*Mus musculus*) (Laurie, 1946) to 22 months in the African elephant (*Loxodonta africana*) (Perry, 1953). Marsupial neonates are born highly altricial and resemble a eutherian foetus (Old & Deane, 2000). This is because eutherians complete

development embryonically through placentation, while marsupials develop in the pouch (the "marsupium" for which they are named) supported by a complex and extended lactation period (Old & Deane, 2000).

Of the seven extant marsupial orders, two orders (Paucituberculata and Microbiotheria) are restricted to South America (Beck, 2023). One order (Didelphimorphia) is found in South and Central America, with a single species (*Didelphis virginiana*) remaining in North America (Beck, 2023). Four marsupial orders are endemic to Australasia (Beck, 2023), comprising 19 families and 269 species (Baker et al., 2023). Two of the most well-recognised Australian marsupial species are the koala (*Phascolarctos cinereus*) and the Tasmanian devil (*Sarcophilus harrisii*), the largest species in the Dasyuridae family (**Figure 1-1**). In addition to their iconic status, both of these species are currently threatened.

The koala is the sole extant species in the Phascolarctidae family (Duchêne et al., 2018). A folivore that feeds exclusively on the leaves of Eucalyptus trees (Shipley et al., 2009), they occupy a range across the east and southeast of Australia (Melzer et al., 2000). Koala populations are declining in the northern parts of their range, and these populations were declared "Endangered" in 2022 by the Australian government (Environment, 2025).

The Tasmanian devil is the largest extant member of the Dasyuridae, a family of small to medium sized marsupial carnivores (García-Navas et al., 2020). The species has been endemic to the island state of Tasmania since its extirpation on the mainland 3,000 – 34,000 years ago (Brown, 2006), and plays a vital role in the ecosystem as an apex carnivore (Hollings et al., 2014). The devil has undergone multiple population crashes due to environmental and anthropogenic factors, resulting in extremely low neutral and functional genetic diversity (Brüniche-Olsen et al., 2014; Cheng et al., 2012; Jones et al., 2004; Miller et al., 2011; Morris et al., 2015).

Dasyurids range in size from the devil (7 – 12 kg) to the Pilbara ningaui (*Ningaui timealeyi*) (2 – 9.4 g) and occupy a variety of habitats including arid deserts, woodlands and

rainforests (Jackson, 2007). The smaller dasyurids are primarily insectivorous, with only the largest species (mulgara [*Dasycercus cristicauda*], kowari [*Dasyuroides byrneii*], quolls [*Dasyurus* spp] and devil) frequently feeding on vertebrates (García-Navas et al., 2020).

Figure 1-1. A. Tasmanian devil (*Sarcophilus harrisii*). Photo: Carolyn Hogg. B. Kowari (*Dasyuroides byrnei*). Photo: Elias Neideck. C. Koala (*Phascolarctos cinereus*). Photo: Carolyn Hogg

1.3 MARSUPIAL DISEASES

Two marsupial diseases of major conservation concern are chlamydiosis in koalas and Devil Facial Tumour Disease (DFTD) in devils. Both diseases have been highly studied due to their devastating effects on populations across their species' geographic range, and DFTD is of particular interest due to its uniqueness as a contagious cancer. Tasmanian devils and

other dasyurids also have a reportedly high susceptibility to non-contagious cancers, although this susceptibility has not yet been studied.

1.3.1. Dasyuridae and Cancer

Studies have repeatedly noted that dasyurids have a particularly high susceptibility to cancer compared to other animals (Attwood & Woolley, 1973; Canfield et al., 1990b; Old & Stannard, 2020; Twin & Pearse, 1986; Vincze et al., 2022). Cancer is an umbrella term that encompasses hundreds of distinct genetic diseases. Due to the molecular and pathological diversity of these diseases, Hanahan and Weinberg proposed using six "hallmarks" as a unifying framework for cancer research (Hanahan & Weinberg, 2000). The hallmarks are phenotypic traits that underpin all cancers. These are continually being revised (Hanahan & Weinberg, 2011) and currently consist of eight hallmarks and two enabling characteristics (Hanahan, 2022). Broadly, the hallmarks characterise a cancer cell by two key traits: (i) rapid, uncontrolled cell division and (ii) the ability to migrate to distant organs (metastasise).

Genetics of Cancer

Cancer is caused by genetic mutations (Stratton et al., 2009). DNA is constantly being exposed to damage, from both endogenous and exogenous sources (Torgovnick & Schumacher, 2015; Wood et al., 2001). Natural products of metabolism (for example, reactive oxygen species and nitrogen-based free radicals) can damage nucleotides and break DNA strands (Hussain et al., 2003). Exogenous chemical and physical agents (polyaromatic hydrocarbons, ultraviolet light, gamma radiation) similarly induce DNA lesions (Marnett & Plastaras, 2001). Eukaryotes have evolved a variety of efficient mechanisms to combat these mutagens, including nucleotide excision repair and mismatch repair (Mu et al., 1997). However, if these mechanisms are overwhelmed, a DNA lesion will become fixed and result in a mutation (Fuchs, 2002). Cancer arises when mutations

accumulate in the three major groups of cancer genes: proto-oncogenes, tumour suppressors or DNA repair genes (Arnal et al., 2016).

Proto-oncogenes are essential for cell growth and differentiation (Anderson et al., 1992). Mutations in these cause a gain of function and transform them into oncogenes (Chan & Feng, 2007), resulting in uncontrolled cell proliferation (Arnal et al., 2016). Tumour suppressor genes are involved in cell cycle regulation (Arnal et al., 2016) and mutations within these genes result in a loss of function, although both alleles must be affected (Levitt & Hickson, 2002). The most famous tumour suppressor gene is *p53* (Levine et al., 1991) – termed "guardian of the genome" as it is mutated in most human cancers (Efeyan & Serrano, 2007). The third group of genes that can influence cancer susceptibility are those involved in the DNA repair (Torgovnick & Schumacher, 2015). Mutations in DNA repair genes result in genome instability, accelerating the rate at which deleterious mutations accumulate (Arnal et al., 2016; Jeggo et al., 2016).

Cancer in Wildlife

Although it is evident that cancer affects most metazoans (Aktipis et al., 2015; Albuquerque et al., 2018; Boddy, Harrison, et al., 2020; Madsen et al., 2017), its specific prevalence in wildlife species is difficult to quantify (McAloose & Newton, 2009). As cancer renders organisms more susceptible to secondary causes of death such as predation or infection, the disease may go undetected (Vittecoq et al., 2013). This is compounded by the lack of diagnostic tools available in the field (Madsen et al., 2017). Most oncological data pertaining to wildlife comes from captive animals (Boddy, Abegglen, et al., 2020; Lombard & Witte, 1959; Vincze et al., 2022), which may not accurately reflect the true prevalence. For example, some species are unable to survive in captivity (Mason, 2010), introducing inherent selection bias. In addition, neoplasia has also been associated with artificial conditions in captivity such as contraception (Harrenstien et al., 1996; McAloose

et al., 2007; Munson & Moresco, 2007) and microchipping (Pessier et al., 1999; Siegal-Willott et al., 2007; Sura et al., 2011).

Despite the limitations in obtaining and interpreting data, it is evident that cancer susceptibility varies significantly between species. As larger animals with more cells experience more cell divisions, they should theoretically be exposed to more mutations and consequently have a higher risk of cancer. While larger individuals from the same species do often have higher cancer rates (Nunney, 2013), Richard Peto observed that this assumption does not always hold in inter-specific comparisons (Peto, 2016). This apparent lack of correlation between body size and cancer susceptibility is now known as Peto's paradox. African and Asian elephants (*Loxodonta africana* and *Elephas maximus*) and the bowhead whale (*Balaena mysticetus*) are three classic examples of species with extremely low cancer rates considering their size (Seluanov et al., 2018). Although this is not always true, as small animals such as bats (Hua et al., 2024), and naked mole rats (*Heterocephalus glaber*) (Aktipis et al., 2015), also have low reported rates of cancer.

At the other end of the spectrum, cancer rates appear to be high in mammalian carnivores, particularly in mammals that consume other mammals (Vincze et al., 2022). Dasyurids are often referred to as having high cancer susceptibility (Boddy, Abegglen, et al., 2020; Griner, 1979; Vincze et al., 2022), with neoplasia being the most common cause of mortality in captive devils (Peck et al., 2019). Spontaneous neoplasms have been documented in range of dasyurids including eastern quolls (*Dasyurus viverinnus*) (Straube & Callinan, 1980; Twin & Pearse, 1986), kowari (*Dasyuroides byrnei*) (Anderson et al., 1990; Hopkins & Gaynor, 1985; Vincze et al., 2022), fat-tailed false antechinus (*Pseudantechinus macdonnellensis*) (Attwood & Woolley, 1973), dibbler (*Parantechinus apicalis*) (Attwood & Woolley, 1973) and many others (Canfield et al., 1990b). (See **Table A1-1** for more details). However, unlike mechanisms of cancer resistance, mechanisms underlying cancer susceptibility are currently unknown.

The genetic mechanisms of cancer resistance can be broadly grouped into two categories: (i) more efficient DNA repair, resulting in a lower somatic mutation rate and (ii) the prevention of damaged DNA replication via mechanisms such as increased tumour suppressor genes (**Figure 1-2**).

Lower somatic mutation rate

As all tumours are caused by mutations (Stratton et al., 2009), it is likely that a species' susceptibility to cancer is largely dependent on its somatic mutation rate (Caulin & Maley, 2011). Studies have shown an inverse correlation between mutation rate and lifespan across many mammalian taxa (Cagan et al., 2022; L. Zhang et al., 2021), and specifically a positive relationship between mutation rate and cancer prevalence (Compton et al., 2025).

This is often due to more efficient DNA repair mechanisms in long-lived organisms. The bowhead whale, the mammal with the greatest absolute longevity (Seim et al., 2014), has mutations and duplications, respectively, in the DNA repair genes *ERCC1* and *PCNA* (Keane et al., 2015). The naked mole rat, the longest living rodent, has more copies of *CEBPG*, a transcription factor involved in DNA repair, and *TINF2*, a gene involved in telomere protection (MacRae et al., 2015). They also have more efficient base and nucleotide excision repair systems than mice (Evdokimov et al., 2018). Bats also have unique DNA repair mechanisms. Genes involved in DNA repair were under positive selection in the bat ancestor (Zhang et al., 2013), while extant bat species express the transporter *ABCB1* at high levels (Koh et al., 2019), involved in drug efflux and protection against genotoxic agents (Koh et al., 2019). It is hypothesized that this improved DNA response in bats evolved in response to the increased metabolic demands of flight (Zhang et al., 2013), as metabolic products such as reactive oxygen species are major contributors to DNA damage (Wiseman & Halliwell, 1996).

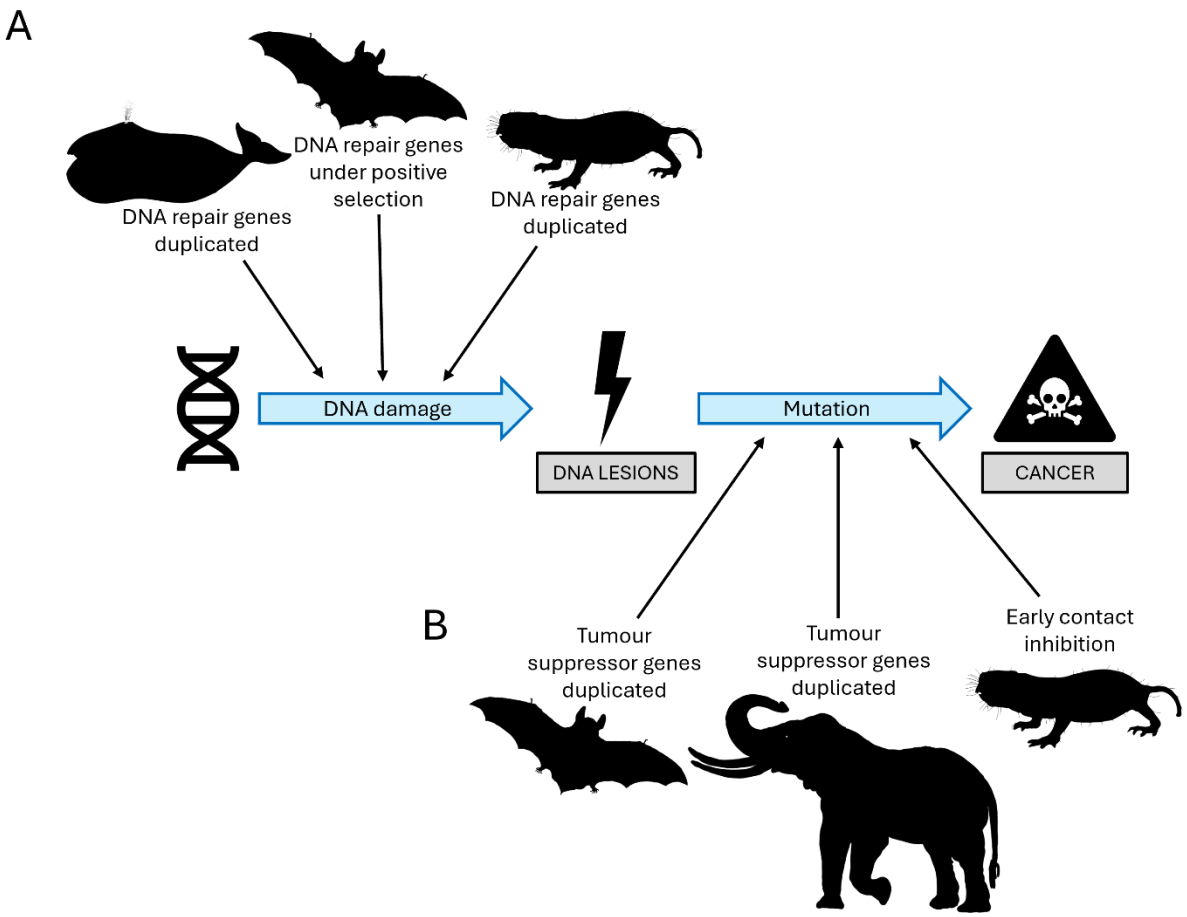


Figure 1-2. DNA damage can result in mutations leading cancer. Cancer resistant mechanisms in various species include (A) duplication or positive selection in DNA repair genes (bowhead whale, bats, naked mole rat) and (B) duplication in tumour suppressor genes (little brown bat, African elephant) and early contact inhibition (naked mole rat). Images from phylopic.org

Increased tumour suppression

If mutations cannot be prevented via DNA repair mechanisms, the mutated cell can still be prevented from replicating. This is primarily achieved via tumour suppressor genes or DNA checkpoints.

Although tumour suppressors and DNA checkpoints are present in most metazoans, many have undergone duplications and positive selection in various cancer resistant species. For example, the little brown bat (*Myotis lucifugus*) has 63 copies of the tumour suppressor

gene *FBXO31* (Caulin et al., 2015) and the African elephant has 20 copies of the tumour suppressor *TP53* (Abegglen et al., 2015). The latter is accompanied by an enhanced damage response pathway – in both African and Asian elephants, the TP53 pathway is activated in response to a lower dose of DNA damaging agents than other species (Sulak et al., 2016). For example, African elephant lymphocytes and fibroblasts exposed to ionizing agents both undergo TP53-dependent apoptosis at higher rates than human cells (Abegglen et al., 2015).

Another method of tumour suppression may be early contact inhibition (Seluanov et al., 2009). Contact inhibition refers to the phenomenon where cells cease proliferation upon reaching a certain density (McClatchey & Yap, 2012) and is regulated by a protein called hyaluronan (Itano et al., 2002). This property is lost under certain conditions, including cancer cells (Abercrombie, 1979), embryonic development and wound healing (Pavel et al., 2018). Interestingly, naked mole rat fibroblasts cease proliferation at a much lower density than other species, which is hypothesized to contribute to their cancer resistance (Seluanov et al., 2009). This may be due to their unique type and distribution of hyaluronan. While mouse and guinea pig hyaluronan is 0.5-3 MDa and human hyaluronan is 0.5-2MDa, naked mole rat hyaluronan has an extremely high molecular mass of 6-12 MDa and is highly expressed in numerous tissues compared to other species (Tian et al., 2013).

Cancer and Life History

As cancer is primarily a disease of aging (DePinho, 2000), it may be considered a manifestation of senescence. Evolutionary theory has long proposed that senescence (and consequently, cancer) is a result of the weakened selection in post reproductive years (Charlesworth, 2000). Williams developed this idea further into the concept of antagonistic pleiotropy: genes that are ultimately deleterious may be under positive selection if they confer an advantage earlier in life (Williams, 2001). A number of cancer-

related genes exhibit antagonistic pleiotropy including *Xmrk* in Xiphophorus fish (Fernandez & Bowser, 2010) and the well-known breast cancer gene *BRCA1/2* in humans (Smith et al., 2012).

Antagonistic pleiotropy is also exhibited in Kirkwood's disposable soma theory (Kirkwood & Holliday, 1979). All organisms divide their energetic resources between three major processes: basic metabolism, reproduction and somatic maintenance (**Figure 1-3**) (Holliday, 2006). The disposable soma theory describes the trade-off that exists between reproduction and maintenance: organisms that allocate more resources to reproduction have less to dedicate to somatic maintenance and are thus more susceptible to aging related degeneration such as cancer (Kirkwood, 1977). This is reflected in some studies that indicate an inverse relationship between lifespan and maximum offspring in many eutherian species (Holliday, 1994).

The relationship between specific life history traits and cancer prevalence can be difficult to resolve, as many of these traits are tightly correlated. They are often considered to fall on a spectrum that ranges from slow (low metabolic rates, long gestation, late maturity, smaller litters, longer lifespan, larger body mass) to fast (high metabolic rate, short gestation, larger litter, shorter lifespan, smaller body mass) (**Figure 1-3**) (Promislow & Harvey, 1990). Life history theory predicts that environments of low extrinsic mortality selects for slow life history strategies (traditionally called K-selection) while stochastic environments select for fast life history strategies (r-selection) (Austad, 1997; Stearns, 1976). This theory appears to confirm the early observations of Peto's paradox: large organisms invest in better somatic maintenance mechanisms such as cancer resistance. It is also supported by some comparative studies that show cancer prevalence is higher in mammals with larger litter sizes (Boddy, Abegglen, et al., 2020; Dujon et al., 2023) and longer lactation periods (Dujon et al., 2023). Dasyurids exhibit a range of these life history traits, including small body size, long lactation periods and large litter sizes.

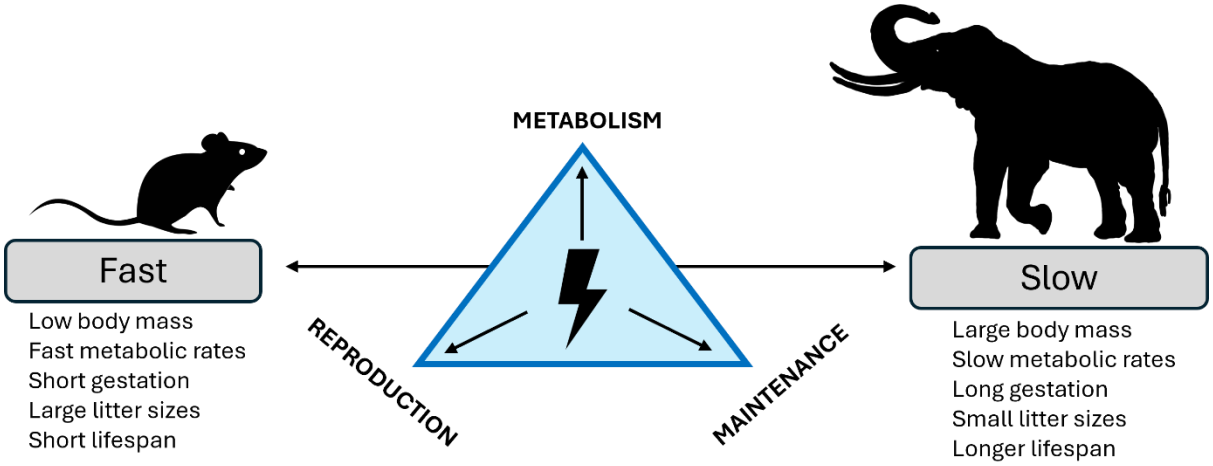


Figure 1-3. Life history traits are often correlated and occupy a continuum ranging from slow to fast. Cancer is often more prevalent in species with fast life history traits that invest less energy in somatic maintenance. Images from phylopic.org

However, other studies have highlighted the limitations of Peto's paradox. One recent study analysed 16,049 necropsy records for 292 species including mammals, amphibians and sauropsids (Compton et al., 2025). The results suggested that gestation period has a significant negative relationship with cancer prevalence; and that when controlling for gestation period, cancer prevalence may in fact increase with body mass (Compton et al., 2025). Interestingly, gestation length is not linked to body mass in some cancer resistant species such as bats, dolphins and whales (Danis & Rokas, 2024).

1.3.2. Devils and Devil Facial Tumour Disease (DFTD)

The devil was classified as "Endangered" in 2008 due to the spread of DFTD (Hawkins et al., 2008). In 2014, a second facial tumour disease was discovered, and these two clonal cancers are now referred to as DFT1 and DFT2, respectively (Pye, Pemberton, et al., 2016). DFT1 originated in north-east Tasmania and has since spread across most of the island, resulting in population crashes of up to 80% (Lazenby et al., 2018). DFT2 originated in the D'Entrecasteaux Channel Peninsula, in south-east Tasmania, and appears to be confined to this area (James et al., 2019).

DFT1 and DFT2 are both contagious cancers that arose from Schwann cells (Patchett et al., 2020). These diseases are transmitted between individuals as an allograft through biting (Hamede et al., 2013) and manifest as large masses around the face and neck (Loh et al., 2006). The growths can interfere with feeding and metastasise, resulting in death of the animal (Pye, Woods, et al., 2016). Although the two diseases are grossly similar, they are histologically and cytogenetically distinct (Pye, Pemberton, et al., 2016). Contagious cancers are exceedingly rare in nature, with only three having been observed in vertebrates (Dujon, Gatenby, et al., 2020). This is because the host immune system ordinarily recognises the MHC-I molecules on cancer cells as foreign and mounts an immune response against the cancer (Siddle, 2017). Originally, the emergence and spread of DFT1 was attributed to the low genetic diversity of devils (Siddle et al., 2007). Early studies showed extremely low MHC diversity in devils (Siddle et al., 2007), with most variation existing at the copy number level rather than the sequence level (Siddle et al., 2010).

However, subsequent studies showed both *in vitro* mixed lymphocyte reaction responses and rejection of *in vivo* skin grafts (Kreiss et al., 2011), demonstrating capacity to distinguish self from non-self, while improved genomic pipelines revealed new functional MHC diversity in devils (Cheng et al., 2022). This indicates that despite the low diversity, sufficient histocompatibility differences between host and tumour do exist. Further studies revealed that DFT1 evades host recognition through epigenetic downregulation of MHC (Siddle et al., 2013). Although DFT2 cells do express MHC-I, the most common alleles are not polymorphic (Caldwell et al., 2018). All further references to DFTD refer to DFT1, as this is the more widespread disease.

DFTD Evolution

Although DFTD was first formally observed in the north-east of Tasmania 1996 (Hawkins et al., 2006), later phylogenetic estimates suggested that it originated in 1986 (Stammnitz et al., 2023). Cytogenic analyses indicated that all tumour cells exhibit a highly rearranged karyotype (Pearse & Swift, 2006). A healthy devil genome consists of six pairs of

autosomes and one pair of sex chromosomes, while a DFTD genome consists of four pairs and one single chromosome – the remainder have been fragmented and rearranged to form five marker chromosomes (M1-M5) (Deakin et al., 2012). This unusual karyotype is a result of chromothripsis, a catastrophic shattering and rejoining of chromosomes (Deakin et al., 2012).

Early on in its history, DFTD diversified into five distinct phylogenetic clades (Clades A – E), with Clade A subsequently splitting into Clade A1 and A2 (Kwon et al., 2020). Clades A – C originated in the north-east (Kwon et al., 2020). Clade A1 spread south and Clade C spread west, while Clades A2 and B spread throughout most of Tasmania (Kwon et al., 2020) (**Figure 1-4**). Clades D and E have failed to persist, with Clade E only documented in a single individual in 2004 (Kwon et al., 2020). Clade E exhibits a highly unusual "hypermutator" phenotype – it contains many more mutations than other tumours sampled at the same time, as well as unusual mutation signatures associated with disrupted DNA repair mechanisms (Stammnitz et al., 2023). Apart from this hypermutator phenotype, it is unknown whether the different clades exhibit significant phenotypic differences.

Many DFTD cells have undergone whole genome duplication, resulting in tetraploid genomes (Stammnitz et al., 2022). This kind of polyploidy is common in asexual organisms like transmissible cancers, as it can mask deleterious mutations (Ujvari et al., 2014). However, other chromosomal abnormalities are uncommon (Deakin et al., 2012). Although early studies suggested four karyotypic strains of the disease (Deakin et al., 2012), further research indicated that these four strains had minimal microsatellite (Pearse et al., 2012) or epigenetic (Ujvari et al., 2013) differences.

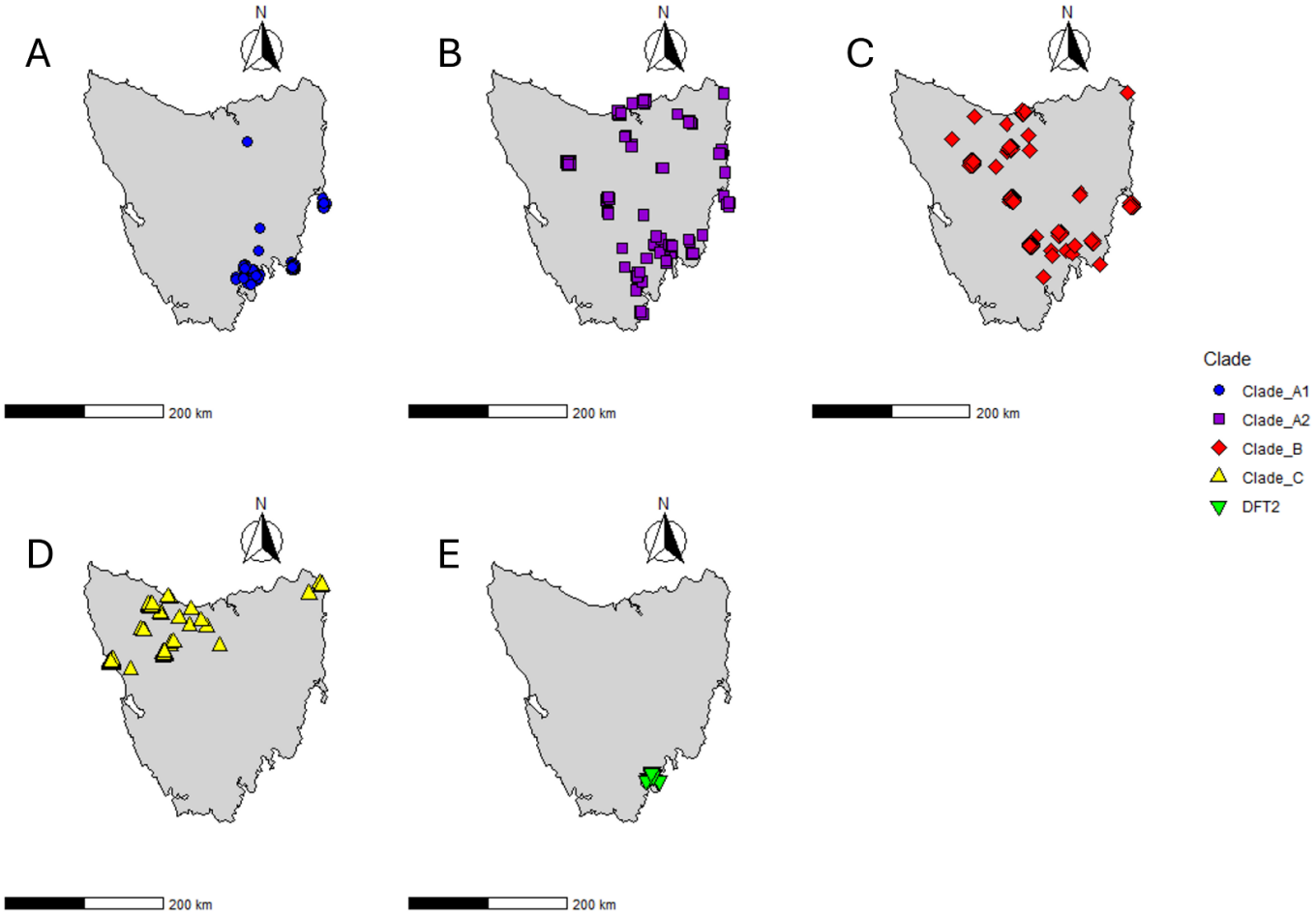


Figure 1-4. Distribution of DFTD across Tasmania: A. Clade A1, B. Clade A2, C. Clade B, D. Clade C and E. DFT2. Figure created from data collected between 2003 and 2018 (Kwon et al., 2020).

Immunogenetics of DFTD regression

DFTD is almost always fatal in devils and there is currently no treatment. However, there have been instances of tumour regression and immune response in a limited number of individuals from the north west (Pye, Hamede, et al., 2016). As DFTD regression is rare, robust statistical analysis is limited due to the small sample size. However, some studies have identified putative genes that may be involved in mounting an immune response.

Overall, no non-synonymous single nucleotide polymorphisms (SNPs) have been identified in candidate genes, with most differences within regulatory regions, suggesting that regression is due to differential expression (Margres et al., 2018). One study revealed that the putative tumour suppressor *RASL11A* was only expressed in regressed tumours, potentially due to a single point mutation in the untranslated region (Margres et al., 2020). *RASL11A* expression is also downregulated in human prostate (Louro et al., 2004) and colorectal (Wangsa et al., 2019) cancers. Another study identified a genomic region associated with DFTD regression, which contained SNPs within introns of the *PAX3* gene (Wright et al., 2017). *PAX3* is involved in angiogenesis pathways and is suggested to play an oncogenic role in a number of human cancers (Arasu et al., 2018; Boudjadi et al., 2018).

Molecular drivers of DFTD

Several molecular drivers of DFTD have been identified, such as the tyrosine kinase receptor ERBB3 (Stammnitz et al., 2018). ERBB3 is a component of the STAT3 signalling pathway (Kosack et al., 2019). Although vital for several cellular processes, when overactive this pathway is implicated in DFTD and other cancers (Gu et al., 2020; Kosack et al., 2019; Venturutti et al., 2016; Xie et al., 2004). The gene has undergone copy gains in DFTD (Taylor et al., 2017), resulting in diseased devils having higher serum levels of ERBB3 than non-diseased devils (Hayes et al., 2017). Another potential driver is the hemizygous deletion of WWC3, an element of the Hippo pathway (Stammnitz et al., 2018). WWC3 can act as a tumour suppressor by inhibiting the transcription factors YAP1

and TAZ (Ferraiuolo et al., 2017; Höffken et al., 2021), and disruption of the Hippo pathway is associated with cancer progression (Han, 2019; Harvey et al., 2013).

DFTD proliferation is also driven by altered cholesterol homeostasis (Ikonomopoulou et al., 2021). Most healthy cells use oxidative phosphorylation to generate energy and only rely on glycolysis in the absence of oxygen (Vander Heiden et al., 2009). In contrast, cancer cells use glycolysis even if oxygen is present ("aerobic glycolysis") (Vander Heiden et al., 2009) . This is known as the Warburg effect and has been shown to promote tumour progression (Barba et al., 2024). In DFTD cells, the switch to aerobic glycolysis is triggered by 24S-hydroxycholesterol (24-OHC) acting on the liver-X nuclear receptor-β (LXR-β) (Ikonomopoulou et al., 2021).

DFTD Therapeutics

Some drugs have shown potential to target these molecular drivers. Therapeutics that target elements of the STAT3 pathway have been effective at inhibiting tumour growth *in vitro* and in xenograft mouse models (Kosack et al., 2019). Atorvastatin (a statin drug used for lowering cholesterol) prevents tumour growth *in vitro* and in mouse models by inhibiting cholesterol synthesis (Ikonomopoulou et al., 2021). Other molecules including imiquimod (Patchett et al., 2016), gomesin (Fernandez-Rojo et al., 2018) and devil cathelicidins (Petrohilos et al., 2023) also show potential for drug development as they have successfully induced death of DFTD cells *in vitro*. However, nothing has yet progressed to clinical trials.

Currently, vaccine development appears to be the most promising method of managing DFTD. The most effective mode of delivery would be through a rabies-style oral bait vaccine (Flies et al., 2020). Early vaccine trials used inactivated DFTD cells that induced an immune response but did not necessarily prevent infection (Kreiss et al., 2015). Other vaccines used adjuvants that upregulated MHC expression (Pye et al., 2018; Tovar et al., 2017). Although these also resulted in antibodies and immune cell infiltration, they were

still not successful in preventing infection (Pye et al., 2021; Tovar et al., 2017). However, tumours in vaccinated devils did regress with immunotherapy (Tovar et al., 2017).

More recently, adenovirus vector-base vaccines are being investigated (Flies et al., 2020; Kayigwe et al., 2022). A human adenoviral vector that encodes devil interferon gamma (IFN-γ) has been successfully developed (Kayigwe et al., 2022). This vector was able to stimulate expression of beta-2 microglobulin, a component of MHC-I, in both devil and DFTD cells *in vitro* (Kayigwe et al., 2022).

A major goal of vaccine development is the identification of stable DFTD antigens (Flies et al., 2020). Prophylactic cancer vaccines work by targeting tumour-associated or tumour-specific antigens (Li et al., 2024; Liu et al., 2022). In the case of DFTD, such antigens may be identified by exploring "trunk" mutations in the disease. There are 1,311 substitutions that arose at the trunk of the DFTD phylogenetic tree and are present in all tumours that have been genotyped and are absent in healthy devil genomes (Stammnitz et al., 2024). Non-synonymous SNPs in these trunk mutations may result in peptides that are only expressed in DFTD cells and consequently are a valuable target for vaccines.

1.3.3. Koalas and Chlamydiosis

One major disease devastating koala populations is chlamydiosis, an infection caused by bacteria in the *Chlamydia* genus (Polkinghorne et al., 2013). *Chlamydiae* are obligate intracellular bacteria with a unique biphasic life cycle (Polkinghorne et al., 2013). In the infectious phase, they exist as extracellular and metabolically inactive particles called elementary bodies (EBs) (AbdelRahman & Belland, 2005). EBs infect host cells and are internalised in membrane bound vacuoles (AbdelRahman & Belland, 2005). Here, they differentiate into metabolically active reticular bodies (RBs) and commence the non-infectious phase of the life cycle (AbdelRahman & Belland, 2005). RBs rapidly replicate inside the host cell before differentiating back into EBs, which are released when the host cell lyses (AbdelRahman & Belland, 2005). Although chlamydiosis is generally accepted to

be a sexually transmitted disease, it may also be transmitted vertically from mother to offspring during birth or through pap-feeding (Nyari et al., 2017).

Two *Chlamydia* species have been reported to infect koalas, *Chlamydia pecorum* and *Chlamydia pneumoniae* (Polkinghorne et al., 2013). However, *C. pecorum* is more prevalent and more pathogenic (Polkinghorne et al., 2013). Infection most commonly affects conjunctiva, urinary and reproductive tracts (Polkinghorne et al., 2013) although there have also been reports chlamydia affecting the gastrointestinal tract (Burach et al., 2014; Phillips et al., 2018) and causing pneumonia (Mackie et al., 2016) and polyarthritis (Burnard et al., 2018). Infection results in a range of clinical presentations, from asymptomatic and subclinical, to severe infections causing keratoconjunctivitis resulting in blindness, urinary tract infections leading to "wet bottom" (incontinence causing staining of the fur around the rump) and reproductive tract infections that cause infertility in both males and females (Polkinghorne et al., 2013).

Immunogenetics and Chlamydia

Chlamydia prevalence and severity in koalas varies widely across their range (Quigley & Timms, 2020). Chlamydia infection in humans is often asymptomatic, and does not always progress to clinical disease (Ziklo et al., 2016). Similarly, koalas also differ in their response to the disease – some infections remain asymptomatic or are resolved without treatment, while others are recurring (Polkinghorne et al., 2013; Robbins et al., 2018). Like most infectious diseases, this variation in response is likely due to a confluence of factors such as age, sex and nutrition (Godbout et al., 2020). However, many studies have shown that immunogenetic diversity plays a strong role.

One study has shown that the MHC II allele DAB*10 was present in more *Chlamydia*-infected koalas than healthy koalas (Lau et al., 2014), although interestingly another study noted that the DAB*10 allele (as well as the MHC I allele UC01:01) was significantly more prevalent in koalas that did not progress to urinary tract disease (Robbins et al.,

2020). Two additional MHC II alleles have been identified (DCB*03 and DBB*04) that were significantly more prevalent in koalas with serious urogenital tract disease (Robbins et al., 2020). The DBB*04 variant was associated with high c-hsp60 (60 kDa chlamydial heat-shock protein) (Lau et al., 2014), while the absence of DBB*03 has been associated with clinical disease (Quigley et al., 2018). Interestingly, there is also a correlation between DBB*03 and neoplasia in koalas (Quigley et al., 2020).

MHC class I genes have also been implicated in chlamydiosis. Two MHC I genes (UA and UC) displayed different allele frequencies between koalas that resolved *Chlamydia* infection and those that did not (Silver et al., 2022). However, unlike previous studies, no association was identified between MHC class II genes and chlamydial disease. These apparently contradictory results may be because all these studies were restricted to examining one or two populations so have limited statistical power.

SNPs have also been identified in 17 candidate genes that are significantly associated with a koala's ability to resolve *Chlamydia* infection, including non-synonymous SNPs in genes involved in immune response (*IFNγ, TLR5, STAT2, RAB35*) (Silver et al., 2022). Cytokines including IFNγ, TNFα and IL10 are also expressed at higher levels in koalas with *Chlamydia* (Mathew, Beagley, et al., 2013; Mathew, Pavasovic, et al., 2013).

Chlamydia Treatment

Chlamydia in koalas is traditionally treated with antibiotics such as chloramphenicol and enrofloxacin (Blanshard & Bodley, 2008). However, chloramphenicol is not always effective in severe cases (Govendir et al., 2012) and koalas treated with enrofloxacin may continue to shed chlamydia following treatment (Black et al., 2014). In addition, many antibiotics commonly used to treat chlamydia in humans and domestic species can cause fatal dysbiosis (altered gut flora) in koalas (Booth & Nyari, 2020).

Due to the limitations of antibiotics, vaccines are also being investigated (Phillips et al., 2019). Early efforts have yielded promising results in inducing cell-mediated and antibody responses (Carey et al., 2010; Kollipara et al., 2012). Vaccine trials have also reduced the number of animals that progress to disease over twelve months (Waugh et al., 2016). However, further vaccine development is required to achieve lasting protection from the disease (Phillips et al., 2019).

1.4 MARSUPIAL IMMUNE SYSTEM

Early studies referred to marsupials as "inferior mammals" (Ashman et al., 1975) and marsupial immunity as "primitive" compared to eutherians (Jurd, 1994). However, advances in immunology have since confirmed that marsupials have a complex immune system comprising all tissues and cells found in eutherians (Belov et al., 2013).

The major difference between the marsupial and eutherian immune systems is in developmental timing. Marsupials are born without immune tissues or cells (Basden et al., 1997; Cutts & Krause, 1982; Yadav et al., 1972) and cannot mount an adaptive immune response (Old & Deane, 2003; Old et al., 2004). In addition, they complete development in a pouch with diverse and abundant bacteria, some of which may be pathogenic (Maidment et al., 2023; Ockert et al., 2024; Weiss et al., 2021). During this time, marsupial young are dependent on immune compounds in the milk and colostrum and rapid development of innate immunity (Edwards et al., 2012).

Unlike eutherians, the marsupial liver is the primary site of haematapoesis (formation of blood cells) for the first month of life, before being replaced by the bone marrow as the primary haematopoietic organ (Borthwick et al., 2014). The first lymphoid tissue to develop in marsupials is the thymus, which is involved in maturation of T cells (Belov et al., 2013). Although many marsupials only have a single thoracic thymus like eutherians, most diprotodonts have a second cervical thymus (Yadav, 1973; Yadav et al., 1972). In species with two thymuses, the cervical thymus is larger and matures first, followed by the

thoracic thymus (Yadav et al., 1972). Transcriptomic analysis indicates that both thymuses have equivalent functions (Wong et al., 2011). Marsupial secondary lymphoid tissues such as spleen and lymph nodes have similar structure and function to eutherians (Peel et al., 2019).

1.5. MARSUPIAL IMMUNOGENETICS AND GENOMICS

Marsupials also possess complete repertoires of all six major mammalian immune gene families: immunoglobulins, Major Histocompatibility Complex (MHC), T-cell receptors, toll-like receptors, cytokines, and natural killer receptors (Borthwick et al., 2014). Although, the gene completement of some families (e.g. MHC and NK receptors) differs to eutherians due to evolutionary divergence and rapid birth/death of immune genes (Nei et al., 1997). For example, marsupials have a fifth TCR chain TCRμ that has been lost in eutherians (Parra et al., 2007), and expansion of NK receptors not found in humans (Peel et al., 2022; van der Kraan et al., 2013). It is also hypothesized that exposure to microorganisms in the pouch during early life has led to expansions in some immune gene families such as antimicrobial peptides (AMPs).

1.5.1. Antimicrobial peptides (AMPs)

Antimicrobial peptides (AMPs) are a diverse group of molecules that are widely expressed across the plant and animal kingdoms (Zasloff, 2002). An ancient component of the innate immune system, they perform a vital role as the first line of defence against a broad spectrum of microbes such as bacteria, fungi, viruses and parasites (Wang GuangShun, 2017). In addition to their antimicrobial activity, they have pleiotropic functions that include both immune (chemotaxis (Yang et al., 2000), wound healing (Mangoni et al., 2016), angiogenesis (Koczulla et al., 2003) and anticancer (Petrohilos et al., 2023)) and non-immune function (e.g. sperm function (Dorin & Barratt, 2014), coat colour in dogs (Candille et al., 2007)).

AMPs are typically short, cationic polypeptides (Jenssen et al., 2006). They are synthesized as inactive precursor molecules that undergo post-translational enzymatic processing to create the active peptide (Haney et al., 2017). Their distinct pockets of hydrophobic and cationic resides allow them to adopt an amphipathic structure (Nguyen et al., 2011) that enables association with microbial cell membranes. AMPs generally cause cell death by disrupting membrane integrity (Hancock & Patrzykat, 2002) but they can also inhibit cellular functions (Chesnokova et al., 2004; Hsu et al., 2005; Park et al., 1998).

AMPs can be classified in various ways such as amino acid sequence, secondary structure, and physicochemical properties but the two major groups expressed by mammals are cathelicidins and defensins (Wang GuangShun, 2017). Although interspecific diversity of both these gene families has been well characterised in marsupials (Peel et al., 2025; Peel et al., 2024), it is not yet known what level of intraspecific AMP variation exists within a single marsupial species.

Cathelicidins

Cathelicidins are a group of AMPs that were first discovered in bovine bone marrow cells (Zanetti et al., 1993). They are synthesized as an inactive prepropeptide that is encoded by a gene containing four exons. The first three exons encode the prepro region comprising the signal peptide (exon 1) and the conserved cathelin like domain (exons 2-3), named for its similarity to the protein cathelin ("cathepsin L inhibitor") (Bals & Wilson, 2003; Zanetti et al., 1995). The prepro region is 99 to 114 amino acids long and exhibits high homology amongst species (Zanetti et al., 1995). The fourth exon encodes the active peptide region which is highly variable in both length and sequence (12 to 100 residues) (Zanetti et al., 1995).

Cathelicidins have been characterised in a range of vertebrates including fish (Maier et al., 2008), amphibians (Hao et al., 2012), reptiles (Zhao et al., 2008), birds (Cheng et al.,

2015) and mammals (Zanetti, 2005). Interestingly, the number of cathelicidins varies drastically between species. While mice (Gallo et al., 1997) and humans (Agerberth et al., 1995) encode only a single cathelicidin gene, marsupials have between five and 19 (Peel et al., 2025). Many of these marsupial cathelicidins are expressed in the pouch, milk, uterus, skin and mouth mucosa, so are hypothesized to play a role in protecting the altricial young (K. M. Morris et al., 2016; Peel et al., 2016).

Cathelicidins are expressed by neutrophils and macrophages upon activation, and constitutively expressed by epithelial cells (Kościuczuk et al., 2012; Van Harten et al., 2018). In immune cells, they are stored in granules and released extracellularly when the cell is activated (Kościuczuk et al., 2012). After secretion or degranulation, the cathelin-like domain is cleaved off by protease 3 (Sørensen et al., 2001) or elastase (Scocchi et al., 1992) to form the active peptide. Like other AMPs, cathelicidins are pleiotropic and exhibit antibacterial (Guthmiller et al., 2001), antifungal (López-García et al., 2005), antiviral (Tripathi et al., 2015), anticancer (Mahmoud et al., 2022) and immunomodulatory activity (Van Harten et al., 2018). Interestingly, some marsupial cathelicidins exhibit activity against diseases that pose a major conservation threat. Four devil cathelicidins (SahaCATH3, 4, 5 and 6) have exhibited anticancer activity against DFTD cells (Petrohilos et al., 2023) and one koala cathelicidin (PhciCATH5) has antimicrobial activity against *Chlamydia pecorum* (Peel et al., 2021).

Defensins

Defensins have been identified in animals, plants and fungi (Shafee et al., 2017). They are disulfide rich peptides that are characterised by a highly conserved six cysteine motif (**Figure 1-5**) (Ganz, 2003). Mammalian defensins can be classified as α, β or θ defensins based on the arrangement of the cysteines (Peel et al., 2024). θ defensins have only been identified in primates (Li et al., 2014) so will not be discussed.

Similar to cathelicidins, defensin gene number varies amongst species. Marsupials have a similar number of defensins to humans, with gene number ranging between 21 to 49 for β defensins and between 1 and 7 for α defensins (Peel et al., 2024). Dasyuromorph species have more α defensins than other marsupial species, suggesting lineage-specific gene expansions (Peel et al., 2024).

β defensins are considered the ancestral defensins and are thought to have arisen prior to the divergence of mammals and birds (Xiao et al., 2004). α defensins are only found in some mammals and are believed to have arisen from β defensins (Lynn & Bradley, 2007; Patil et al., 2004). Both α and β defensins are typically encoded by genes that contain two exons, although some α defensins are encoded by three (Semple et al., 2006). The first exon encodes the signal sequence and an anionic propiece (Ganz, 2003). The second exon encodes the active peptide which ranges in length from 29-35 amino acids for α defensins and 38-42 for β defensins (Ganz, 2003; Ganz & Lehrer, 1994; Lehrer & Ganz, 2002). Alpha and beta defensins can be differentiated by the disulfide bonding pattern between cysteine pairs within the active peptide. For beta defensins, disulfide bonds are formed between C1-C5, C2-C4, and C3-C6 (Torres & Kuchel, 2004). While for alpha defensins, bonds are formed between C1-C6, C2-C4 and C3-C5 (**Figure 1-5**) (De Smet & Contreras, 2005). Both alpha and beta defensins are expressed in neutrophils, epithelial cells and Paneth cells (Ganz, 2003), while beta defensins are also expressed in epithelial cells in the respiratory, gastrointestinal and reproductive tracts (Patil et al., 2005). Like cathelicidins, the mature peptide is cleaved from the precursor by proteases (Peel et al., 2024). Defensins are also pleiotropic, with broad spectrum activity against different types of microbes (Daher et al., 1986; Dhople et al., 2006; Sathoff & Samac, 2019; Silva et al., 2014), immunomodulatory (Semple & Dorin, 2012) and wound healing properties (Niyonsaba et al., 2007), as well as non-immune functions such as coat colour in dogs (Candille et al., 2007). Although marsupial defensins are expressed in a range of tissues (Peel et al., 2024), their function has not been yet been tested *in vitro*.

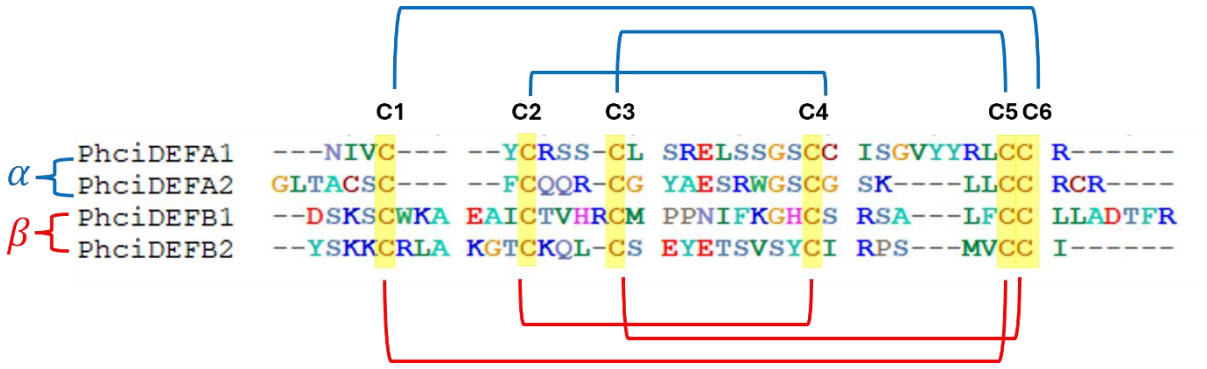


Figure 1-5. Defensin sequences for two koala a defensins and two koala b defensins. The conserved cysteines are highlighted in yellow and the disulfide bonds shown in blue (for α defensins) and red (for β defensins).

1.5.2. Marsupial Genomic Resources

Marsupial immune genes were first characterized in the 1990s by using probes to isolate cDNA (Belov et al., 1998; Belov, Harrison, et al., 1999a, 1999b; Belov et al., 2001; Belov, Harrison, Rosenberg, et al., 1999; Belov et al., 2004; Schneider et al., 1991). These methods were limited to studying one gene at a time, mostly conserved regions of immunoglobulins. More recently, the boom in genomic and transcriptomic data has allowed the entire suite of immune genes to be characterized in numerous marsupials (Belov et al., 2007; Johnson et al., 2018; Peel et al., 2022). Genomics as a discipline emerged in the latter half of the 20th century and rapidly grew from the discovery of the double helix structure of DNA (Watson & Crick, 1953) to the deciphering of the genetic code (Nirenberg & Leder, 1964) to the invention of automated sequencing (Smith et al., 1986). These developments culminated in the sequencing of the human genome in 2003 – an achievement that took 13 years and cost $2.7 billion (Wetterstrand).

The field of genomics has since grown exponentially. The reduction of cost has far exceeded Moore's law – the observation that computing power halves in cost every two years (Hayden, 2014). A draft quality genome can now be assembled for under US$1000 (Lightbody et al., 2019). The first marsupial genome to be sequenced was the grey, short-

tailed opossum (*Monodelphis domestica*) in 2007 (Mikkelsen et al., 2007), followed by the tammar wallaby (Renfree et al., 2011) and Tasmanian devil in 2011 (Miller et al., 2011). Since then, a number of consortia have been established to generate reference genomes for a plethora of non-model organisms – for example, the Earth BioGenome Project (Lewin et al., 2018), the Vertebrate Genomes Project (Rhie et al., 2021), the Bat1K Project (Teeling et al., 2018), and Oz Mammals Genomics (Eldridge et al., 2020). Reference genomes now exist for 384 marsupial species, 18 of which are at chromosome level (Challis et al., 2023).

Until recently, genome assembly was performed using data from next generation sequencing (NGS), which often necessitated a trade-off between read length and accuracy (Hon et al., 2020). Newer technologies (known as third generation sequencing) can now produce long reads (10-25kbp) that are over 99% accurate (Hon et al., 2020). However, resolving overall genomic organization remains a challenge. High levels of heterozygosity can result in haplotigs (different versions of a single locus being assembled as duplicated loci) (Puritz et al., 2024), while repetitive regions can cause local genome assembly collapse (the repeated regions collapse into a single region) (Tørresen et al., 2019). These problems can be addressed by combining third generation sequencing such as PacBio high fidelity (HiFi) with scaffolding methods such as Hi-C to enable more contiguous and accurate genome assembly (Belton et al., 2012). Although reference genomes are now available for many marsupial species, many of these are only draft quality and have not been scaffolded.

Whole-genome resequencing involves sequencing the genome of several individuals is sequenced and aligned them a reference genome (Ellegren, 2014). This is a valuable tool that allows us to analyse functional variation across multiple gene families across the entire genome (Fuentes-Pardo & Ruzzante, 2017).

1.5.3. Bioinformatic Tools

As large amounts of data began to accumulate in databases such as Genbank (Sayers et al., 2020), EMBL-EBI (Thakur et al., 2024) and UniProt (Consortium, 2019), genomics converged with information technology to form the discipline now known as bioinformatics. There are now countless bioinformatics tools available that allow researchers to investigate the interplay between immunogenetics and wildlife disease, including homology-based search algorithms and phylogenetic analysis and machine learning.

Homology-based search algorithms such as BLAST (Altschul et al., 1990) and HMMR (Eddy, 2009) are useful for annotating short, complex and highly duplicated genes such as immune gene families that are not detected by automated pipelines (Peel et al., 2022). These programs predict homology, or shared ancestry, between two sequences by identifying regions of higher similarity than can be expected by chance (Pearson, 2013). The predicted homology can then be verified by manually searching for conserved motifs – for example, the conserved cathelin-like domain of cathelicidins or the cysteine residues of defensins. Manual annotation has enabled both MHC (Silver, Hogg, et al., 2024) and AMP genes (Peel et al., 2016; Peel et al., 2021; Peel et al., 2024) to be identified across nearly all marsupial orders. In the case of cathelicidins, the peptides were then synthesized and tested *in vitro* to identify candidates for drug development.

Phylogenetic analysis has also been used to provide valuable insight into marsupial genomics. Phylogenetic analysis refers to inferring the evolutionary relationships between taxa and many programs have been developed to perform the major steps of the analysis: sequence alignment (Katoh et al., 2002; Larkin et al., 2007); substitution model (Kalyaanamoorthy et al., 2017); tree building and evaluation, using either maximum likelihood (Minh et al., 2020) or Bayesian inference methods (Bouckaert et al., 2014; Drummond & Rambaut, 2007). Phylogenetic analyses have been used to reconstruct the evolutionary history of DFTD (Kwon et al., 2020; Stammnitz et al., 2022) , predict the

sequences of ancestral marsupial AMPs (Peel et al., 2025) and identify gene families that have undergone rapid expansions and contractions in the antechinus (Brandies et al., 2020) and bilby (Hogg et al., 2024).

1.6. AIMS

This thesis aims to use comprehensive genomic and transcriptomic data and bioinformatic algorithms to investigate the genetic interplay of marsupials and their diseases. To achieve these aims, I complete the following:

Chapter 2. I examine marsupial genetic diversity at the **interspecific level** to interrogate cancer susceptibility in dasyurids. I generate reference genomes for the kowari and eastern barred bandicoot and use these in conjunction with other reference genomes to investigate the evolution of cancer-related genes in dasyurids to determine if their susceptibility to neoplasia has a genetic basis.

Chapter 3. I examine marsupial genetic diversity at the **intra-specific level** in koalas. I use 418 whole resequenced genomes to characterise population level differences in both cathelicidins and defensins across the entire geographic range of koalas and identify potential variants that may influence susceptibility to chlamydia.

Chapter 4. I investigate whether **genetic diversity** translates into **phenotypic diversity** in DFTD. I generate one of the largest DFTD transcriptomic datasets to date and perform clustering analysis to determine if phenotypic differences exist between the different genotypic clades of the disease.

This work will help inform management decisions such as translocations and lay the groundwork for future studies in marsupial disease and immunogenetics.

CHAPTER TWO:

WHEN CELLS REBEL: A COMPARATIVE GENOMICS INVESTIGATION INTO MARSUPIAL CANCER SUSCEPTIBILITY

CHAPTER 2. WHEN CELLS REBEL: A COMPARATIVE GENOMICS INVESTIGATION INTO MARSUPIAL CANCER SUSCEPTIBILITY

2.1 BACKGROUND

The research presented in Chapter 2 is currently under review at *Molecular Biology and Evolution*. The Dasyuridae are a family of carnivorous marsupials with a reportedly high cancer susceptibility. I investigated whether this susceptibility has a genetic basis by generating reference genomes for two marsupials and annotating cancer-related genes in eleven marsupial species. The analysis revealed large expansions of Ras genes in the Dasyuridae and Peramelemorphia that were almost exclusively expressed in reproductive organs.

I designed the study and performed comparative genomics analyses with guidance from Emma Peel and Luke W. Silver. Luke W. Silver and I generated genome assemblies for the kowari and Eastern barred bandicoot and genome annotations for all species. Rachel J. O'Neill and Patrick G. S. Grady generated the tammar wallaby genome. Carolyn J. Hogg and Katherine Belov sourced funding and undertook project management and supervision. I wrote the main manuscript text, with feedback and revisions on the manuscript provided by all authors.

2.2 MANUSCRIPT

When Cells Rebel: a comparative genomics investigation into marsupial cancer susceptibility

Cleopatra Petrohilos[1,2], Emma Peel[1,2], Luke W. Silver[1,2], Rachel J. O'Neill[3,4], Patrick G. S. Grady[3,5], Carolyn J. Hogg[1,2] and Katherine Belov[1,2]

[1] *School of Life and Environmental Science, University of Sydney, NSW, Australia*

[2] *Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science*

[3]*Department of Molecular and Cell Biology and Institute for Systems Genomics, University of Connecticut, Storrs, CT 06269, USA*

[4]*Department of Genome Sciences, University of Connecticut Health, Farmington, CT 060300, USA*

[5]*Colossal Biosciences, Dallas, Texas, US*

Corresponding author: carolyn.hogg@sydney.edu.au

Keywords: cancer, comparative genomics, marsupials, dasyurid

Abstract

Cancer is ubiquitous in multicellular life, yet susceptibility varies significantly between species. Previous studies have shown a genetic basis for cancer resistance in many species, but few studies have investigated the inverse: why some species are particularly susceptible to cancer. The Dasyuridae are a family of carnivorous marsupials that are frequently reported as having high rates of cancer prevalence. We hypothesised that this high susceptibility also has a genetic basis. To investigate this, we generated reference genomes for the kowari (*Dasyuroides byrnei*), a dasyurid species with one of the highest rates of reported cancer prevalence among mammals, and a non-dasyurid marsupial, the eastern barred bandicoot (*Perameles gunnii*). We used these to perform a comparative genomics analysis alongside nine previously assembled reference genomes: four dasyurid species and five non-dasyurid marsupial species. Genomes were annotated using FGENESH++ and

assigned to orthogroups for input to CAFE (Computational Analysis of gene Family Evolution) analysis to identify gene families that had undergone significant expansions or contractions in each lineage. In the dasyurids, we identified large expansions in Ras genes, a family of oncogenes. Interestingly, a similar expansion of Ras genes was also identified in the bandicoot and bilby. These genes were primarily expressed in tissues such as testes, ovaries and yolk sac, so we hypothesise they serve a reproductive role. Future work is required to identify the potential roles of oncogene expansions in cancer susceptibility in these marsupial species.

Introduction

Cancer encompasses a multitude of diseases that are all characterised by uncontrolled cell growth and metastasis (Hanahan, 2022). They arise when mutations accumulate in the three major groups of driver genes: tumour suppressors, proto-oncogenes, and DNA damage response (Arnal et al., 2016). Mutations in tumour suppressors result in a loss of function (Levitt & Hickson, 2002) whilst mutated DNA repair genes cause genome instability, accelerating the rate at which deleterious mutations accumulate (Arnal et al., 2016; Jeggo et al., 2016). Mutated proto-oncogenes cause a gain of function (Chan & Feng, 2007) and result in uncontrolled cell proliferation (Arnal et al., 2016).

Cancer is ubiquitous in multicellular life (Aktipis et al., 2015). Phylostratigraphic analysis suggests that the emergence of tumour suppressors and oncogenes coincided with the emergence of metazoans (Domazet-Lošo & Tautz, 2010). Cancer susceptibility varies greatly between species (Boddy, Abegglen, et al., 2020; Madsen et al., 2017; Vincze et al., 2022). As mutations arise from imperfect cell replication, larger animals with greater lifespans who undergo a greater number of cell divisions should theoretically be at a higher risk of mutations and consequently cancer. Yet many large-bodied mammals, such as whales (Tollis et al., 2019) and elephants (Abegglen et al., 2015), have extremely low rates of cancer. This lack of correlation between body size and cancer risk is known as Peto's paradox (Nunney et al., 2015) and has been supported by a number of studies

(Abegglen et al., 2015; Boddy, Abegglen, et al., 2020; Bulls et al., 2022; Compton et al., 2025; Vincze et al., 2022). A more recent model of cancer mortality contradicts Peto's paradox (Butler et al 2025), highlighting that larger mammals often have higher rates of malignancy. However, this relationship was logarithmic, meaning observed cancer rates were still lower than expected relative to body size, suggesting larger mammals may have evolved more efficient anticancer mechanisms.

This differential cancer susceptibility is often attributed to life history trade-offs and antagonistic pleiotropy (Boddy, Harrison, et al., 2020). Life history theory predicts that there is a trade-off between reproduction and maintenance (Kirkwood, 1977). Organisms that allocate more resources to reproduction have less to dedicate to somatic maintenance, which renders them more susceptible to aging-related degeneration such as cancer. This appears to be supported by studies that have shown mammals with large litter sizes and long lactation periods are at greater risk of cancer (Dujon et al., 2023). Antagonistic pleiotropy refers to genes that are ultimately deleterious but confer an advantage earlier in life (Ljubuncic & Reznick, 2009). An example is the *Xmrk* oncogene in *Xiphophorus* fish. Although it causes melanoma (Wittbrodt et al., 1989), *Xmrk* is also associated with traits that increase reproductive success such as a larger size (Fernandez & Bowser, 2010) and a spotted caudal melanin pattern (Fernandez & Morris, 2008).

Comparative genomic studies have shown that there is often a genetic basis for cancer resistance. Some of the most well studied oncogenes are the three canonical Ras genes (*H-Ras, K-Ras* and *N-Ras*). These are responsible for 15-20 % of human cancers (Quinlan & Settleman, 2009) and only require a single mutation at codon 12, 16 or 31 to become activated (Prior et al., 2012). First discovered in the 1960s (Harvey, 1964), *H-Ras, K-Ras* and *N-Ras* were the first identified members of the much larger Ras superfamily, with over 150 genes found across all forms of eukaryotic life (Goitre et al., 2014). The Ras superfamily is divided into five major gene subfamilies: Ras (which includes the three canonical genes), Rho, Rab, Ran and Arf (Goitre et al., 2014).

Duplication of tumour suppressors has also been linked to cancer resistance. For example, the microbat (*Myotis lucifugus*) has 63 copies of *FBXO31* (Caulin et al., 2015). This is a tumour suppressor gene that induces cell cycle arrest in response to DNA damage (Santra et al., 2009). Bats also express high levels of the transporter *ABCB1*, which provides protection against genotoxic agents (Koh et al., 2019). Similar expansions have been observed in the African elephant (*Loxodonta africana)*, which has 20 copies of the tumour suppressor gene *TP53* (Abegglen et al., 2015). The TP53 pathway is activated in the presence of genotoxic agents and, interestingly, in elephants is activated at lower doses of these agents than other species (Sulak et al., 2016). The bowhead whale (*Balaena mysticetus*), one of the longest living mammals, also has duplications of *PCNA* and species-specific mutations in *ERCC1*, both genes involved in DNA repair (Keane et al., 2015). The famously cancer-resistant naked mole rat (*Heterocephalus glaber*) has higher copy numbers of *TINF2* and *CEBPG*, two genes involved in telomere protection and DNA repair respectively (MacRae et al., 2015), and is documented to have better base and nucleotide excision repair systems than other rodents (Evdokimov et al., 2018).

However, the inverse has rarely been explored – why are some taxa particularly susceptible to cancer? One example of taxa with high cancer susceptibility is the Dasyuridae, a family of small to medium sized marsupial carnivores (García-Navas et al., 2020). Like other marsupials, they give birth to altricial neonates who complete development in the pouch (Old & Deane, 2000). Many dasyurid species give birth to supernumerary young (more offspring than can be supported by the number of available teats) (Gemmell et al., 2002; Parrott & Edwards, 2023). Cancer has been reported in many species of dasyurids (Anderson et al., 1990; Attwood & Woolley, 1973; Hopkins & Gaynor, 1985; Straube & Callinan, 1980; Twin & Pearse, 1986), with the family frequently being referred to as having a particularly high susceptibility to cancer (Attwood & Woolley, 1973; Canfield et al., 1990b; Old & Stannard, 2020; Twin & Pearse, 1986; Vincze et al., 2022). Vincze et al. (2022) observed that the kowari (*Dasyuroides byrnei*) had the highest cancer mortality risk out of the 191 mammalian species included in their

comparative study – a value so high that the species was excluded due to concerns that its high leverage would bias the model. Not only is cancer the most common cause of mortality for Tasmanian devils (*Sarcophilus harrisii*) in captivity (Peck et al., 2019), but they are one of only two vertebrates to suffer from contagious cancers (devil facial tumour disease, DFTD), and the only vertebrate to suffer from two separate contagious cancers (DFT1 and DFT2) (Metzger & Goff, 2016). However, it is unknown whether this high incidence of cancer amongst dasyurids arises from a genetic predisposition to cancer, or that they are frequently housed in captivity where artificial conditions and veterinary care lead to increased lifespans compared to individuals in the wild (Jackson, 2007; Obendorf, 1993). It is therefore possible that cancer is simply reported more often in these genera, rather than occurring more often. Another marsupial species, the koala (*Phascolarctos cinereus*), has also shown elevated rates of lymphoma and leukaemia, although this is often linked to koala retrovirus (KoRV) (McEwen et al., 2021; Tarlinton et al., 2005). Retroviruses such as KoRV are known to cause cancer in various species (Hartmann, 2012; Ruprecht et al., 2008; Tarlinton & Greenwood, 2024).

The increasing availability of scaffolded or chromosome level reference genomes has enabled a range of comparative evolutionary studies, including for marsupials, with high quality reference genomes currently available for 18 species (Challis et al., 2023). In order to conduct a more comprehensive comparative study across the marsupial family tree, we generated a long read scaffolded reference genome for the kowari (highest reported marsupial cancer susceptibility) (Vincze et al., 2022) and eastern barred bandicoot (*Perameles gunnii*). The bandicoot was chosen as representing one of the two Peramelemorphia families (Thylacomyidae and Peramelidae). The Peramelemorphia are an interesting lineage as they have an unusual reproductive biology for a marsupial (chorioallantoic placenta) and their position in the marsupial phylogeny is unresolved. Here we aimed to use these genomes, in conjunction with nine other high quality marsupial genomes, to investigate the evolution of cancer related genes in marsupials using a comparative framework.

Results and Discussion

Genome Assemblies

We sequenced the kowari and eastern barred bandicoot genome with PacBio HiFi and scaffolded using Hi-C, with the resulting assemblies 3.21Gb and 3.94 Gb in size, containing 851 and 1,462 scaffolds for the kowari and eastern barred bandicoot, respectively. In addition, we downloaded nine publicly available genomes (accession numbers and genome information, **Table A1-2Table A1-1**). Benchmarking universal single-copy orthologs (BUSCO) analysis revealed highly complete assemblies for all 11 genomes used in this study, with over 89.5% (ten above 94.7%) of single copy mammalian genes identified (**Table A1-2**). For more information on the genome assemblies, see **Supplementary Material to Chapter 2**.

Cancer Prevalence

We collated data on reports of neoplasia in marsupials from five published sources (Canfield et al., 1990a, 1990b; Effron et al., 1977; Ladds, 2009; Ratcliffe, 1933) and the Australian Wildlife Health Information System (eWHIS) (**Table A1-1** and **Table A1-3**). Dasyurids had the highest number of reported neoplasia amongst marsupials, comprising 37.35% (282/757) of all reports. This was even though eWHIS may not include all cases of neoplasia in captive Tasmanian devils (Cox-Witton, K., *pers. comm.*). Within dasyurids, the highest reports of neoplasia occurred in the Tasmanian devil, three quoll species (*Dasyurus hallucatus*, *Dasyurus maculatus*, *Dasyurus viverrinus*) and kowari (**Table A1-3**). Koalas had the second highest rates amongst marsupials (27.6%, 209/757), although most of these were from the most recent source (eWHIS, 2008-2024). Nearly all reports (105 out of 113) were koalas from Queensland and New South Wales, which may reflect the higher prevalence of KoRV in these northern regions (Simmons et al., 2012). For this reason, cancer genes were investigated in both dasyurid and koala ancestors.

Gene Family Evolution

First, we annotated 11 marsupial genomes using FGENESH++v7.2.2 (Solovyev et al., 2006), which annotated between 28,365 and 76,963 genes in all 11 species (**Table A1-4**). We then looked for orthologs of known cancer genes within rapidly evolving orthogroups to gain insight into genomic mechanisms underpinning cancer risk. We identified orthologs of all genes with mutations implicated in human cancer ($n = 753$), from the Cosmic Cancer Gene Census v101 GRCh38 (Sondka et al., 2024) using a reciprocal blast best hit analysis. We identified between 576 and 674 orthologs of these cancer related genes in each marsupial genome (**Table A1-4**).

Genes were also assigned to orthogroups (sets of genes descended from a single ancestral gene) using Orthofinder v2.4.0. In total, 260,261 annotated genes (84.3%) were assigned to 24,416 orthogroups amongst the 11 species. 7,640 (31.29%) of these orthogroups had all species present and 2,999 (12.28%) were single copy orthologs. The opossum had the lowest percentage of genes (68.8 %) in orthogroups, which is expected as it was the phylogenetic outgroup (**Table A1-4**).

We undertook a Computational Analysis of gene Family Evolution (CAFE) (Mendes et al., 2020) analysis to identify rapidly evolving orthogroups across marsupial lineages. Across the five marsupial orders (Diprotodontia, Dasyuromorphia, Peramelemorphia, Microbiotheria, Didelphimorphia), CAFE identified 229 orthogroups that had undergone statistically significant expansions and 115 that had undergone statistically significant contractions (**Figure 2-1A**). The orthogroups in the koala and dasyurid ancestors were then examined in more detail because of the high cancer prevalence in these two lineages.

CAFE identified 41 significant expansions and 10 significant contractions in the koala ancestor. However, only five gene families contained cancer related genes from the COSMIC database and all were from the major immune gene families T cell receptors

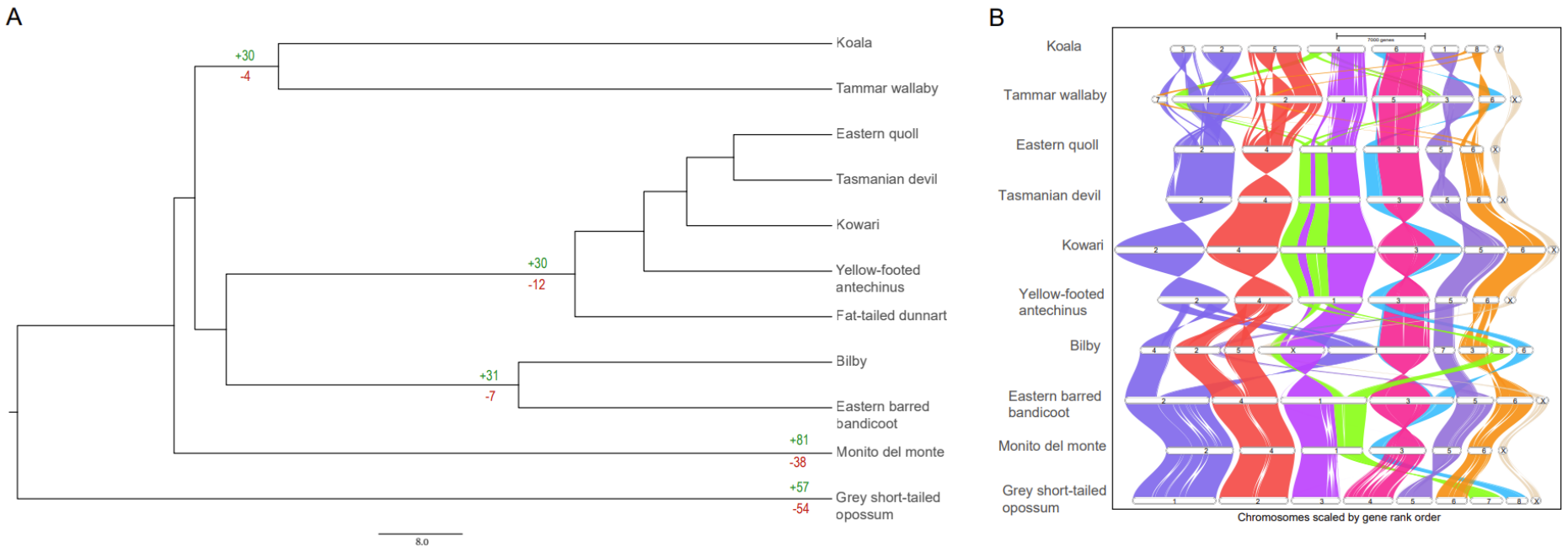


Figure 2-1A. For each marsupial order, the number in green above the branch represents the number of orthogroups deemed to have undergone a statistically significant expansion by CAFE and the number in red beneath the branch represents the number of orthogroups deemed to have undergone a statistically significant contraction. The figure was created by manually annotating CAFE output onto the time calibrated phylogeny (see methods for more detail about how phylogenetic tree was generated). The scale bar represents 8 million years. **B**. Synteny plots generated by GENESPACE. The coloured blocks represent chromosome scaffolds for each species. The opossum was chosen to represent chromosome order as a model for the ancestral species. The dunnart is not included in this plot as the genome was fragmented.

(TCRs) and immunoglobulins (IGs). Two of the significant gene families in the dasyurid ancestor also involved TCRs and IGs (see below).

Automated pipelines, such as FGENESH++, are known to be unreliable at annotating immune genes, especially those that are highly polymorphic and rapidly evolving such as the TCR and IG variable regions (Peel et al., 2022). The major immune gene families have also been well characterised and compared in marsupials and the constant region gene numbers are conserved. Hogg et al. (2024) identified similar numbers of IG and TCR constant genes in the bilby, devil, short-tailed opossum, koala and woylie (ranging from 11-20 for IGs and 9-14 for TCRs). Both TCRs and IGs consist of a small number of constant regions. They also contain a number of variable and joining regions that undergo somatic recombination. This results in a high level of diversity that enables them to recognise and bind to a wide variety of antigens. Previous studies have indicated the koala has a higher number of IG variable region genes (289 compared to 226 in the woylie and 116 in the bilby), although this high number is attributed to genome quality rather than gene expansion (Peel et al., 2022). Further, gene expansions and contractions are how immune gene families evolve. As immune genes have a basic biological function, it is difficult to disentangle this from their involvement in cancer. For these reasons, orthogroups containing IGs and TCRs were not investigated further.

In the dasyurid ancestor, CAFE identified 30 significant expansions and 12 significant contractions. Only six of these gene families contained cancer related genes from the COSMIC database, and two of these involved TCRs and IGs and so were not investigated further. Three of the remaining four orthogroups underwent contractions in subsequent lineages. For example, orthogroup HOG000767 contained putative orthologs of *ATRX*, a tumour suppressor. Although this underwent a significant expansion in the dasyurid ancestor, it subsequently underwent contractions in the ancestor to the quoll and devil (**Figure A1-3**). Orthogroup HOG0001427 contained putative orthologs of *SLC34A2*, a tumour suppressor also involved in oncogenic fusion. This underwent contractions in the

quoll and kowari (**Figure A1-4**). Orthogroup HOG0002169 contained orthologs of *TCEA1*, a gene involved in oncogenic fusions and subsequently underwent contraction in the devil (**Figure A1-5**). As these orthogroups were not consistently high in all dasyurid species, they were not investigated further.

The remaining orthogroup containing cancer genes (HOG0000752) underwent significant expansion in the dasyurid ancestor and had much higher numbers of genes in all extant dasyurids (average 12 ± 4.3 standard deviation) compared to the other marsupials in this study (average 0.67 ± 1.2) (**Figure 2-2**). In this context, statistical significance was determined by the CAFE analysis. This orthogroup contained genes annotated as *K-ras*, a well-known oncogene in the Ras gene family.

K-ras is highly conserved across all jawed vertebrate species and is usually present in a single copy within the genome and transcribed in two isoforms (*K-Ras4A* and *K-Ras4B*). All dasyurids and tammar wallaby were predicted to encode between three and 18 copies of genes within this orthogroup (**Figure 2-2**). In addition to containing more than one copy, the marsupial genes in this orthogroup showed much higher differentiation compared to *K-ras* from other mammals, birds and amphibians (**Figure 2-3**). The amino acid sequence similarity amongst non-marsupials ranged from 94.15% (between the African clawed frog and zebra fish) to 100% (between the human and chicken). The sequence similarity amongst marsupials ranged from 68.75% (between the dunnart and the quoll) to 97.12% (between the devil and the kowari). The similarity between marsupials and non-marsupials ranged from 58.54% (between the dunnart the African clawed frog) to 68.75% (between the antechinus and the human, chicken and zebrafish).

Novel Ras Genes Discovered in Marsupials Annotation

As CAFE identified expansions in the *K-Ras* gene family in the dasyurid ancestor, we chose to further investigate other members of the Ras subfamily. First, to check if the

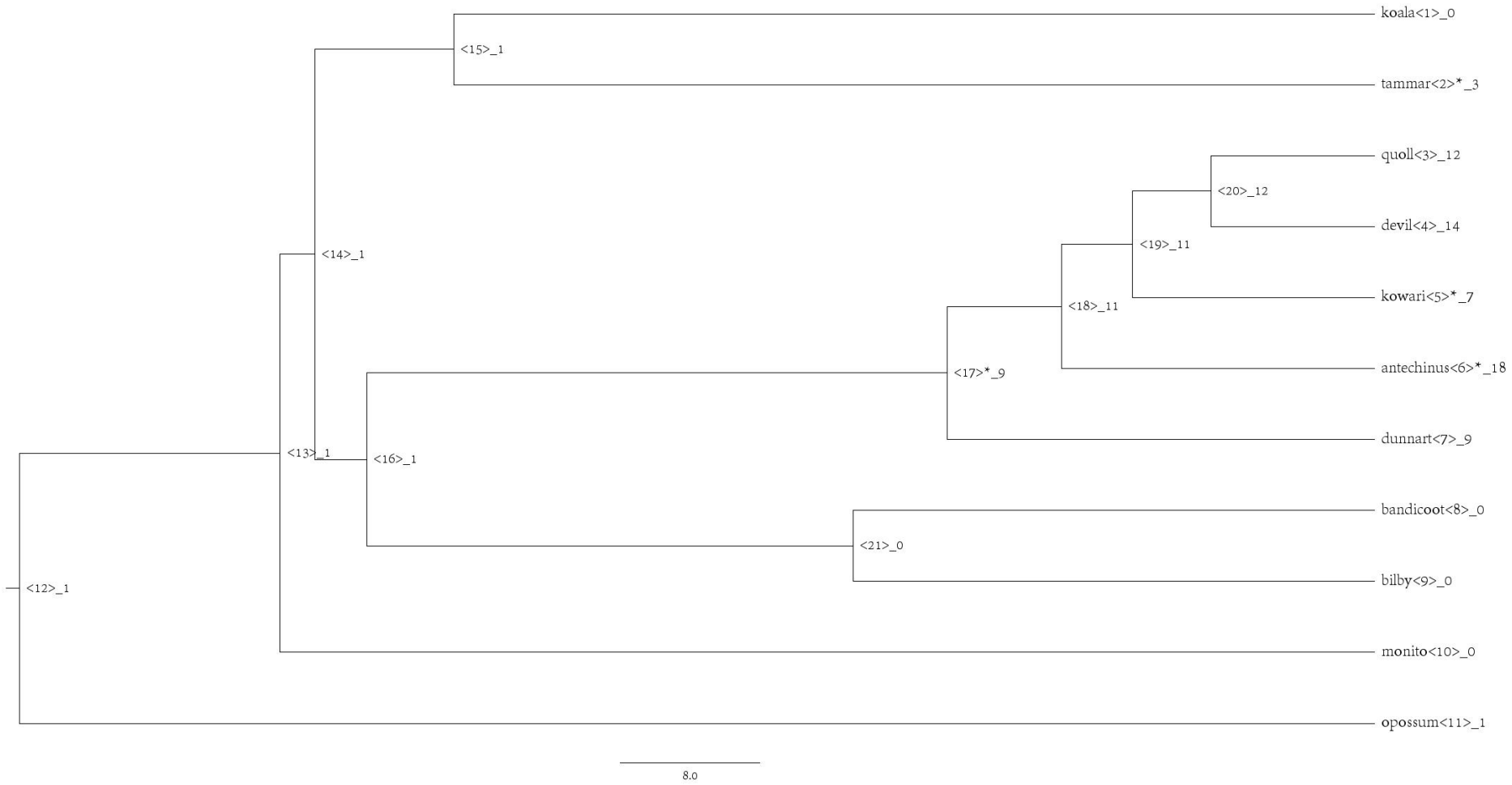


Figure 2-2. Gene tree for orthogroup HOG0000752 annotated as K-ras. The number after the underscore represents the number of genes within the orthogroup encoded in the genome of each species (or is predicted to encode, in the case of ancestral nodes). Asterisk indicates that a statistically significant expansion or contraction occurred in this lineage, as seen in the dasyurid ancestor.

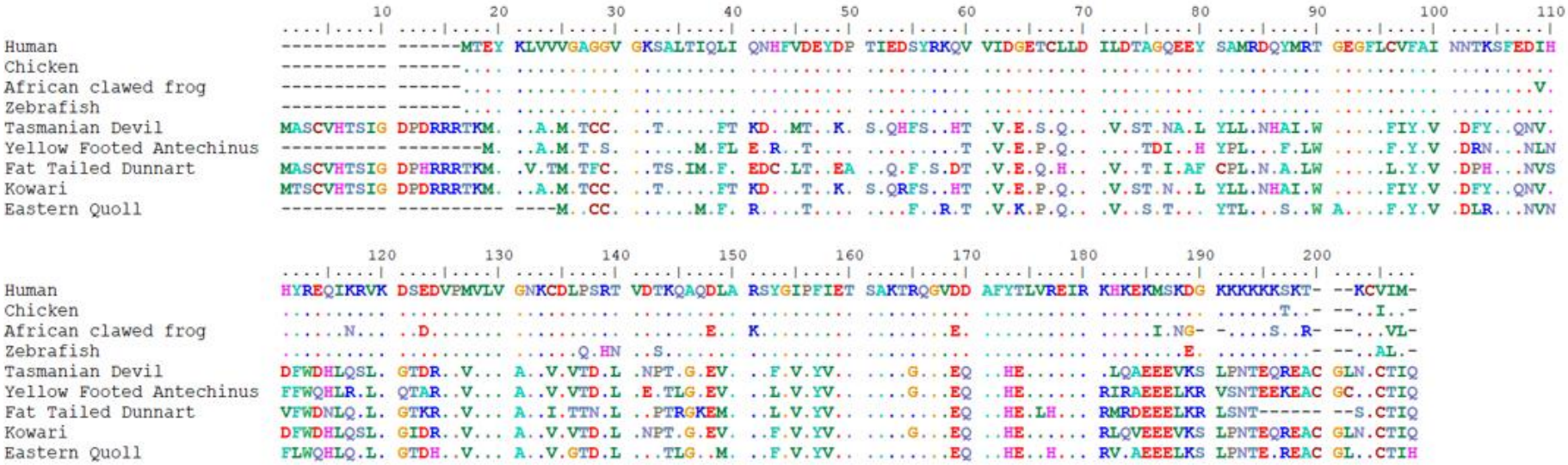


Figure 2-3. Alignment of a subset of marsupial genes annotated as K-ras by FGENESH++ with K-ras genes from eutherian, amphibian, avian and fish species: humans (*Homo sapiens*, NP_001356716.1), chicken (*Gallus gallus*, NP_001243091.1), African clawed frog (*Xenopus laevis*, NP_001081316.1) and zebrafish (*Danio rerio*, NP_001003744.1). The marsupial species are the Tasmanian devil, yellow footed antechinus, fat-tailed dunnart, kowari and eastern quoll. Dots represent 100% amino acid identity to human. All sequences showed high similarity (> 58.6% similarity using the BLOSUM62 similarity matrix) but there were distinct differences between the marsupial sequences and those from other taxa. Only a subset of genes is shown for ease of viewing, as the orthogroup contained 60 dasyurid genes.

automated annotations were correct, we manually annotated the four canonical Ras genes (*HRas, NRas, KRas4A* and *KRas4B*) in all 11 study species. A single copy of *HRas, NRas, KRas4A* and *KRas4B* were identified in all marsupial species in this study (included in Supplementary File 1 available online here). In addition to these, an almost exact duplicate of *Kras4B* was also annotated in the opossum, with only a single amino acid change (K>Q at 101). This duplicate was encoded by a single exon as opposed to the four exons in the other *K-ras* genes. However, the marsupial genes that FGENESH++ had annotated as *K-Ras* were not amongst these canonical genes, indicating they may be novel members of the Ras gene family.

To confirm that these genes were novel and not orthologs of other members of the Ras gene subfamily, we chose to manually annotate the remaining 33 genes in the Ras subfamily in the Tasmanian devil (the Ras subfamily being one of the five subfamilies that comprise the Ras superfamily). We selected the devil as a representative marsupial as it had the most complete genome amongst the dasyurids. Out of these 33 other genes in Ras subfamily, 29 were identified in the devil (included in Supplementary File 1 available online here), the four we were unable to identify were *ERas, RERGL, RHEBL1* and *DIRAS3*. *ERas* is an unusual gene within the Ras subfamily as it is only encoded by a single exon, the only Ras gene without paralogs and has so far only been characterised in eutherian mammals (De Falco et al., 2022; Roperto et al., 2017; Takahashi et al., 2003; Tanaka et al., 2009). Our results provide further support that *Eras* is not only mammal specific but eutherian specific.

In addition to these genes, we identified a distinct phylogenetic clade of putative novel marsupial *Ras* genes that belonged to the Ras subfamily but did not show orthology with any of the other vertebrate *Ras* genes (**Figure 2-5**). Some of these had been incorrectly annotated by FGENESH++ as one of the classical Ras genes (*H-Ras, K-Ras* or *N-Ras*). The genes in this clade were all encoded by a single exon and ranged in length from 155 to 390 amino acids. All putative novel Ras genes contained an open reading frame, a RAS smart domain and at least partial matches for four out of the five conserved G box motifs

characteristic of the Ras superfamily (**Figure A1-6**). The top reciprocal BLAST hit for all genes was a canonical Ras gene, yet the novel genes contained a more diverse sequence compared to eutherian sequences (**Figure 2-3**). These genes were almost exclusively expressed in marsupial gonads (**Table 2-1**, Supplementary File 2 available online here) and so were named in the order in which they were found using the first letters of the genus and species then MgRas (for marsupial gonad Ras gene). The protein sequences for all genes are included in Supplementary File 1 available online here.

Phylogenetic analysis revealed strong support for the MgRas genes being within the Ras subfamily of the Ras superfamily (**Figure 2-5**). The MgRas genes and the majority of non-canonical Ras genes each form a monophyletic clade with strong bootstrap support (100% bootstrap) and form a single clade (98% bootstrap support) that sits sister to ERas genes (100% bootstrap). This combined group is, in turn, sister to the classical Ras genes *H-Ras*, *K-Ras* and *N-Ras* (100% bootstrap). Within each clade (canonical, ERas, MgRas, and non-canonical), genes cluster in orthologous groups amongst species, particularly evident for Ras subfamily and canonical Ras clades. Although, within the MgRas clade, distinct order-specific expansions were also evident in the Dasyuromorphia (antechinus, devil, dunnart, kowari and quoll) and the Peramelemorphia (bandicoot and bilby), with all other marsupial species having either two or three MgRas orthologs (**Figure 2-5**).

Genomic Organisation and Synteny

There was strong support (>98% bootstrap) for three phylogenetic clades (termed A, B and C) (**Figure 2-5**) that were largely encoded in two main clusters in the genome of each species (**Figure 2-6**), as is common in genes that evolve by tandem duplication (Pan & Zhang, 2008). Genes from all three clades showed similar expression patterns.

Cluster 1 was encoded in syntenic chromosomes in all species, although in the diprotodonts (koala and tammar wallaby) and monito del monte a small number of orphan genes were identified on different chromosomes (**Figure 2-4B**). Cluster 1 contained

representatives from all phylogenetic clades, although not all clades were present in all species (**Figure 2-6A**) Fig 5A). Genes from Clade A were not identified in the monito del monte and Peramelemorphia, and genes from Clade C were not identified in Cluster 1 in the Dasyuridae (instead being found in Cluster 2). In the Peramelemorphia, Clade B underwent an extremely large expansion and Clade C underwent a smaller expansion, relative to other marsupials.

Cluster 2 was only found in dasyurids on chr 5 (scaffold 89 in the dunnart) (**Figure 2-6B**). This cluster also contained representatives from all phylogenetic clades, with Clade A having undergone a large expansion and Clade C having undergone a smaller expansion in some species.

All species had some combination of genes from all three clades either flanked or near *DLC1* (Deleted in Liver Cancer 1) and *TRMT9B* (probable tRNA methyltransferase 9B) genes, so the most parsimonious explanation is that the ancestral state included a three gene cluster in this region, which underwent subsequent expansions and contractions in the different marsupial lineages. We hypothesise the entire three gene cluster was duplicated and translocated to chromosome 5 in the dasyurid lineage, where clades A and C subsequently underwent expansions to form Cluster 2.

Both clusters were encoded near potential tumour suppressors (*DLC1, TRMT9B* and GRAM domain containing 4, *GRAMD4*) (Begley et al., 2013; John et al., 2011; Zhang & Li, 2020), with Cluster 1 entirely flanked by tumour suppressors in most species (**Figure 2-6**). Oncogenes and tumour suppressors are often in close proximity within the genome (Antonio & Widegren, 2005), and oncogenes without a neighbouring tumour suppressor (within 1.46 Mbp) are more prone to amplification (Wu et al., 2017). The two clusters both occurred near chromosome ends, which is also common in oncogenes (Antonio & Widegren, 2005; Lima-de-Faria et al., 1991).

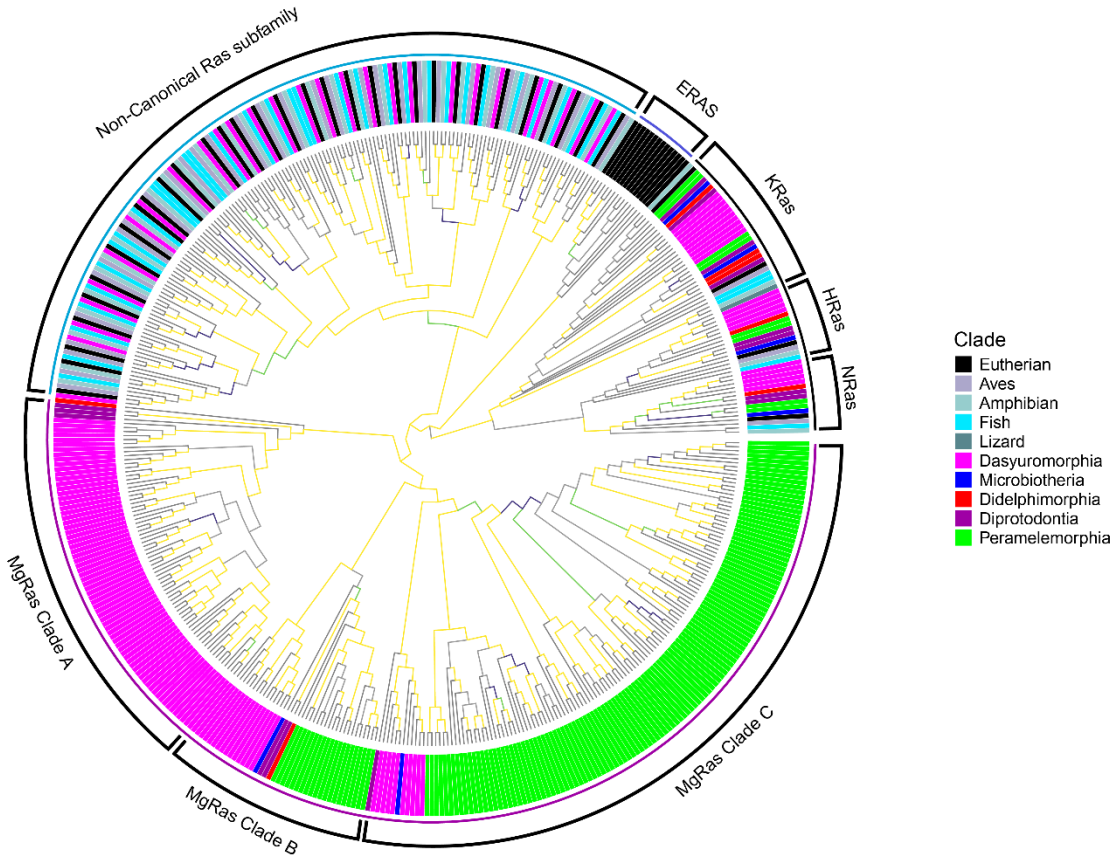


Figure 2-5. Phylogenetic relationships among Ras subfamily genes in marsupials, eutherians, birds, amphibians, and fish. ERas genes were eutherian specific and form a monophyletic clade, MgRas genes were marsupial specific and formed a monophyletic clade with three subclades (Clade A, Clade B, Clade C). The rest of the Ras subfamily were found in all vertebrate species. Yellow branches indicate bootstrap values between 95 and 100, green branches indicate bootstrap values 90-94, dark blue branches indicate bootstrap values between 80 and 89, grey branches indicate bootstrap values under 80. Terminal branches are in black.

Table 2-2. Number of MgRas genes identified amongst the 11 marsupials in this study and their tissue expression. The first number represents the total number of genes, followed by how many genes belonged to each phylogenetic clade (A, B and C). N/A indicates that gonad transcriptome was not available for that species. NCBI accession numbers for genomes and RNAseq data are included in **Table A1-2**. For the opossum, additional RNAseq data from PRJNA193216 was used (Wang et al., 2014). Note: not all genes were expressed in available transcriptomes.

Order	Species	Number of genes	Number of genes expressed in gonads	Number of genes expressed in other tissues
Diprotodontia	Koala	3 (A: 1, B: 1, C: 1)	3	1 (Clade B - pouch)
	Tammar wallaby	3 (A: 2, B: 0, C: 1)	N/A	0
Dasyuromorphia	Eastern quoll	22 (A: 16, B: 2, C: 4)	N/A	0
	Tasmanian devil	16 (A: 14, B: 1, C: 1)	15	0
	Kowari	21 (A: 15, B: 2, C: 4)	N/A	1 (Clade A - spleen)
	Yellow footed antechinus	23 (A: 17, B: 2, C: 4)	4	1 (Clade A -liver)
	Fat tailed dunnart	14 (A: 10, B: 2, C: 2)	10	1 (Clade A - proximal yolk sac, distal yolk sac)
Peramelemorphia	Greater bilby	23 (A:0, B: 17, C: 6)	14	0
	Eastern barred bandicoot	117 (A: 0, B: 103, C: 14)	77	1 (Clade B -brain)
Microbiotheria	Monito del Monte	2 (A: 0, B: 1, C: 1)	N/A	0

Order	Species	Number of genes	Number of genes expressed in gonads	Number of genes expressed in other tissues
Didelphimorphia	Grey short-tailed opossum	2 (A:1, B:0, C: 1)	N/A	1 (Clade C - extra embryo membrane, embryo brain)

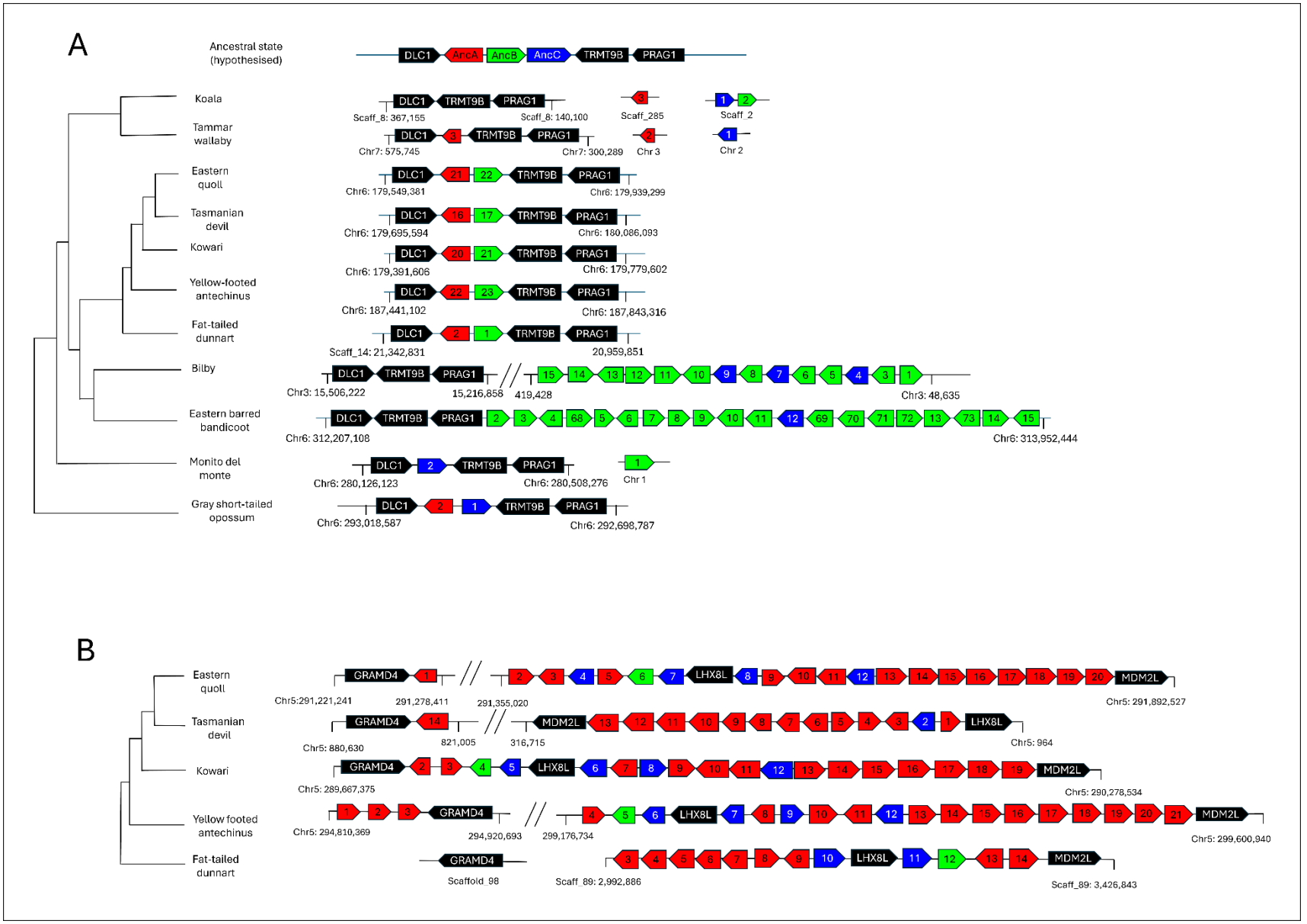

A
Ancestral state (hypothesised)
DLC1 AncA AncB AncC TRMT9B PRAG1
Koala
Tammar wallaby
Eastern quoll
Tasmanian devil
Kowari
Yellow-footed antechinus
Fat-tailed dunnart
Bilby
Eastern barred bandicoot
Monito del monte
Gray short-tailed opossum
Scaff_8: 367,155
Scaff_8: 140,100
Scaff_285
Scaff_2
Chr7: 575,745
Chr7: 300,289
Chr 3
Chr 2
Chr6: 179,549,381
Chr6: 179,939,299
Chr6: 179,695,594
Chr6: 180,086,093
Chr6: 179,391,606
Chr6: 179,779,602
Chr6: 187,441,102
Chr6: 187,843,316
Scaff_14: 21,342,831
20,959,851
Chr3: 15,506,222
15,216,858
419,428
Chr3: 48,635
Chr6: 312,207,108
Chr6: 313,952,444
Chr6: 280,126,123
Chr6: 280,508,276
Chr 1
Chr6: 293,018,587
Chr6: 292,698,787
B
Eastern quoll
Tasmanian devil
Kowari
Yellow footed antechinus
Fat-tailed dunnart
GRAMD4 LHX8L MDM2L
Chr5:291,221,241
291,278,411
291,355,020
Chr5: 291,892,527
Chr5: 880,630
821,005
316,715
Chr5: 964
Chr5: 289,667,375
Chr5: 290,278,534
Chr5: 294,810,369
294,920,693
299,176,734
Chr5: 299,600,940
Scaffold_98
Scaff_89: 2,992,886
Scaff_89: 3,426,843

Figure 2-6. MgRas genes occurred in two main clusters. (A) Cluster 1 was found in all marsupials on syntenic chromosomes. Not all bandicoot genes are depicted in figure as the cluster contained 115 genes. The hypothesised ancestral state depicted represents the most parsimonious scenario. (B) Cluster 2 was found only in dasyurids. We hypothesise that the ancestral three gene cluster was duplicated and translocated to chromosome 5 in the Dasyuridae. MgRas genes are coloured according to clade (red for A, green for B, blue for C) and flanking genes are in black. The numbers represent the gene number given to each gene.

Marsupial Ras Genes and Cancer

All genes in the Ras superfamily are molecular switches that regulate a diverse range of functions, although those in the Ras subfamily are primarily involved in cell proliferation, cell growth and the cell cycle (Goitre et al., 2014). Of the 36 mammalian Ras subfamily genes, five are listed in the COSMIC census as oncogenes (*K-Ras, H-Ras, N-Ras, Rap1B, R-Ras2*), with a number of others also implicated in cancer (Berger et al., 2014; Khalil & Nemer, 2020; Suarez-Cabrera et al., 2021; Thies et al., 2021). We hypothesise that, similar to these members of the Ras subfamily, at least some MgRas genes may also function as oncogenes. The expansion observed in dasyurids may render the species more susceptible to cancer, as a greater number of genes increases the probability of incurring a mutation in one.

Interestingly, some members of the Ras subfamily (eg. *NKIRAS1*, *DIRAS3)* appear to have the opposite role and function as tumour suppressors (Bildik et al., 2022; Postler et al., 2023). MgRas gene expansions were also observed in bandicoots and bilbies, two species that had low reported rates of cancer but noting both species have short lifespans (Lynch, 2008). As the major expansion in the Peramelemorphia was not orthologous to the major expansion in the dasyurids (**Figure 2-5**) it is possible the different groups of genes have evolved different functions. We recommend future studies use marsupial cancer biopsies to determine (i) if these genes are expressed, and (ii) if these genes (or regulatory regions) have been mutated in cancer tissue.

Oncogenes are retained as they play vital roles in organism function (Spandidos & Anderson, 1989). As the MgRas genes were almost exclusively expressed in the gonads, we hypothesise that their role may be in reproduction. The expansions observed in the Dasyuromorphia and Peramelemorphia may be due to the unique reproductive biology of these two orders. The Dasyuromorphia species in our study all give birth to supernumerary young – an excess number of offspring than the number of teats that can support them (Guiler, 1970; Morton & Fletcher, 1989; Woolley, 1971). Increased litter size has frequently been hypothesized to contribute to cancer risk in species due to antagonistic pleiotropy (Boddy et al 2020), and studies have found that cancer is more likely to be detected in species with larger litter sizes (Dujon et al., 2023)

The Peramelemorphia also have a unique reproductive biology in that they are the only marsupial species to have a chorioallantoic placenta (Renfree, 2010; Tyndale-Biscoe & Renfree, 1987). Although this is more energetically demanding pregnancy, more invasive placentation is interestingly not correlated with increased cancer risk (Dujon et al., 2023). Ultimately future studies should use transcriptomes from reproductive organs taken at different times to determine if gene expression changes throughout reproduction, noting that the species in this order are all currently listed as threatened species so sampling will be difficult.

Conclusion

Here, we aimed to characterise the evolution of cancer related genes in marsupials to determine if cancer susceptibility in certain lineages may have a genetic basis. Our results show a novel marsupial specific lineage of Ras genes that has undergone unique family specific expansions in both the Dasyuromorphia and the Peramelemorphia species (bandicoot and bilby). As these genes are expressed in the reproductive organs, we theorise that they have expanded as a result of the supernumery offspring in the Daysuromorphia, and invasive placentation in the Peramelemorphia. Genomic organisation suggests some of these may function as oncogenes with the potential to

influence cancer mortality risk in some species and recommend future studies to verify gene function.

Materials and Methods

Eastern Barred Bandicoot and Kowari Reference Genome Assemblies

Three frozen kowari tissue samples were obtained from the Australian Biological Tissue Collection (ABTC, South Australian Museum): liver from one individual (Reg No ABTC7551) and heart and spleen from a second individual (Reg No ABTC7553). Four bandicoot tissues samples were obtained from two individuals: heart and kidney from a female (University of Melbourne); and brain and gonad from a male. Both bandicoots were wild individuals from Victoria (the female was found deceased at Mount Rothwell, the male was euthanized after a vehicle strike on Philip Island). The kidney tissue was flash frozen and stored at -80° C. The brain and gonad tissue were stored at -80° C.

High molecular weight DNA was extracted from the kowari heart and bandicoot kidney tissues using the Circulomics Nanobind Tissue Big DNA Kit (Circulomics) (NB-900-701-001) and DNA concentration was measured using Qubit fluorometer (ThermoFisher Scientific). DNA was then pooled for each species and sent to Australian Genome Research Facility (Brisbane, Australia) for PacBio HiFi library prep and Revio sequencing on a 25M SMRT cell. Heart tissue for the kowari and bandicoot was sent for HiC Armina 2.0 library preparation and sequenced on NovaseqX at Biomolecular Resource Facility (Australian National University).

Total RNA was extracted from all three kowari tissues (liver, heart and spleen) and two bandicoot tissues (brain and gonad) using a Qiagen RNeasy mini kit (Qiagen, Cat. No. 74104) with on-column DNAse digestion using the DNase I set (Qiagen). RNA quality was assessed using the RNA nano 6000 kit on the Bioanalyzer (Agilent) then submitted to Ramaciotti Centre for Genomics (The University of New South Wales) for Illumina

stranded mRNA prep. The RNA was sequenced as 150 bp paired end reads on a NovaSeq X Plus 10B flowcell, resulting in 76 to 120 million reads per sample.

Both reference genomes were assembled on the Galaxy Australia webserver (https://usegalaxy.org.au/) using the Vertebrate Genome Project genome assembly pipeline (Lariviere et al., 2024). Briefly, HiFi reads were quality trimmed and reads containing adapters removed using Cutadapt v4.9 (Martin, 2011), genome size and k-mers were estimated using Meryl v1.3 (Rhie et al., 2020) and GenomeScope v2.0 (Ranallo-Benavidez et al., 2020). Genome assembly was performed using the hifiasm v2.1 in HiC mode (Cheng et al., 2021; Price, 2022) using HiFi reads to assemble contigs and Hi-C reads to identify haplotypes, with the assembly quality assessed using gfastats v1.3.9 (Formenti et al., 2022) and BUSCO v5.8.0 with mammlia_odb10 lineages (Simão et al., 2015). Hi-C reads were then used to scaffold contigs of the primary assembly, scaffolding was performed using YaHS v1.2a.2 (Zhou et al., 2023) with contact maps generated with Pretext v0.1.9 (Harry).

To identify sex chromosomes in the kowari, HiFi reads were mapped to the assembly using minimap2 v2.22-r1101 (Li, 2018) and secondary alignments excluded using Samtools v1.9 (Danecek et al., 2021). The genomes were divided into 1 Mbp windows and the mean coverage for each window was calculated using Bedtools v2.31.0 (Quinlan & Hall, 2010). We also performed NBLASTN and TBLASTN using Blast v2.2.30 (Altschul et al., 1990) search for the sex-linked SRY gene on the Y chromosome, using the human (NP_003131.1 and NM_003140.3) and devil (XP_031801149.1 and XM_031945289.1) SRY genes as query sequences.

Genome Annotations

Nine additional marsupial genomes were used in this study: Monito del monte (*Dromiciops gliroides*) (Rhie et al., 2021), gray short-tailed opossum (*Monodelphis domestica*) (Rhie et al., 2021), greater bilby (*Macrotis lagotis*) (Hogg et al., 2024), koala

(*Phascolarctos cinereus*) (Damas et al., 2022), tammar wallaby (*Notamacropus eugenii*) (O'Neill, R., *pers. comm.*), fat-tailed dunnart (*Sminthopsis crassicaudata*) (Ibeh et al., 2024), yellow-footed antechinus (*Antechinus flavipes*) (Tian et al., 2022), Tasmanian devil (*Sarcophilus harrisii*) (Stammnitz et al., 2023) and Eastern quoll (*Dasyurus viverrinus*) (Hartley et al., 2024). See **Table A1-2** for further information about these genomes. All eleven genomes were annotated using the same method to ensure consistency and reliability of results. For more details about the methods used for genome annotation, see **Supplementary Material to Chapter 2**. A synteny map was created using GENESPACE v1.3.1(Lovell et al., 2022)

Cancer Gene Prevalence and Orthologs

To estimate cancer prevalence in different marsupial taxa, we collated data from previous published summaries that contained comparisons of different marsupial species. We also obtained data from the Wildlife Health Information System (eWHIS).

The Cancer Gene Census (v100_GRCh38) was downloaded from the COSMIC (Catalogue of Somatic Mutations in Cancer) website https://cancer.sanger.ac.uk/cosmic (Sondka et al., 2024). We extracted coding sequences for these genes from the human reference genome (GCF_000001405.40_GRCh38.p14_cds_from_genomic.fna).

To identify orthologs of these genes in the marsupial species, we performed a reciprocal blast best hit analysis. We used BLASTP v2.2.30 (Altschul et al., 1990) using the marsupial annotation files as queries against the human annotation file GCF_000001405.40_GRCh38.p14 (Schneider et al., 2017) and an e-value cutoff of 0.003. We then performed the reverse i.e. the human annotation file was used as a query against each marsupial annotation file. Two genes (e.g. *A* and *B*) were then deemed to be orthologs if gene *A* was the best hit for gene *B* in the first blast search, and gene *B* was also the best hit for gene *A* in the second blast search. Orthologs of the genes from the Cancer Gene Census were then flagged for easy identification in downstream analysis.

Gene Family Analysis

To remove potential pseudogenes, we excluded any predicted genes from the FGENESH++ genome annotations that had no mRNA evidence and did not have any hits with any known proteins. We also excluded any genes annotated as repetitive elements (SINE, LINE, L1TD1, transposase, retroposon, retrotransposon, reverse transcriptase). All eleven filtered genome annotations were used as input for Orthofinder v2.4.0 (Emms & Kelly, 2019) to infer orthogroups (sets of genes descended from a single ancestral gene).

We then created a rooted, ultrametric phylogenetic tree. We used a species tree inferred by Orthofinder (SpeciesTree_rooted_at_outgroup_6) as the topology of the tree matched the current marsupial phylogeny consensus (Duchêne et al., 2018; Westerman et al., 2016). We calibrated this tree using r8s v1.81 (Sanderson, 2003) with divergence dates obtained from timetree.org (Kumar et al., 2022) (**Table A1-5**).

Highly variable orthogroups with more than 100 genes in a single lineage were analysed separately, as recommended by the CAFE manual (Mendes et al., 2020). We then ran CAFE5 v5.1.0 (Mendes et al., 2020) to estimate an error model for all future analyses (-e flag). We choose a gamma model with k=2 (meaning gene families can belong to one of two different evolutionary rate categories) with two λ values (λ_1 for Australian marsupials and λ_2 for American marsupials). The λ parameter refers to the gene family evolution rate, or the probability that a gene will be gained or lost in a particular lineage. For more information about why this model was chosen, see **Supplementary Material to Chapter 2**. Gene families that had undergone significant expansions or contractions in either the koala or dasyurid ancestral lineages (the two lineages with high cancer prevalence) were then manually examined for interesting patterns.

Manual Gene Annotation

As the CAFE analysis indicated potential expansion of Ras genes in some lineages, we verified the automated gene annotation through manual annotation. Human sequences for the four canonical RAS genes (*H-Ras, K-Ras4A, Kras4B* and *N-Ras*) were downloaded from Uniprot and used as query sequences for TBLASTN v2.2.30 against all the marsupial genomes and transcriptomes. Putative genes were extracted using Bedtools v2.29.2 (Quinlan & Hall, 2010) then inspected using Integrative Genomics Viewer v2.16.0 (Robinson et al., 2011) and exon boundaries identified using either the GT/AG or GC/AG convention. Genes were confirmed to be orthologs if they exhibited the extremely conserved sequences of canonical Ras genes in other vertebrate (García-España & Philips, 2023).

Once orthologs of the canonical genes had been identified in each marsupial species, a hidden Markov model (HMM) was constructed using the nucleotide sequences for each species. The HMM was then used to search all genomes using HMMR v3.3.2 (Eddy, 2009). For each marsupial species, this process (TBLASTN and HMMR) was repeated iteratively, incorporating new sequences into the queries until no new hits were found.

Putative hits were first checked for an open reading frame (including start codon and stop codon) and aligned using ClustalW in BioEdit (Hall, 1999). They were initially retained if they contained partial sequences for at least four out of the five conserved G box motifs of Ras genes (GXXXXGKS/T, T, DXXG, T/NKXD, C/SAK/L/T) (Colicelli, 2004). For each sequence, BLASTP was then performed against non-redundant sequences in *Homo sapiens* using the webserver https://blast.ncbi.nlm.nih.gov/ (*Homo sapiens* was used as it is well annotated). The SMART webserver (http://smart.embl-heidelberg.de/) (Letunic et al., 2021) was also used to identify domains in the sequences. Putative genes were then considered to be Ras genes if the best BLAST hit was a canonical Ras gene (*H-Ras, K-Ras, N-Ras*) and the SMART domain was also Ras. We used featureCounts in the subread package v1.5.1 (Liao et al., 2014) to determine if these genes were expressed in any of the

tissues in the available transcriptomes. Genes were considered to be expressed if they had at least three reads present in a transcriptome.

To determine the position of the newly annotated genes in RAS phylogeny, we also annotated 29 out of the 33 Ras subfamily genes in the devil genome using the same method (see **Table A1-6** for details of the query sequences). We then aligned nucleotide sequences using MAFFT 7.522 (Katoh & Standley, 2013) and generated a phylogenetic tree with IQTree v2.2.2 (Minh et al., 2020) using ModelFinder (Kalyaanamoorthy et al., 2017) to determine the best model. For branch support, we used 1000 replicates of ultrafast bootstrap (Hoang et al., 2018) and an SH-aLRT test (Guindon et al., 2010) with 1000 replicates. The tree was rooted using the three canonical Ras genes (*K-Ras*, *N-Ras* and *H-Ras*) as these are the founding members of the gene family. Accession numbers for genes used to create the tree are included in **Table A1-7**.

In order to identify patterns of synteny, we annotated flanking genes around Ras expansions by manually inspecting the devil and antechinus genomes in NCBI's Genome Data Viewer (Rangwala et al., 2021). We then used these as query sequences in TBLASN against the other genomes.

ChatGPT 3.5 (https://chatgpt.com/) was used for assistance in some coding and debugging, and all scripts were carefully evaluated and remain the responsibility of the authors.

Acknowledgements

We would like to thank Wildlife Health Australia for facilitating provision of data from the national Wildlife Health Information System (eWHIS). Wildlife health data were provided by: Adelaide Koala and Wildlife Centre, Agriculture Victoria, Australia Zoo Wildlife Hospital, Bonorong Wildlife Sanctuary, Currumbin Wildlife Sanctuary, Department of Natural Resources and Environment Tasmania, Department of Primary Industries and

Regions SA, Healesville Sanctuary, James Cook University, Kingston Animal Hospital, Melbourne Zoo, NSW Department of Primary Industries, Perth Zoo, Queensland Department of Agriculture and Fisheries, RSPCA Qld, Taronga Western Plains Zoo, Taronga Zoo Wildlife Hospital, University of Queensland, WA Department of Primary Industries and Regional Development, WA Wildlife, and Zoos SA. We would like to thank the South Australian Museum for the kowari samples, A. Weeks and D. Sutherland for the eastern barred bandicoot samples. We thank the Vertebrate Genome Project for the generation of the Monito del Monte genome, with sequencing and genome assembly conducted by the Max Planck Institute in Dresden, led by Gene Meyers, and coordinated with Olivier Fedrigo and Erich D. Jarvis; and the Gray short-tailed opossum genome, with sequencing and genome assembly conducted at the Vertebrate Genomes Lab (VGL) at the Rockefeller University, led by Olivier Fedrigo and Erich D. Jarvis, with support from Ed Lien and Trygve Bakken at the Allen Institute for Brain Science.

Funding

This work was supported by funding from the ARC Centre of Excellence for Innovations in Peptide and Protein Science (CE200100012) and the NCRIS-funded Bioplatforms Australia Threatened Species Initiative.

Data Availability

The genome assemblies and raw transcriptome files are available on NCBI and the SRA, under BioProject PRJNA1301011 (kowari) and PRJNA1301027 (Eastern barred bandicoot).

Author Contributions

CP designed the study and performed comparative genomics analyses with guidance from EP and LWS. CP and LWS generated genome assemblies for the kowari and Eastern

barred bandicoot and genome annotations for all species. RJO and PGSG generated the tammar wallaby genome assembly. CJH and KB sourced funding and undertook project management and supervision. CP wrote the main manuscript text, with feedback and revisions on the manuscript were provided by all authors.

CHAPTER THREE:

AMPED UP IMMUNITY: 418 WHOLE GENOMES REVEAL INTRASPECIFIC DIVERSITY OF KOALA ANTIMICROBIAL PEPTIDES

CHAPTER 3. AMPED UP IMMUNITY: 418 WHOLE GENOMES REVEAL INTRASPECIFIC DIVERSITY OF KOALA ANTIMICROBIAL PEPTIDES

3.1 BACKGROUND

Chapter 3 comprises the following published manuscript:

Petrohilos, C., Peel, E., Silver, L. W., Belov, K., & Hogg, C. J. (2025). AMPed up immunity: 418 whole genomes reveal intraspecific diversity of koala antimicrobial peptides. *Immunogenetics*, 77(1), 11.

This chapter investigates intraspecific diversity in koala antimicrobial peptides across the entire east coast of Australia. The results reveal diversity at both the nucleotide and copy number level, including non-synonymous SNPs that are predicted to alter peptide function.

I designed the study with guidance from Emma Peel and Luke W. Silver. Luke W. Silver and I performed all data analysis with assistance from Emma Peel. Carolyn J. Hogg and Katherine Belov sourced funding and undertook project management and supervision. I wrote the main manuscript text with feedback and revisions provided by all authors.

3.2 MANUSCRIPT

AMPed Up Immunity: 418 Whole Genomes Reveal Intraspecific Diversity of Koala Antimicrobial Peptides

Cleopatra Petrohilos[1,2], Emma Peel[1,2], Luke W. Silver[1,2], Katherine Belov[1,2] and Carolyn J. Hogg[1,2]

[1]School of Life and Environmental Sciences, The University of Sydney, Sydney, New South Wales, Australia

[2]Australian Research Council Centre of Excellence for Innovations in Peptide & Protein Science, The University of Sydney, Sydney, NSW, Australia

Corresponding Author: Prof Carolyn Hogg, The University of Sydney, Sydney, NSW, 2006, Australia

Email: carolyn.hogg@sydney.edu.au

Keywords: cathelicidins, defensins, immune genes, conservation genomics

Abstract

Characterising functional diversity is a vital element to understanding a species' immune function yet many immunogenetic studies in non-model organisms tend to focus on only one or two gene families such as the major histocompatibility complex (MHC) or toll-like receptors (TLR). Another interesting component of the eukaryotic innate immune system are the antimicrobial peptides (AMPs). The two major groups of mammalian AMPs are cathelicidins and defensins, with the former having undergone species-specific expansions in marsupials. Here, we utilised data from 418 koala whole genomes to undertake the first comprehensive analysis of AMP diversity across a mammalian wildlife species' range. Overall allelic diversity was lower than other immune gene families such as MHC, suggesting that AMPs are more conserved, although balancing selection was observed in PhciDEFB12. Some non-synonymous SNPs in the active peptide are predicted to change AMP function through stop gains, change in structure, and increase in peptide charge.

Copy number variants (CNVs) were observed in two defensins and one cathelicidin. Interestingly, the most common CNV was the duplication of PhciCATH5, a cathelicidin with activity against chlamydia, which was more common in the southern part of the species range than the north. AMP copy number is correlated with expression levels, so we hypothesise that there is a selective pressure from chlamydia for duplications in PhciCATH5. Future studies should use phenotypic metadata to assess the functional impacts of gene duplication.

Introduction

Immune genes are some of the fastest evolving in the genome due to the co-evolutionary arms race between hosts and pathogens (Barreiro & Quintana-Murci, 2010; Trowsdale & Parham, 2004). Although balancing selection frequently results in many immune genes being highly polymorphic, this is not uniform across all immune gene families (Mukherjee et al., 2009; Vinkler et al., 2023). Characterising functional diversity is an important component of understanding species' biology and is particularly important in threatened species management where immunogenetic variation has been shown to influence disease susceptibility in many species (Elbers et al., 2018; Morrison et al., 2020). However, much of the research in non-model organisms to date has only focused on a small number of immune gene families such as the major histocompatibility complex (MHC) and toll-like receptors (TLR) (Bagheri & Zahmatkesh, 2018; Brouwer et al., 2010; Grueber et al., 2014; Minias et al., 2019; Minias & Vinkler, 2022; Vinkler et al., 2023).

More recently, other immune genes such as antimicrobial peptides (AMPs) have been investigated (Chapman et al., 2016; Hellgren, 2015; Schmitt et al., 2017). AMPs are small, cationic molecules with broad spectrum antimicrobial activity. They are an ancient component of the eukaryotic immune system and are expressed by both plants and animals (Zasloff, 2002). The two major families of mammalian AMPs are cathelicidins and defensins (Gallo et al., 2002). Both are encoded as precursor molecules composed of three domains: a signal sequence, a propeptide sequence and an antimicrobial domain. Upon

activation, the antimicrobial domain is cleaved off to create the active mature peptide (Selsted & Ouellette, 2005; Zanetti, 2005).

AMPs generally exhibit lower overall diversity than other immune genes, with only one or two haplotypes being common across a population (Chapman et al., 2016; Gilroy et al., 2016; Hellgren, 2015; Schmitt et al., 2017). Immune genes such as MHC benefit from high levels of polymorphism as this enables them to bind to a wide variety of antigenic peptides (Spurgin & Richardson, 2010). In contrast, AMPs are characterised by their broad-spectrum activity against a diverse range of microbes. This is largely mediated by their physicochemical properties such as net cationic charge and the ability to adopt an amphipathic structure (Zasloff, 2002). These properties enable the peptides to electrostatically bind to negatively charged bacterial membranes and disrupt their integrity, leading to cell death (Zasloff, 2002). However, they also likely constrain nucleotide polymorphisms given physiochemical properties are tightly linked to function.

Characterising diversity in AMPs is important as even minimal amino acid changes can drastically alter the physicochemical and antimicrobial properties of peptides (Hellgren et al., 2010; Higgs et al., 2007; Meade et al., 2008). To date, intraspecific AMP diversity has only been investigated in birds and agricultural animals (Brahma et al., 2015; Chapman et al., 2016; Gillenwaters et al., 2009; Gilroy et al., 2016; Hellgren, 2015; Ishige et al., 2021; Monteleone et al., 2011; Schmitt et al., 2017). Cathelicidin intraspecific diversity has not yet been investigated in wildlife, with most studies instead exploring defensin diversity in one or two populations (with a total sample size ranging from 5 – 160) (Gilroy et al., 2016; Hellgren, 2015; Schmitt et al., 2017). The most comprehensive study has involved mallards (*Anas platyrhynchos*) from global populations (n=274 from a single location in Sweden and n=190 from 16 other geographic locations), although only five genes were included in this analysis (Chapman et al., 2016). Collectively, these studies suggest that most defensins are either monomorphic or primarily consist of one major allele that is predominant across a population.

However, some diversity has been observed within AMPs, at both nucleotide and copy number level (**Table A2-1**). For example, the mallard defensin AvBD10 consists of 41 alleles. Only two alleles are present at high frequencies in most populations but their frequencies can vary between geographic regions (Chapman et al., 2016). The domestic water buffalo (*Bubalus bubalis*) expresses seven cathelicidins with one (CATHL4) having high polymorphism at both the nucleotide and copy number level (Brahma et al., 2015). As this study only involved 25 individual animals from a single slaughterhouse, it is unknown whether this differs between populations. Defensins also exhibit high variation in copy number between individuals in humans (Hollox et al., 2003; Linzmeier & Ganz, 2005), with variation shown to influence disease susceptibility (Hollox, 2009; Hollox et al., 2008).

All these studies have relied on species-specific primers to amplify genes of interest. This means they have been restricted to studying a small number of genes (typically four to six) that represent only a fraction of a species' AMP repertoire. As whole genome sequencing (WGS) costs have decreased our ability to investigate functional diversity in non-model organisms across multiple gene families has increased (Fuentes-Pardo & Ruzzante, 2017). WGS has enabled exploration of the entire range of AMP genes in a species' genome rather than being limited to a small number of genes using more conventional analytical methods.

Marsupials are a particularly interesting model species as they encode a diverse repertoire of AMPs. The Tasmanian devil (*Sarcophilus harrisii*) has seven cathelicidins (Peel et al., 2016), gray short-tailed opossum (*Monodelphis domestica*) has 19 (Cho et al., 2020) and the koala (*Phascolarctos cinereus*) has ten (Peel et al., 2021), whilst humans and mice only have a single cathelicidin gene (Ramanathan et al., 2002). This high diversity in marsupials is often attributed to their unique reproductive biology (Peel et al., 2017), characterised by a short gestation period and birthing of highly altricial young that are immunologically naive (Old & Deane, 2000). Marsupial neonates are often exposed to a variety of pathogens as they complete their development in the non-sterile pouch environment (Maidment et

al., 2023; Weiss et al., 2021). During this period they depend on passive immunity conferred by the mother, including AMPs delivered through pouch secretions, maternal licking, and milk (Edwards et al., 2012). The selective pressure of this environment may have encouraged the species-specific expansion observed in the cathelicidins of marsupials.

Many marsupial AMPs exhibit antimicrobial activity *in vitro* against a range of bacteria, fungi, and viruses, including multidrug resistant strains (Cho et al., 2020; Peel et al., 2016; Peel et al., 2017; Wang et al., 2011). Interestingly, marsupial cathelicidins are also active against pathogens of conservation concern. For example, the koala cathelicidin PhciCATH5 is active against the bacterium *Chlamydia pecorum* (Peel et al., 2021), the main causative agent of chlamydiosis and a major threatening process for the species (Polkinghorne et al., 2013). Chlamydiosis is an endemic disease with devastating effects on koala populations that can lead to blindness and infertility (Polkinghorne et al., 2013). Similarly, Tasmanian devil cathelicidins exhibit anticancer activity against Devil Facial Tumour Disease (DFTD), a contagious cancer that has decimated populations throughout the species' range (Petrohilos et al., 2023). These findings make marsupial AMPs promising candidates for drug discovery and development against a range of pathogens of concern, for both humans and wildlife.

In 2021, the Koala Genome Survey sequenced 430 koala genomes across the eastern coast of Australia to improve understanding of the species' genomic diversity (Hogg et al., 2023). Here, we used 418 individuals from this unique dataset to conduct the first comprehensive analysis of AMP diversity across a wildlife mammal species' range, ~700 000 km^2 in the instance of koalas. Our aims were to: (i) characterise baseline level of nucleotide and copy number diversity amongst koala AMPs across their geographic range, and (ii) use *in silico* methods to predict functional effects of SNPs and identify targets for future research.

Methods

We downloaded 418 aligned koala BAM files from the Amazon Web Services (AWS) Open Data Program (https://awgg-lab.github.io/australasiangenomes/species/Phascolarctos_cinereus.html). Captive koalas (n = 12) were excluded from our dataset as they were deemed not to represent natural genetic diversity. A multi sample VCF file was created using the Dragen gVCF genotyper in the non-iterative mode on a Dragen V4 server. The Dragen Joint Genotyping Pipeline (v3.9.5) was used to run joint genotyping on Illumina's Basespace portal.

Published koala defensin (Jones et al., 2017) and cathelicidin (Peel et al., 2021) sequences were annotated in the current genome assembly (GCA_002099425.1_phaCin_unsw_v4.1) (Johnson et al., 2018) using BLAST v2.2.30 (Altschul et al., 1990). Exons were extracted using bedtools (version 2.29.2) and manually checked to ensure consensus with published sequences (Jones et al., 2017; Peel et al., 2021). In total, 38 AMPs were used in this study: seven cathelicidins, two alpha defensins (including one partial sequence) and 29 beta defensins (including 21 partial sequences). Although it is possible these partial sequences represent pseudogenes, we included them as very short sequences are difficult to identify using homology-base algorithms. The first exon of marsupial defensins has an average length of 69 bp and so many marsupial defensins have only been annotated as partial sequences (Peel et al., 2024). As the second exon contains the active peptide sequence that is of most interest, many studies have only focused on this region when characterising diversity (Chapman et al., 2016). The PhciDEFB10 sequence annotated in this study differs to the published sequence (Jones et al., 2017), which we attribute to the different versions of the genome assembly. The coordinates and sequences of the AMPs used in this study are listed in **Table A2-2 –Table A2-3**. The predicted active peptide sequences are listed in **Table A2-4**.

The whole genome joint genotyped multi sample VCF file was filtered to only include biallelic variants within the AMP genes using gatk (version 4.2.0.0) SelectVariants. Gatk

(version 4.2.0.0) VariantFiltration was then used to exclude variants with QUAL < 40, MQ < 40.00, MQRankSum > |12.5|, FS > 60.00, QD < 1.5, ReadPosRankSum > |8.00|. AMPs that did not contain any SNPs were excluded from further analysis. Variants were annotated using Annovar (version 20180416) and classified as intronic or exonic, with exonic variants categorized as synonymous or non-synonymous.

Alleles were determined by first using gatk (version 4.2.1.0) FastaAlternateReferenceMaker to obtain the nucleotide sequence for all AMPs in each individual. Seqphase (Flot, 2010) was used to convert sequences into PHASE format, PHASE (version 2.1.1) was then used to determine alleles and estimate allele frequencies for each gene. Seqphase was then used to convert the output into fasta format. Alleles that occurred in less than two individuals were excluded as potential errors. Alleles for each gene are saved in Supplementary File 3 available online here. Sequences with unresolved alleles were removed and Tajima's D was calculated using DnaSP 6 (Rozas et al., 2017).

Alleles that contained non-synonymous SNPs in the active peptide region (exon four of cathelicidins and exon two of defensins) were tested *in silico* to predict the functional effect. The webserver https://protcalc.sourceforge.net/ was used to calculate change in peptide charge, the CSM-peptides webserver (Rodrigues et al., 2022) was used to predict antibacterial, antiviral, anticancer and anti-inflammatory activity and the AntiFungal webserver (J. Zhang et al., 2021) was used to predict antifungal activity. Three dimensional structures of the active peptides with non-synonymous SNPs were predicted using RoseTTAFold (Baek et al., 2021) and visualized using Chimera 1.18 (Pettersen et al., 2004) (**Figure A2-1 – Figure A2-15**).

Datamonkey webserver (Delport et al., 2010; Pond & Frost, 2005; Weaver et al., 2018) was used to test for residues under positive or negative selection using three tests: fixed effects likelihood (FEL) (Kosakovsky Pond & Frost, 2005), mixed effects model evolution (MEME) (Murrell et al., 2012), and A Fast, Unconstrained Bayesian AppRoximation for

Inferring Selection (FUBAR) (Murrell et al., 2013). As Datamonkey requires a minimum of three sequences, only AMPs with three or more alleles were tested for selection. Only the coding sequence was used for testing and sites were considered under selection if they were supported by two or more of the tests.

CNVnator (version 0.4.1) was used to call copy number variants (CNVs) on the BAM files using a bin size of 1000bp. Low quality calls were removed, including those with q0 > 0.5 (indicating that more than 50% of the reads supporting the CNV had zero mapping quality), e-value > 0.05 and size < 1 kb or > 5 Mb. Any CNVs that overlapped with regions containing AMPs were collated into a single file (**Table A2-5**). Samples were then divided into six geographic regions to determine the distribution of CNVs: North Queensland (QLD) (samples collected north of Brisbane River) (n=47); South QLD (samples collected south of Brisbane River) (n=54); North New South Wales (NSW) (samples collected north of Clarence River) (n=76); Mid NSW (samples collected south of Clarence River and north of Hunter Valley (n=57); South NSW (samples collected south of Hunter Valley) (n=113); and Victoria (n=72) (**Figure 3-1**). These geographic subregions were selected as the Hunter Valley, Brisbane and Clarence Rivers have been identified as biogeographic barriers to gene flow in koalas (Johnson et al., 2018). Figures were created using ggplot2 (v3.4.2) (Wickham, 2016) in R 4.1.3 (Team, 2022).

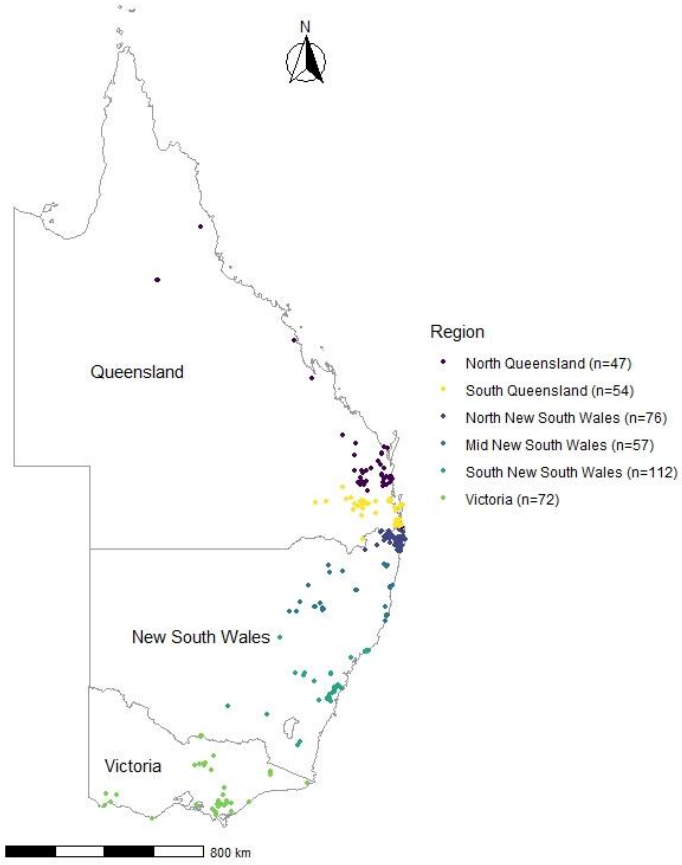


Figure 3-1. Sampling locations coloured by geographic subregions.

Results

Nucleotide Diversity and Selection Analysis

We investigated diversity in 38 AMP genes (7 cathelicidins, 31 defensins) across 418 individuals. 609 SNPs were retained following filtering. From these 59 SNPs were exonic, out of which 39 (66%) were non-synonymous (**Table 3-1**). For cathelicidins, all seven genes contained exonic SNPs of which 8 were synonymous and 13 were non-synonymous. Most of the non-synonymous SNPs (12 out of 13) were in the prepro region rather than the active mature peptide.

For defensins, 8 synonymous and 26 non-synonymous exonic SNPs were identified amongst the 31 genes studied. Non-synonymous SNPs in the active peptide were more common in defensins compared to cathelicidins, with 13 out 31 genes containing at least one non-synonymous SNP (**Table 3-1**). This included deleterious mutations such as stop gains in alpha defensin PhciDEFA1, and beta defensins PhciDEFB12 and PhciDEFB4. In particular, the stop gain in PhciDEFA1 removes some of the conserved cysteines involved in the disulfide bonds responsible for peptide structure. Similarly, PhciDEFB16 has a SNP that affected one of the conserved cysteine residues. As these cysteine residues are responsible for forming the three disulfide bonds that characterise defensins, this would likely lead changes in structure and function.

Two out of seven cathelicidins had between two and three sites and one defensin had a single site under negative selection (**Table 3-1**). Tajima's D was only significant for one AMP (PhciDEFB12), suggesting that this defensin may be under balancing selection (**Table 3-1**).

Allelic Diversity and Spatial Structuring

The AMPs investigated here in this study contained between 1 to 9 alleles, with a mean of 2.6 alleles per AMP (**Table 3-1**). Eleven defensins, over a quarter of all peptides investigated, were invariant in all populations.

Allelic diversity decreased along a geographic gradient. Of the 27 peptides that did consist of multiple alleles, two were monomorphic in North QLD; four were monomorphic in South QLD; six were monomorphic in North NSW; six were monomorphic in Mid NSW; eight were monomorphic in South NSW and fourteen in Victoria (**Table A2-6**). Most individuals were homozygous for most of the genes (**Table A2-7**).

Overall, allelic diversity was low: eleven peptides consisted of only two alleles, with a single predominant allele having a frequency > 0.70 in all regions (**Table A2-6**).

However, there were some exceptions to this. For example, although PhciDEFB10_Hap1 was the major allele in all geographic regions, it had a much lower frequency in Mid NSW (0.526) than other populations (**Figure 3-2E**).

For PhciCATH6, PhciDEFA1, PhciDEFB9 and PhciDEFB26, the major allele differed between geographic regions. PhciCATH6_Hap1 was predominant in all regions except for North NSW (**Fig.3-2A**). PhciDEFA1_Hap4 was predominant in QLD and South NSW, PhciDEFA1_Hap3 was predominant in Victoria, while North and Mid NSW contained a balance of two or three alleles (**Figure 3-2B**). PhciDEFB9_Hap1 was the most common in northern regions (QLD, North and Mid NSW), and PhciDEFB9_Hap4 was the most common allele in South NSW and Victoria (**Figure 3-2D**). PhciDEFB26_Hap1 was most common in northern regions (QLD, North NSW), PhciDEFB26_Hap2 was most common in southern regions (South NSW and Victoria) while Mid NSW contained similar frequencies of both (0.491 of PhciDEFB26_Hap1 and 0.447 of PhciDEFB26_Hap2) (**Figure 3-2H**).

In Silico *Predictions*

14 AMPs contained non-synonymous SNPs in the active mature peptide region that were predicted to influence function (**Table 3-2**). This included eight peptides that had an increase in charge of one or greater and six peptides with a change in predicted antimicrobial or immunomodulatory activity.

Table 3-1. Summary of nucleotide and allelic diversity. For Tajima's D values, an asterisk indicates a p-value < 0.05. Phci is the precursor indicating the species, in this case koala. CATH is cathelicidin, DEF are the defensins with DEFA indicating the alpha-defensins and DEFB indicating the beta-defensins. Splice refers to a SNP that is within 2bp from an exon/intron boundary.

Name of gene	Total SNPs	Intronic SNPs	Exonic SNPs	Non-Synonymous Exonic SNPs	Non-Synonymous SNPs in active peptide	Stop gain	Splice	Number of alleles	Sites under negative selection	CNV?	Tajima's D
PhciCATH1	24	23	1	1	0	0	0	2	0	N/A	-0.50013
PhciCATH2	22	18	4	3	0	0	0	5	0	N/A	1.39557
PhciCATH3	24	21	3	2	1	0	0	3	0	N/A	1.02439
PhciCATH5	29	27	2	0	0	0	0	4	2	duplication	1.93808
PhciCATH6	26	23	3	2	0	0	0	4	0	N/A	0.03613
PhciCATH7	27	25	2	2	0	0	0	3	0	N/A	-0.81107
PhciCATH8	68	62	6	3	0	0	1	9	3	N/A	0.17117
PhciDEFA1	3	0	3	2	2	1	0	4	0	N/A	1.92027
PhciDEFA2	20	19	1	0	0	0	0	1	0	N/A	-0.84986
PhciDEFB1	69	69	0	0	0	0	0	1	0	N/A	N/A
PhciDEFB10	124	123	1	1	1	0	0	2	0	N/A	1.39771
PhciDEFB11	2	0	2	2	2	0	0	3	0	N/A	0.67435
PhciDEFB12	5	0	5	4	4	1	0	5	0	N/A	2.66593*
PhciDEFB13	0	0	0	0	0	0	0	1	0	N/A	N/A
PhciDEFB14	40	39	1	0	0	0	0	2	0	N/A	1.12422
PhciDEFB15	0	0	0	0	0	0	0	1	0	N/A	N/A
PhciDEFB16	21	20	1	1	1	0	0	2	0	N/A	-0.65634
PhciDEFB17	0	0	0	0	0	0	0	1	0	N/A	N/A
PhciDEFB18	1	0	1	1	0	0	0	2	0	N/A	-0.79334

Name of gene	Total SNPs	Intronic SNPs	Exonic SNPs	Non-Synonymous Exonic SNPs	Non-Synonymous SNPs in active peptide	Stop gain	Splice	Number of alleles	Sites under negative selection	CNV?	Tajima's D
PhciDEFB19	2	0	2	2	2	0	0	4	0	N/A	0.49335
PhciDEFB20	3	0	3	2	2	0	0	3	1	N/A	0.325
PhciDEFB21	0	0	0	0	0	0	0	1	0	N/A	N/A
PhciDEFB22	0	0	0	0	0	0	0	1	0	N/A	N/A
PhciDEFB23	0	0	0	0	0	0	0	1	0	deletion	N/A
PhciDEFB24	1	0	1	1	1	0	0	2	0	duplication	-0.39911
PhciDEFB25	0	0	0	0	0	0	0	1	0	N/A	N/A
PhciDEFB26	3	0	3	0	0	0	0	4	0	N/A	0.59922
PhciDEFB27	1	0	1	1	1	0	0	2	0	N/A	-0.232
PhciDEFB28	1	0	1	1	1	0	0	2	0	N/A	2.06845
PhciDEFB29	0	0	0	0	0	0	0	1	0	N/A	N/A
PhciDEFB3	26	24	2	1	1	0	0	3	0	N/A	0.39386
PhciDEFB30	0	0	0	0	0	0	0	1	0	N/A	N/A
PhciDEFB4	3	0	3	2	0	1	0	4	0	N/A	0.75082
PhciDEFB5	58	57	1	1	0	0	0	2	0	N/A	-0.63003
PhciDEFB6	0	0	0	0	0	0	0	1	0	N/A	N/A
PhciDEFB7	2	0	2	1	1	0	0	3	0	N/A	-0.81084
PhciDEFB8	1	0	1	0	0	0	0	2	0	N/A	-0.76548
PhciDEFB9	3	0	3	3	3	0	0	4	0	N/A	1.8104

Only one cathelicidin PhciCATH3 contained non-synonymous SNPs in the active mature peptide (**Table 3-2**). PhciCATH3_Hap2 and PhciCATH3_Hap3 both had a G > R substitution which increased the charge at pH7 from +8.3 to +9.3. While PhciCATH3_Hap1 was predicted to have antiviral activity (**Table 3-2**), the other two alleles are not. PhciCATH3_Hap1 is the predominant allele present across the entire geographic range and the only allele present in Victoria. PhciCATH3_Hap3 showed the highest frequency in QLD, where it had a frequency of 0.234 in North QLD and 0.157 in South QLD.

Although koala defensins have not yet been tested *in vitro*, many are predicted to have antifungal activity (PhciDEFA1, PhciDEFB9, PhciDEFB10, PhciDEFB11, PhciDEFB12, PhciDEFB16, PhciDEFB19, PhciDEFB24, PhciDEFB28) (**Table 3-2**). Some defensins contained variants that altered this predicted activity. For example, PhciDEFB10_Hap2 had a K > E substitution that resulted in a lower charge at pH 7 (+ 5.7) than PhciDEFB10_Hap1 (+ 7.7). While PhciDEFB10_Hap1 was predicted to have antifungal activity, PhciDEFB10_Hap2 was not (**Table 3-2**). PhciDEFB10_Hap1 is also the predominant allele across the entire range (**Figure 3-2E**) and the only allele present in Victoria. PhciDEFB10_Hap2 has a higher frequency in Mid NSW (0.474) and South NSW (0.353).

PhciDEFA1_Hap3 contains a stop codon resulting in the loss of two of the conserved cysteine residues. Interestingly, although this allele was not predicted to be anti-fungal like the other three alleles, it still had predicted anti-bacterial, anti-inflammatory and anti-viral activity (**Table 3-2**). PhciDEFA1_Hap3 is rare across most of the range, but it is the major allele in Victoria (**Figure 3-2B**).

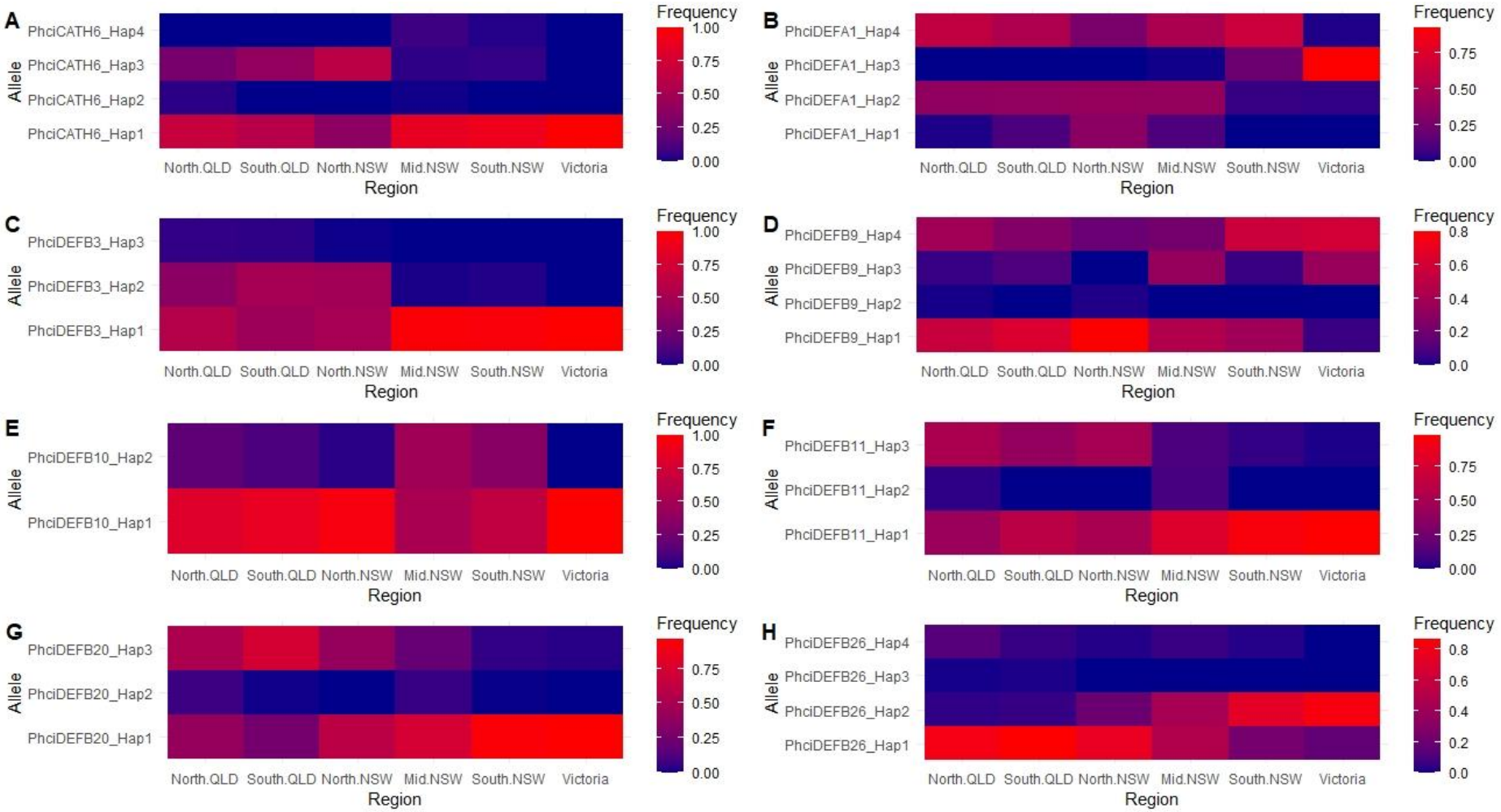


Figure 3-2. Heatmaps of allele frequencies in different geographic regions of (A) PhciCATH6, (B) PhciDEFA1, (C) PhciDEFB3, (D) PhciDEFB9, (E) PhciDEFB10, (F) PhciDEFB11, (G) PhciDEFB20 and (H) PhciDEFB26.

Table 3-2. *In silico* predictions of the effect of non-synonymous SNPs on AMP function, with predicted changes in function highlighted in red (negative) or green (positive). Changes in charge greater than 1 are highlighted in yellow. Italics indicate that the two haplotypes have the same amino acid sequence in the active peptide region. Haplotype frequency was calculated using PHASE (version 2.1.1), anti-bacterial, anti-inflammatory and anti-viral activity predicted using CSM-peptides webserver (Rodrigues et al., 2022) and anti-fungal activity predicted using the AntiFungal webserver (J. Zhang et al., 2021). The numbers and percentages in brackets indicate the probability of activity as either a frequency or a percentage depending on the webserver used. "Positive" indicates a probability value > 0.50.

Haplotype	Charge at pH7	Anti-bacterial	Anti-inflammatory	Anti-viral	Anti-fungal
PhciCATH3_Hap1	8.3	negative (0.4)	positive (0.79)	positive (0.52)	negative (22%)
PhciCATH3_Hap2	9.3	negative (0.43)	positive (0.79)	negative (0.49)	negative (20%)
PhciCATH3_Hap3	9.3	negative (0.43)	positive (0.79)	negative (0.49)	negative (20%)
PhciDEFA1_Hap1	1.7	positive (0.68)	positive (0.88)	positive (0.63)	positive (80%)
PhciDEFA1_Hap2	2.7	positive (0.7)	positive (0.88)	positive (0.64)	positive (87%)
PhciDEFA1_Hap3	0.8	positive (0.63)	positive (0.91)	positive (0.6)	negative (21%)
PhciDEFA1_Hap4	2.7	positive (0.7)	positive (0.83)	positive (0.68)	positive (98%)
PhciDEFB3_Hap1	3.2	positive (0.5)	positive (0.77)	negative (0.47)	negative (31%)
PhciDEFB3_Hap2	3.2	positive (0.5)	positive (0.77)	negative (0.47)	negative (31%)
PhciDEFB3_Hap3	3.2	positive (0.53)	positive (0.78)	negative (0.47)	negative (35%)
PhciDEFB7_Hap1	-1	positive (0.5)	positive (0.89)	positive (0.55)	negative (24%)
PhciDEFB7_Hap2	-1	positive (0.5)	positive (0.89)	positive (0.55)	negative (24%)
PhciDEFB7_Hap3	-0.3	negative (0.5)	positive (0.91)	negative (0.49)	negative (13%)
PhciDEFB9_Hap1	3.2	positive (0.52)	positive (0.68)	negative (0.46)	positive (83%)
PhciDEFB9_Hap2	2.2	negative (0.48)	positive (0.69)	negative (0.48)	positive (79%)
PhciDEFB9_Hap3	4.3	positive (0.55)	positive (0.68)	negative (0.44)	positive (71%)
PhciDEFB9_Hap4	4	positive (0.55)	positive (0.69)	negative (0.46)	positive (83%)
PhciDEFB10_Hap1	7.7	positive (0.61)	positive (0.74)	negative (0.47)	positive (71%)
PhciDEFB10_Hap2	5.7	positive (0.6)	positive (0.68)	negative (0.49)	negative (46%)

Haplotype	Charge at pH7	Anti-bacterial	Anti-inflammatory	Anti-viral	Anti-fungal
PhciDEFB11_Hap1	7.4	negative (0.39)	positive (0.85)	negative (0.34)	positive (85%)
PhciDEFB11_Hap2	7.4	negative (0.39)	positive (0.87)	negative (0.35)	positive (81%)
PhciDEFB11_Hap3	7.2	negative (0.41)	positive (0.87)	negative (0.38)	positive (87%)
PhciDEFB12_Hap1	15.9	negative (0.38)	positive (0.78)	negative (0.35)	positive (82%)
PhciDEFB12_Hap2	15.9	negative (0.34)	positive (0.76)	negative (0.34)	positive (83%)
PhciDEFB12_Hap3	15.9	negative (0.34)	positive (0.75)	negative (0.3)	positive (84%)
PhciDEFB12_Hap4	14.9	negative (0.33)	positive (0.75)	negative (0.32)	positive (71%)
PhciDEFB12_Hap5	14.9	negative (0.39)	positive (0.77)	negative (0.33)	positive (85%)
PhciDEFB16_Hap1	9	positive (0.59)	positive (0.82)	negative (0.38)	positive (90%)
PhciDEFB16_Hap2	9	positive (0.58)	positive (0.82)	negative (0.4)	positive (95%)
PhciDEFB19_Hap1	9.7	positive (0.66)	positive (0.83)	positive (0.5)	positive (92%)
PhciDEFB19_Hap2	8.7	positive (0.65)	positive (0.83)	positive (0.53)	positive (98%)
PhciDEFB19_Hap3	9.7	positive (0.65)	positive (0.84)	positive (0.53)	positive (99%)
PhciDEFB19_Hap4	8.7	positive (0.64)	positive (0.83)	positive (0.54)	positive (100%)
PhciDEFB20_Hap1	5	negative (0.32)	positive (0.77)	positive (0.71)	negative (29%)
PhciDEFB20_Hap2	5.2	negative (0.33)	positive (0.8)	positive (0.73)	negative (15%)
PhciDEFB20_Hap3	5	negative (0.32)	positive (0.77)	positive (0.71)	negative (29%)
PhciDEFB24_Hap1	1.7	positive (0.53)	positive (0.68)	positive (0.6)	positive (96%)
PhciDEFB24_Hap2	1.7	positive (0.53)	positive (0.68)	positive (0.62)	positive (89%)
PhciDEFB27_Hap1	2.7	positive (0.5)	negative (0.48)	positive (0.61)	negative (41%)
PhciDEFB27_Hap2	1.7	negative (0.46)	negative (0.48)	positive (0.58)	negative (37%)
PhciDEFB28_Hap1	2.7	positive (0.55)	negative (0.45)	positive (0.54)	positive (57%)
PhciDEFB28_Hap2	1.7	positive (0.6)	positive (0.73)	positive (0.55)	positive (55%)

1

Except for PhciDEFB27, all the AMPs were predicted to have anti-inflammatory properties but not all alleles (**Table 3-2**). For example, although PhciDEFB28_Hap2 is predicted to have anti-inflammatory properties but not PhciDEFB28_Hap1. PhciDEFB28_Hap2 is also the predominant allele, occurring at frequencies between 0.606 and 0.667 in every geographic region.

Of the four PhciDEFB9 alleles, all were predicted to have antifungal activity and all but one (PhciDEFB9_Hap2) were predicted to have antibacterial activity. PhciDEFB9_Hap2 was extremely rare and only occurred at very low frequencies in North Queensland (0.011) and North NSW (0.020).

CNVs

In total, 177 copy number variants (CNVs) were retained following filtering (152 duplications and 25 deletions) (**Table A2-5**). The most common CNV involved a duplication of PhciCATH5. Although this was found across the entire geographic range, it was most common in southern areas (**Figure 3-3**). In QLD, it only occurred in 9 out of 101 individuals (9%) and in Northern NSW it only occurred in 12 out of 76 (15%). However, in Mid NSW it was found in 20 out of 57 (35%), in South NSW it was found in 71 out of 112 (63%) and in Victoria it was found in 32 out of 72 (44%). The normalised read depth of this region ranged from 1.4 to 2.5, with an average of 1.87 (**Figure 3-3**). A normalised read depth of one represents the reference genome (ie a single copy of the gene, or two copies in a diploid organism) so these values suggest that whole gene duplications are common.

One defensin (PhciDEFB24) also exhibited duplications but it was much rarer. The duplication occurred in six individuals from Mid NSW, two from South QLD and one from North QLD (**Table A2-5**). In these individuals, normalised read depth ranged from 1.6 to 2.7 also indicating a whole gene duplication.

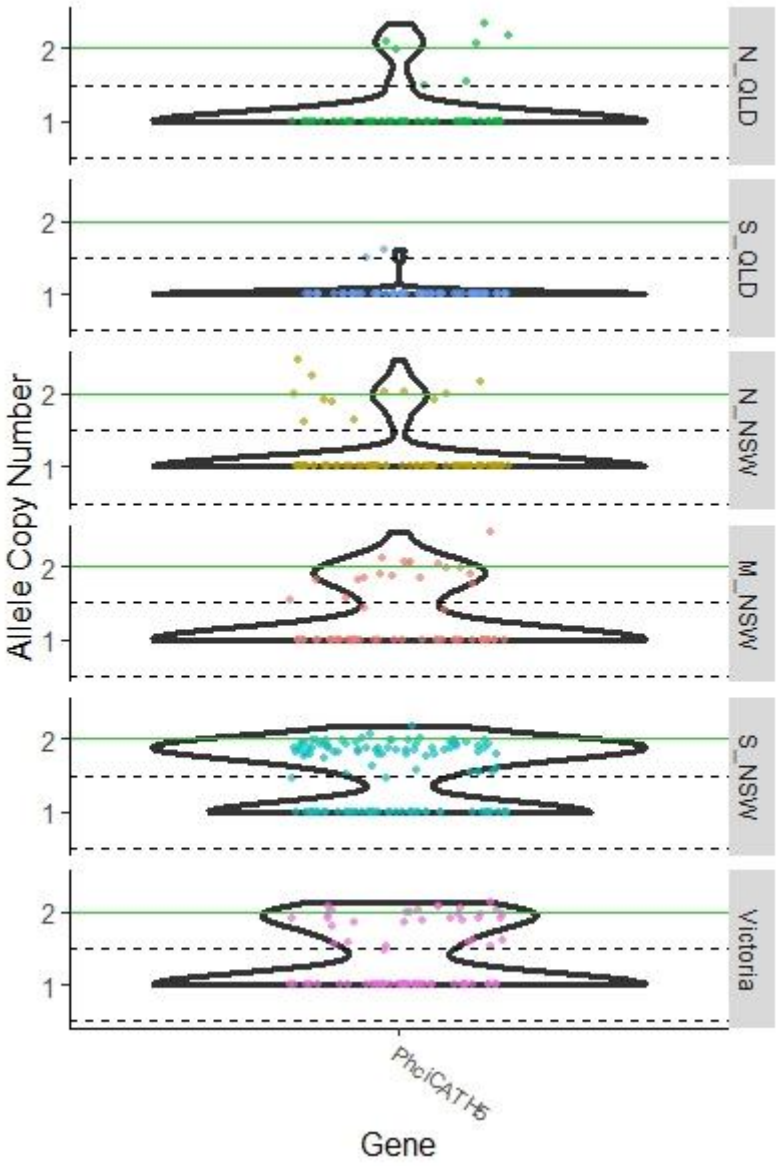


Figure 3-3. Violin plot depicting the most common CNV in koala AMPs (PhciCATH5) based on CNVnator analysis. Each dot represents an individual. Any individuals that did not have a statistically significant different normalised read depth are represented as having a normalised read depth equal to 1. Dotted lines represent the boundaries of a hemizygous deletion or duplication, and the green line represents the boundary of a whole gene duplication.

PhciDEFB23 exhibited deletions in 25 individuals from across the entire range: five from North QLD, two from South QLD, 12 from North NSW, two from Mid NSW, 3 from

South NSW and 1 from Victoria. They had a normalised read depth around 0.5 (half the coverage of the reference genome) suggesting a hemizygous deletion (**Table A2-5**).

Discussion

Here, we characterized AMP diversity across most of the koala range using 418 whole genomes from the east coast of Australia. We found non-synonymous SNPs in 14 AMPs that are predicted to change the function of the active mature peptide. We also found copy number variants involving duplications in two AMPs and deletions in one. Although AMP diversity has previously been investigated in birds and agricultural species, previous studies have been limited to studying a small number of genes and most of these have only looked at one or two populations. This represents the first comprehensive exploration of intraspecific AMP diversity in a wild mammal species across its range.

Overall AMP allelic diversity was lower than in other immune families we have investigated such as MHC (Silver, McLennan, et al., 2024) and toll like receptors (TLR) (Cui et al., 2025). This is not surprising given AMPs are not involved in antigen recognition, but rather utilise physiochemical properties such as cationic charge to exert their antimicrobial effects. It is also consistent with previous studies that have indicated ancient genes of the innate immune system are more constrained than genes of the adaptive immune system (Chapman et al., 2016; Clark & Wang, 1997).

The higher rate of SNPs observed in the defensins mirrors the findings of other studies (**Table A2-1**). Our findings that most non-synonymous SNPs in cathelicidins are in the prepro region rather than the active sequence are also consistent with the results of similar studies on pig and cattle cathelicidins (Ahn et al., 2022; Gillenwaters et al., 2009). This supports the idea that there are more constraints on the mature peptide than the prepro region.

Interestingly, no non-synonymous SNPs were identified in PhciCATH5, the only cathelicidin with documented antimicrobial activity *in vitro* (Peel et al., 2021) . This may indicate strong selective pressure to conserve this function. Selection analysis provided further evidence of the constraints on cathelicidin evolution, with mostly negative selection shown (**Table 3-1**). This is similar to other studies that have shown negative selection to be more common than positive selection in cathelicidins in birds (Cheng et al., 2015). However, we note that due to the short length of AMP sequences, tests of selection have limited statistical power.

Some AMPs did, however, exhibit higher levels of polymorphism. This mirrors the results of a similar study in mallards, which showed that one allele is predominant for most defensins but some have higher diversity (Chapman et al., 2016). Chapman et al investigated five defensin genes in mallards and other species of Anatidae (ducks, geese, and swans). Out of these five genes, the study only found one (*AvBD3b*) with allele frequencies that differed between geographic regions. In contrast, we investigated 38 genes and found allelic frequency differences in seven defensins (PhciDEFA1, PhciDEFB3, PhciDEFB9, PhciDEFB11, PhciDEFB12, PhciDEFB20 and PhciDEFB26) and one cathelicidin (PhciCATH6) (**Table A2-7**). The high Tajima's D value for PhciDEFB12 suggests this gene may be under balancing selection. Balancing selection is well documented in MHC genes but not ubiquitous in other immune genes (Minias & Vinkler, 2022). For example, while some TLR and AMP genes show signs of balancing selection, many others do not (Chapman et al., 2016; Chapman et al., 2019; Cui et al., 2025; Hollox & Armour, 2008; Kloch et al., 2018; Podlaszczuk et al., 2020). Our study provides evidence that although balancing selection is rare in koala AMPs it does occur.

Overall, northern populations contained more alleles than southern populations, which is consistent with previous studies that have shown southern koalas have lower genetic diversity (Johnson et al., 2018; Lott et al., 2022). This is the result of population crashes caused by the fur trade in the late 19th century, leading to a genetic bottleneck (Lee et al.,

2011). Northern and southern koala populations also display phenotypic variation across the range, including morphological (body size, fur depth) (Briscoe et al., 2015) and disease susceptibility (Legione et al., 2016; Patterson et al., 2015; Polkinghorne et al., 2013). Interestingly, despite their low genetic diversity, southern populations also tend to have reduced disease severity. This includes a lower prevalence of koala retrovirus (KoRV) (Simmons et al., 2012) and reduced severity of chlamydiosis (Legione et al., 2016; Patterson et al., 2015; Polkinghorne et al., 2013).

We recommend future studies investigate the functional effects of the different alleles identified here through population studies using disease metadata. Some of the AMPs that may have functional impacts include PhciDEFA1, PhciDEFB9, PhciDEFB10 and PhciCATH6. PhciDEFA1_Hap3 allele has a stop codon which is predicted to have a deleterious function. Interestingly, although our *in silico* results predict that this allele has lost its antifungal activity, it was the most common allele in Victoria. As PhciDEFA1 is expressed in the lactating mammary gland and likely present in the milk, we hypothesise that it may be involved in protecting the young (Peel et al., 2024).

PhciDEFB9_Hap1 is the most common allele in northern regions (QLD, North and Mid-NSW), while PhciDEFB9_Hap4 is most common in southern regions (South NSW and Victoria). Although our *in silico* testing predicted the different alleles have both anti-bacterial and anti-fungal activity, it is possible they vary in specificity and potency. We recommend future studies test these alleles *in vitro*.

PhciDEFB16_Hap2 has a SNP that changes one of the conserved cysteines that may lead to change in structure and consequently function. This haplotype was rare and only found in northern regions (frequency of 0.011 in North QLD and 0.083 in South QLD) (**Table A2-6**). The properties of PhciDEFB16 have not yet been characterised although our *in silico* testing predict both alleles have antibacterial, antifungal and anti-inflammatory properties (**Table 3-2**). It is an ortholog of the mouse defensin DEFB33, which is

expressed in the testes and may play a role in reproduction (Patil et al., 2005). It is possible that PhciDEFB16 may play a similar role in koalas, however further studies should investigate its gene expression in koala testes transcriptomes to confirm this.

Naturally occurring AMPs have long been deemed promising candidates for drug development (Lazzaro et al., 2020). In many cases, the AMP sequence serves as a template that can be modified for optimised activity such as increased activity or specificity. This is often achieved through minor amino acid substitutions (Higgs et al., 2007; Molhoek et al., 2010; Porto & Alencar, 2023). For example, substituting lysine for arginine has increased the antiviral properties of the scorpion peptide BmKn2, as well as the amphibian peptides maximim H5 and dermaseptin S9 (Chen et al., 2012; Wang et al., 2010). While substitutions leading to an increase in net peptide charge sometimes improve antimicrobial potency and specificity (Higgs et al., 2007; Molhoek et al., 2010; Porto & Alencar, 2023), an excessive increase can have the opposite effect by altering structure (Wang et al., 2019). By identifying different alleles of the various peptides in addition to the reference sequence, we have expanded the number of template sequences that can tested and optimised for drug development.

In addition to SNPs, CNVs such as duplications and deletions represent a major source of genetic variability between individuals. CNVs are often more common in immune genes than other regions of the genome (Bickhart et al., 2012; De Smith et al., 2009). The pleiotropic nature of AMPs means that such mutations can influence a wide variety of traits, including disease susceptibility, fertility, coat colour and production traits (da Silva et al., 2016; Machado & Ottolini, 2015; Yue et al., 2013; Zhang et al., 2009). Defensins in particular are recognised as being highly polymorphic in copy number, with duplications often involving whole clusters of genes. For example, the 8p23.1 locus in humans has been identified as a hotspot of genetic variation due to its high mutation rate (Bakar et al., 2009). This region encodes multiple defensins (DEFB4, DEFB103, DEFB104, DEFB105, DEFB106, DEFB107), with individual copy number ranging from two to twelve (Hollox

et al., 2003). High copy numbers in this region have been associated with inflammatory diseases such as psoriasis and hidradenitis suppurativa (Giamarellos-Bourboulis et al., 2016; Hollox et al., 2008).

CNVs involving cathelicidins have also been reported in livestock (Bickhart et al., 2012; Jeon et al., 2019) although these are less common than defensins. One example is the cattle cathelicidin indolicidin, encoded by the CATHL4 gene. Indolicidin induces cell death in the parasite *Leishmana donovani*, the pathogen responsible for Leishmaniasis (Bera et al., 2003). This gene has reported to have undergone recent duplications in indicine cattle, who have a higher copy number than taurine cattle (Bickhart et al., 2012). Interestingly, indicine cattle also have a higher parasite resistance than taurine cattle (Berman, 2011).

Similarly, CNVs were less common in koala cathelicidins than defensins. The only cathelicidin duplication we found was in PhciCATH5, a cathelicidin with activity against a range of microbes including *C. pecorum* (Peel et al., 2021). Chlamydia can manifest clinically as ocular and urogenital infections, resulting in blindness and infertility (Polkinghorne et al., 2013). Although *C. pecorum* affects koalas across their entire geographic range, disease prevalence and severity are both greater in northern populations (Legione et al., 2016; Patterson et al., 2015; Polkinghorne et al., 2013). Unlike koalas from NSW and QLD, *C. pecorum* has not been detected in ocular swabs from Victorian koalas (Patterson et al., 2015). Although mild "wet-bottom" disease (urine staining the rump due to urinary tract infection) has been reported in Victorian koalas, ocular disease has not (Patterson et al., 2015). Interestingly, the PhciCATH5 duplication was more common in southern NSW and Victoria, where it was found in 63% and 44% of the sequenced genomes respectively. PhciCATH5 duplications may be linked to the selective pressure of chlamydia, similar to the expansions observed in cattle indolicidin. However, additional work is required to confirm this pattern. We recommend population studies

using disease metadata to determine if there is a significant association between PhciCATH5 duplications and disease susceptibility.

Conclusion

In conclusion, we have characterised AMP diversity across a species range for the first time in mammalian wildlife and identified differences between populations at the nucleotide, amino acid and copy number level. We have also identified variants of interest for further investigation through population studies and *in vitro* testing.

Statements and Declarations

C.J.H. is a member of the NSW Expert Panel for Koalas, an advisory panel to the NSW government. The authors declare no conflict of interest.

Acknowledgements

This work was supported by the Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science (CE200100012). The Koala Genome Survey was funded by the NSW Government and the Australian Government's Bushfire Recovery for Wildlife and their Habitats program (GA2000526). Further support was provided by The University of Sydney, Amazon Web Services Open Data Sets, Ramaciotti Centre for Genomics, and Illumina.

Author Contributions

CP designed the study with guidance from EP and LWS. CP and LWS performed all data analysis with assistance from EP. CJH and KB sourced funding and undertook project management and supervision. CP wrote the main manuscript text with feedback and revisions provided by all authors.

CHAPTER FOUR:

ATTACK OF THE CLONES: NO EVIDENCE FOR DISTINCT TRANSCRIPTOMIC SUBGROUPS OF DEVIL FACIAL TUMOUR DISEASE (DFTD)

CHAPTER 4. ATTACK OF THE CLONES: NO EVIDENCE FOR DISTINCT TRANSCRIPTOMIC SUBGROUPS OF DEVIL FACIAL TUMOUR DISEASE (DFTD)

4.1 BACKGROUND

Chapter 4 comprises the following published manuscript:

Petrohilos, C., Peel, E., Batley, K. C., Fox, S., Hogg, C. J., & Belov, K. (2025). No Evidence for Distinct Transcriptomic Subgroups of Devil Facial Tumor Disease (DFTD). *Evolutionary Applications*, *18*(4), e70091.

DFTD is a contagious cancer that is comprised of six genotypic clades. This chapter uses transcriptomics and unsupervised clustering to determine if these various clades have different gene expression profiles. The results indicate there are no distinct transcriptomic subgroups, which has implications for vaccine development and wildlife management.

Carolyn J. Hogg and Katherine Belov conceptualised and sourced funding for the study. Emma Peel, Kimberley C. Batley, Carolyn J. Hogg, Katherine Belov and I designed the study. Samantha Fox either collected, or coordinated collection, of the samples and provided access to individual metadata. Kimberley C. Batley and I performed RNA extractions. I performed analyses and wrote the main manuscript text with feedback and revisions provided by all authors.

4.2 MANUSCRIPT

No Evidence for Distinct Transcriptomic Subgroups of Devil Facial Tumor Disease (DFTD)

Cleopatra Petrohilos[1,2], Emma Peel[1,2], Kimberley C Batley[1], Samantha Fox[3], Carolyn J Hogg[1,2], Katherine Belov[1,2]

[1]School of Life and Environmental Sciences, The University of Sydney, Sydney, New South Wales, Australia
[2]Australian Research Council Centre of Excellence for Innovations in Peptide & Protein Science, The University of Sydney, Sydney, NSW, Australia
[3]Save the Tasmanian Devil Program, Department of Natural Resources and Environment, Hobart, TAS, Australia
Corresponding Author: Prof Carolyn Hogg, The University of Sydney, Sydney, NSW, 2006, Australia
Email: carolyn.hogg@sydney.edu.au

Keywords: wildlife disease, contagious cancer, unsupervised clustering

Abstract

Contagious cancers represent one of the least understood types of infections in wildlife. Devil Facial Tumour Disease (comprised of two different contagious cancers, DFT1 and DFT2) has led to an 80% decline in the Tasmanian devil (*Sarcophilus harrisii*) population at the regional level since it was first observed in 1996. There are currently no treatment options for the disease and research efforts are focused on vaccine development. Although DFT1 is clonal, phylogenomic studies have identified different genetic variants of the pathogen. We postulated that different genetic strains may have different gene expression profiles and would therefore require different vaccine components. Here, we aimed to test this hypothesis by applying two types of unsupervised clustering (hierarchical and k-means) to 35 DFT1 transcriptomes selected from the disease's four major phylogenetic clades. The two algorithms produced conflicting results and there was low support for either method individually. Validation metrics, such as the Gap statistic method, the Elbow

method, and the Silhouette method, were ambiguous, contradictory, or indicated that our dataset only consisted of a single cluster. Collectively, our results show that the different phylogenetic clades of DFT1 all have similar gene expression profiles. Previous studies have suggested that transcriptomic differences exist between tumours from different locations. However, our study differs in that it considers both tumour purity and genotypic clade when analysing differences between DFTD biopsies. These results have important implications for therapeutic development as they indicate that a single vaccine, or treatment approach, has the potential to be effective for a large cross section of DFT1 tumours. As one of the largest studies to use transcriptomics to investigate phenotypic variation within a single contagious cancer, it also provides novel insight into this unique group of diseases.

Introduction

As the biodiversity and climate crises deepen, infectious wildlife diseases are becoming more prevalent (El-Sayed & Kamel, 2020). One of the rarest and least understood modalities of infection is contagious cancer, tumour cells that can be transmitted from one individual to another (Metzger & Goff, 2016). To date, transmissible cancers have only been documented in eleven species (two vertebrates and nine invertebrates), although the true number is estimated to be much higher (Bruzos et al., 2023; Dujon et al., 2021; Hallmann et al., 2022; Hammel et al., 2022; Hart et al., 2023; Metzger & Goff, 2016; Metzger et al., 2016; Santamarina et al., 2024; Yonemitsu et al., 2023).

Contagious cancers have devastating impacts on the hosts' ecological function and commercial industries such as aquaculture (Dujon, Schofield, et al., 2020). A contagious cancer that is currently of particular conservation concern is Devil Facial Tumour Disease (DFTD). This disease affects the largest marsupial carnivore, the Tasmanian devil (*Sarcophilus harrisii*), and has caused population crashes of up to 80% across the species' range (Lazenby et al., 2018). There are now two forms of DFTD: DFT1 that first arose in

the 1990s in the northeast of Tasmania (Loh et al., 2006), and DFT2 that was first detected in the southeast of Tasmania in 2014 (Pye, Pemberton, et al., 2016). DFT1 has spread across the island state of Tasmania, whilst DFT2 remains restricted to the southeast of Tasmania (James et al., 2019; Lazenby et al., 2018). Here we focus on DFT1.

DFT1 is almost always fatal with very few regressions of the disease documented (Margres et al., 2020; Wright et al., 2017). A genome wide association analysis indicated that regression may be caused by variants near the *PAX3* gene that disrupt angiogenesis to tumours (Wright et al., 2017), while a comparative genomics approach also identified a mutation in the putative tumour suppressor *RASL11A* (Margres et al., 2020). There are currently no treatments for either forms of DFTD, although candidate therapeutics have been tested *in vitro* (Fernandez-Rojo et al., 2018; Kosack et al., 2019; Patchett et al., 2018; Petrohilos et al., 2023; Stammnitz et al., 2018) and in murine models for DFT1 (Ikonomopoulou et al., 2021). There are also research efforts focused on developing a vaccine (Flies et al., 2020; Kayigwe et al., 2022; Pye et al., 2021; Pye et al., 2018; Tovar et al., 2017) A number of immunostimulatory adjuvants have been identified (Patchett et al., 2017; Pye et al., 2018) that may be used in combination with recombinant DFT proteins to trigger an immune response (Flies et al., 2020). The aim is to deliver this via an oral bait vaccine platform similar to the successful rabies vaccine (Flies et al., 2020), with trials having shown that both captive and wild devils will consume placebo baits (Dempsey et al., 2022).

Cancer – even tumours that affect the same tissue type – is an umbrella term rather than a single disease. For this reason, classifying tumours into subtypes based on molecular differences is important for therapeutic development and maximising treatment efficacy (Collisson et al., 2019; Zhao et al., 2020). However, transmissible cancers are unique in this respect. Contagious cancers are clonal and so do represent a single disease. Previous studies have observed that DFT1 is constantly evolving yet remains remarkably stable for a cancer (Deakin & Belov, 2012; Deakin et al., 2012; Ingles & Deakin, 2015; Kwon et al.,

2020; Murchison et al., 2012). Although it has a highly rearranged genome, additional mutations such as translocations and aneusomy are rare in primary tumours (Pearse et al., 2012). Interestingly, these mutations are common in metastases and long-term cell culture of DFT1, which may indicate a strong selective pressure *in vivo* (Pearse et al., 2012). This may indicate that mutations exceeding a particular threshold result in cells that are either not viable or not transmissible (Pearse et al., 2012). The one exception to this is tetraploidy, with whole genome duplications occurring multiple times in the disease's history (Kwon et al., 2020). Polyploidy is likely advantageous for DFT1 as it can mask Muller's ratchet – the accumulation of deleterious mutations in asexual organisms that leads to genomic erosion (Ujvari et al., 2014).

Early evidence suggested four distinct karyotypic strains of DFT1 (Deakin et al., 2012), categorised by the presence or absence of five "marker" chromosomes (Marker 1 – Marker 5), (Deakin et al., 2012). One study suggested the strains had differential growth rates *in vitro*, however, this conclusion was based on a low number of replicates (n=1 to 3 amongst the four strains) (Pearse et al., 2012). Subsequent studies have since shown minimal microsatellite (Pearse et al., 2012), epigenetic (Ingles & Deakin, 2015; Ujvari et al., 2013) or cytogenetic (Deakin et al., 2012) differences between the four strains. In 2020, Kwon et al showed that the marker characterizing strains two, three and four (Marker 5) is highly unstable and has been lost at least 27 times between 2003 and 2018 (Kwon et al., 2020). This suggests that classifying tumours into these four strains may not be a biologically meaningful way to categorise the disease due to the unreliability of observing the defining markers.

More recently, phylogenomic methods have identified six phylogenetic clades (A1, A2, B, C, D and E) of DFT1, although the latter two (D and E) have failed to persist in the wild and clade A1 has not been detected since 2012 (Kwon et al., 2020; Stammnitz et al., 2023). There does not appear to be any association between phylogenetic clades and karyotypic strains. Kwon et al. (2020) used a robust combination of mitochondrial

variants, nuclear variants, and copy number variants to build one of the largest tumour phylogenies to date (over 600 genomes collected between 2003 and 2018). Although such phylogenomic methods are useful for tracing the evolutionary trajectory of the disease, genetic differences may not necessarily translate to functional differences. Many variants may be effectively neutral, for example those that are synonymous, intronic or in genes that are not expressed. Instead, phenotypic differences between tumour subtypes can be characterised using transcriptomics. RNAseq data is closely linked with phenotype (Guinney et al., 2015) as it only measures the genes that are expressed within a sample. Unlike genomic variation, transcriptomic differences do equate to functional differences. RNAseq data has been used to identify potential antigens for immunotherapy in a range of human cancers (Wu, Duan, et al., 2022; Wu, Qin, et al., 2022) and transcriptomic assays such as MammaPrint, Oncotype DX and PAM50 are frequently used to guide treatment of breast cancer (Chaudhuri et al., 2021). In the case of DFT1, the number of transcriptomic subtypes may influence how many neoantigens must be targeted by a vaccine. It may also aid in conservation management decisions such as translocating diseased animals to reduce the spread of different strains.

RNAseq has been used to investigate phenotypic variation within DFT1. One study investigated transcriptomic differences in tumours from different geographic regions (Kozakiewicz et al., 2021) but did not seek to identify molecular subtypes. The dataset was also limited to 19 samples from 3 locations. Studies have also raised concerns about the effectiveness of a single DFT1 vaccine due to the potential heterogeneity of the disease (Pearse et al., 2012). Here, we generated a much larger RNAseq dataset: 35 DFT1 samples from 12 locations across central Tasmania where DFT1 has been present since 2003. These samples represent all four phylogenomic clades and karyotypic strains (Kwon et al., 2020). Our aim was to determine if DFT1 has distinct patterns of gene expression that categorise tumours into different subgroups, and if so, do these align to those noted in the phylogenomic studies to inform vaccine and therapeutic treatment development.

Methods

Data collection

Biopsies from DFT1 primary tumours were collected by the Save the Tasmanian Devil Program between 2006 and 2015 from multiple sites across Tasmania as part of their annual monitoring program and shared with us for the purposes of this study (**Table A3-1**). The biopsies selected for this study were from tumours that had previously been genotyped and assigned to a phylogenetic clade, as well as to a karyotypic strain (Kwon et al., 2020). They included representative samples from each of the four major clades: 11 samples from clade A1, 10 from clade A2, 10 from clade B and 4 from clade C (the smallest of the four major clades). Samples from clades D and E were not included in our study as these clades have failed to persist in the wild (Kwon et al., 2020). Our dataset also contained representatives from all major karyotypic strains (1-4).

Total RNA was extracted from DFT1 biopsies using a Qiagen RNeasy mini kit (Qiagen, Cat. No. 74104). RNA quality was assessed using the RNA nano 6000 kit on the Bioanalyzer (Agilent) and samples with an RNA integrity score (RIN) greater than seven were submitted to Ramaciotti Centre for Genomics (The University of New South Wales) for sequencing. All samples underwent TruSeq stranded mRNA library prep (Illumina) and were sequenced as paired-end 150bp reads across an SP flowcell on the Illumina NovaSeq6000. This resulted in 50 to 107 million read pairs per sample.

As a control for comparison, raw RNAseq reads from Tasmanian devil healthy tissue and DFT1 biopsies were downloaded from NCBI (BioProject PRJEB34650 (Stammnitz et al., 2023) and PRJEB28680 (Patchett et al., 2020)). The details of the samples downloaded are included in Supplementary File 1A of the published manuscript here.

Data Analysis

All samples were quality assessed using FastQC v0.11.8 (Andrews, 2010) and trimmed of low quality and adaptor sequences using Trimmomatic v0.39 (Bolger et al., 2014). Reads were then aligned to the Tasmanian devil reference genome mSarHar1.11 (NCBI: GCF_902635505.1) (Stammnitz et al., 2023) using STAR v2.7.8a (Dobin et al., 2013). Default parameters were used for the DFT1 reads from this study and the samples from Patchett et al. (2020). As the samples from Stammnitz et al. (2023) had a much shorter read length (75bp) than those generated in this study (150bp), we adjusted the alignment parameters to improve mapping rate and make the samples comparable (--sjdbOverhang 74 -outFilterScoreMinOverLread 0.1 --outFilterMatchNminOverLread 0.1).

Some of the reads downloaded from NCBI had very low alignment rates (Supplementary File 1A available online here) which may indicate poor sequencing accuracy or DNA contamination (Conesa et al., 2016). For this reason, only data that consisted of at least two biological replicates with at least 75% uniquely mapped reads was retained for further analysis. This resulting dataset consisted of 54 samples from 9 different tissues including: 37 DFT1 samples (including the 35 generated here and two from Patchett et al. (2020)), two axillary nerve (Stammnitz et al., 2023), two bone marrow (Stammnitz et al., 2023), two brain (Patchett et al., 2020), two cerebellum (Stammnitz et al., 2023), two cerebrum (Stammnitz et al., 2023), two spleen (Patchett et al., 2020), three testes (Patchett et al., 2020; Stammnitz et al., 2023) and two trigeminal nerve (Stammnitz et al., 2023).

Alignments were summarised into gene counts using featureCounts in the subread package v1.5.1 (Liao et al., 2014). Gene counts were then input into R v4.1.3 (Team, 2022). As the samples in Patchett et al. (2020) consisted of technical replicates, these were summed prior to further analysis. Lowly expressed genes with a total count less than 50 across all samples were excluded from the analysis. Normalisation factors were calculated using

trimmed mean of M values (TMM) to account for differences in raw library sizes (edgeR v3.36.0 (Robinson et al., 2010)).

To ensure that the different method of sequencing used in the different studies did not bias the results, a redundancy analysis (RDA) was performed using the vegan package v 2.6-6.1 (Oksanen et al., 2018) in R. Gene counts (normalised to log counts per million) were used as the response variable, and tissue type and study were both used as the predictor variables in the model. The significance of each term was tested using the anova.cca function in vegan with 999 permutations. Variance partitioning was then conducted using the varpart function in vegan to assess the percentage of variance explained by each response variable (ie tissue type and study). The significance of the variance partitioning was also assessed using anova.cca with 999 permutations.

Multidimensional scaling (MDS) was used to check variation across the samples (limma v3.50.3 (Ritchie et al., 2015)). Four DFT1 samples were identified as outliers using two different methods (MDS and hierarchical clustering) (**Figure A3-1 – Figure A3-2**). These four samples clustered more closely to the spleen samples (N=2) on the first dimension and the bone marrow samples (N=2) on the second dimension than other DFT1 samples (**Figure A3-3**). Both spleen and bone marrow are tissues that contain a high proportion of immune cells, so these results suggest the section of the biopsy from which RNA was extracted contained a higher proportion of immune cells than DFT1 cells. For this reason, these four outliers were excluded from subsequent analysis.

Purity Estimation

Tumour purity was further estimated by using the somatic substitution variant allele fraction (VAF) distributions of each sample as used previously for DFTD (Kwon et al., 2020; Stammnitz et al., 2023). VAF is the proportion of reads covering a particular variant and is calculated by n_s / N_s (where n_s is the number of reads containing the variant and N_s is the total number of reads) (Dentro et al., 2021). VAF reflects the zygosity of a locus:

homozygous reference loci should have a VAF around 0, heterozygous loci around 0.5 and homozygous variant loci around 1 (Strom, 2016). As most mutations in DFT1 exist in a heterozygous state, the average VAF of a clonal population of DFT1 cells is expected to be 0.50 (Stammnitz et al., 2023). For this reason, VAF_{HET} (defined as either the mode or median VAF of all heterozygous variants in a sample) has been used to measure the purity of DFT1 samples in previous studies (Kwon et al., 2020; Stammnitz et al., 2023).

First, samples were preprocessed with Opossum (Oikkonen & Lise, 2017). Many variant callers do not perform optimally when applied directly to RNAseq data, as splice junctions cause the reads to be split and lose information (Oikkonen & Lise, 2017). Opossum modifies the reads prior to splitting to ensure that information is retained and improve variant calling sensitivity. We then called variants using Platypus v0.1.5 (Rimmer et al., 2014) with settings minPosterior=0, minBaseQual=30, badReadsThreshold=30, badReadsWindow=15, minFlank=0 and minReads=1. As Platypus does not work on chromosomes longer than 536 Mb and chromosome 1, 2 and 3 of the reference genome (NCBI GCF_902635505.1) are 611 to 716 Mb, we split these chromosomes into 500 Mb windows prior to variant calling.

Single nucleotide variants (SNVs) were extracted, and any variants flagged as badReads, sb (strand bias), MMLQ (median minimum base quality for bases around variant) < 30 and QUAL < 20 were excluded. We also excluded any variants within 5bp from a simple repeat region (as annotated by TandemRepeatsFinder v4.09.1 (Benson, 1999)), within 500bp from contig start/end or within 1000bp of scaffold start/end.

Variants were considered to be somatic if they were one of the 1,311 "trunk" variants identified by Stammnitz et al. (2023). Of these 1,311 mutations, 421 occur in exonic regions. These "trunk" variants are present in all DFT1 tumours but absent from all healthy Tasmanian devil genomes examined so far. We classified a variant as "present" in our samples if it was supported by >3 reads.

For each sample, we calculated the VAF of each variant by dividing TR (Total reads supporting the variant) by TC (Total coverage at the locus). For example, if there were 3 reads supporting a particular SNV and 5 total reads at that locus, the VAF of that SNV would be 0.6. We plotted the VAF distribution for each sample and defined VAF_{HET} as the maximum density of the heterozygous peak similarly to Stammnitz et al. (2023). An example distribution plot is included in **Figure A3-4**. We then estimated each sample's purity (ρ) using the $\rho = 2 * VAF_{Het}$ formula that has been used by Kwon et al. (2020) and Stammnitz et al. (2023). Distribution plots for all samples are included in **Figure A3-5**.

While this method for estimating tumour purity has its limitations due to the low number of trunk variants in expressed genes, it represents the best available method due to the nature of the data. Ideally, single-cell RNA sequencing would be used to accurately assess purity by generating gene expression profiles for tumour and multiple host cell types to assign cell populations. However, single-cell RNAseq data is not available for DFTD or Tasmanian devils. The logistics of collecting biopsies from wild animals currently renders such a method difficult at this time as single-cell RNAseq requires fresh biopsies.

Only samples with a purity > 80% were retained for further analysis (n=27). This threshold was chosen as it represented a balance between maximizing accuracy and maximizing sample size. Purity estimates for all samples are included in **Table A3-2**. For each sample, the number of reads supporting the reference and alternative alleles for each variant are included in Supplementary File 1C available online here.

For comparison, the analysis was also run separately on the full dataset (excluding the four outliers).

Unsupervised clustering

Upper quartile normalization was applied to account for differences across sequencing lanes (EDAseq v2.28.0 (Risso et al., 2011)) and transcripts per million (TPM) was calculated to account for differences in gene lengths (scater v1.22.0 (McCarthy et al., 2017)). Counts were then log-transformed for variance stabilization across the samples.

Unsupervised clustering encompasses a variety of machine learning algorithms that reveal meaningful groups in unlabeled data (Dalmaijer et al., 2022). They are commonly used to identify biologically relevant subsets of data such as cancer subtypes (Guinney et al., 2015; Lapointe et al., 2004; Lei et al., 2013; Oh et al., 2018; Robertson et al., 2020; Sekiguchi et al., 2020; Tan et al., 2011). One caveat of these methods is that they will identify clusters, regardless of whether they exist, so it is critical to test the robustness of the resulting subsets (Adolfsson et al., 2019). Common methods include: data visualisation (Tan et al., 2011), using multiple algorithms to identify a consensus of subtypes (Guinney et al., 2015; Lei et al., 2013) and clustering validation indices (Xiong et al., 2018).

Here, we used two methods of dimension reduction for preliminary data visualization: MDS and t-Distributed Stochastic Neighbor Embedding (t-SNE) (Van der Maaten & Hinton, 2008). t-SNE was performed using Rtsne v.0.16 with 30 initial principal components and a perplexity (a hyperparameter that reflects the density of the data) of 9. This was run 500 times and the optimal run identified as the one with the lowest KL-divergence. We used heatmaps to visualise potential patterns in the data. We first used the 1000 most variable genes (those with the highest standard deviation), then iteratively halved this number to use progressively fewer genes (500, 250, 100). This was to test for a genetic signal driven by a small number of genes that may be lost in the noise of the entire dataset.

The 1000 most variable genes were then used as input to two different clustering algorithms: hierarchical clustering (using Euclidean distance and Ward's D) and k-means

clustering (**Figure A3-6**). For each method, the optimum number of clusters in the data was determined using three different cluster validation metrics in factoextra v1.0.7 (Kassambara & Mundt, 2017): Silhouette method (Rousseeuw, 1987), elbow method and Gap statistic method (Tibshirani et al., 2001). K=4 was chosen for k-means clustering based on the results of the Silhouette method (see results).

The hierarchical clustering results were then further tested using two methods: pvclust v2.2-0 (Suzuki & Shimodaira, 2006) was used for bootstrap analysis using 10,000 bootstraps, and sigclust2 v1.2.4 (Kimes et al., 2017) was used to perform a Monte Carlo simulation based significance testing.

The results from the two clustering methods (hierarchical and k-means) were compared visually using the R package dendextend 1.17.1 (Galili, 2015).

Results

Over 97% of all reads were retained following trimming, with a mean of 78 million reads per sample (range: 13 – 142 million reads). The reads generated in this study had high rates of uniquely mapped reads (75% - 86%). The full list of number of reads and mapping statistics for each sample are included in Supplementary File 1B available online here.

The RDA and variance partitioning analyses indicated that different sequencing methods did not significantly influence estimation of gene counts (**Table A3-3**). While tissue type was significant as a predictor variable (p=0.001), study was not (p=0.2). This shows that sequencing type does not introduce any bias in gene expression data and combining data from multiple studies is an appropriate method for maximising sample size.

Exploratory analysis showed a difference in gene expression between DFT1 and healthy tissue, with DFT1 samples largely clustering together in the MDS plot, and minimal differences between the various types of healthy tissue biopsies, including those that were

sequenced in different studies. This was achieved after four outliers (**Figure A3-1 – Figure A3-2**) and other samples with purity <80% were removed from the DFT1 dataset (**Figure A3-3**). In particular, the four main outliers clustered most closely with the spleen and bone marrow samples. As both spleen and bone marrow are haemopoetic organs that contain immune cells, we hypothesise that the sections of the biopsies used in this study showed a high level of immune cell infiltration.

Neither MDS (**Figure 4-1A**) nor t-SNE (**Figure 4-1B**) suggested any underlying pattern in our DFT1 dataset. If distinct phenotypic subsets existed, we would expect to see datapoints with similar gene expression clustering together. Clade A2 exhibited the tightest clustering of all groups in the MDS plot (**Figure 4-1A**), however one Clade A2 sample showed tighter clustering with a Clade B sample. Clade A2 also did not exhibit tight clustering in t-SNE (**Figure 4-1B**), indicating a lack of consistency. Similarly, none of the heatmaps show the type of clear mosaic pattern associated with distinct phenotypic groups, even when the data were reduced to the 100 most variable genes (**Figure 4-2**).

The lack of phenotypic subgroups amongst the DFT1 tumour sequences was further confirmed by the lack of a clear consensus between the hierarchical and k-means clustering on the number of clusters within the dataset. However, the three validation indices (Gap, elbow, and Silhouette) generated contradictory values for k when applied to both hierarchical and k-means clustering. The Gap statistic indicated the dataset contained no clusters (**Figure 4-3A-B**), the elbow method results were ambiguous (**Figure 4-3C-D**), and the Silhouette method indicated that the optimum number of clusters was 10 (the maximum) when applied to hierarchical clustering but k = 4 for k-means clustering (**Figure 4-3E-F**). When k-means clustering was performed with k set at 4 (**Figure 4-4**), each cluster contained samples from at least two different clades.

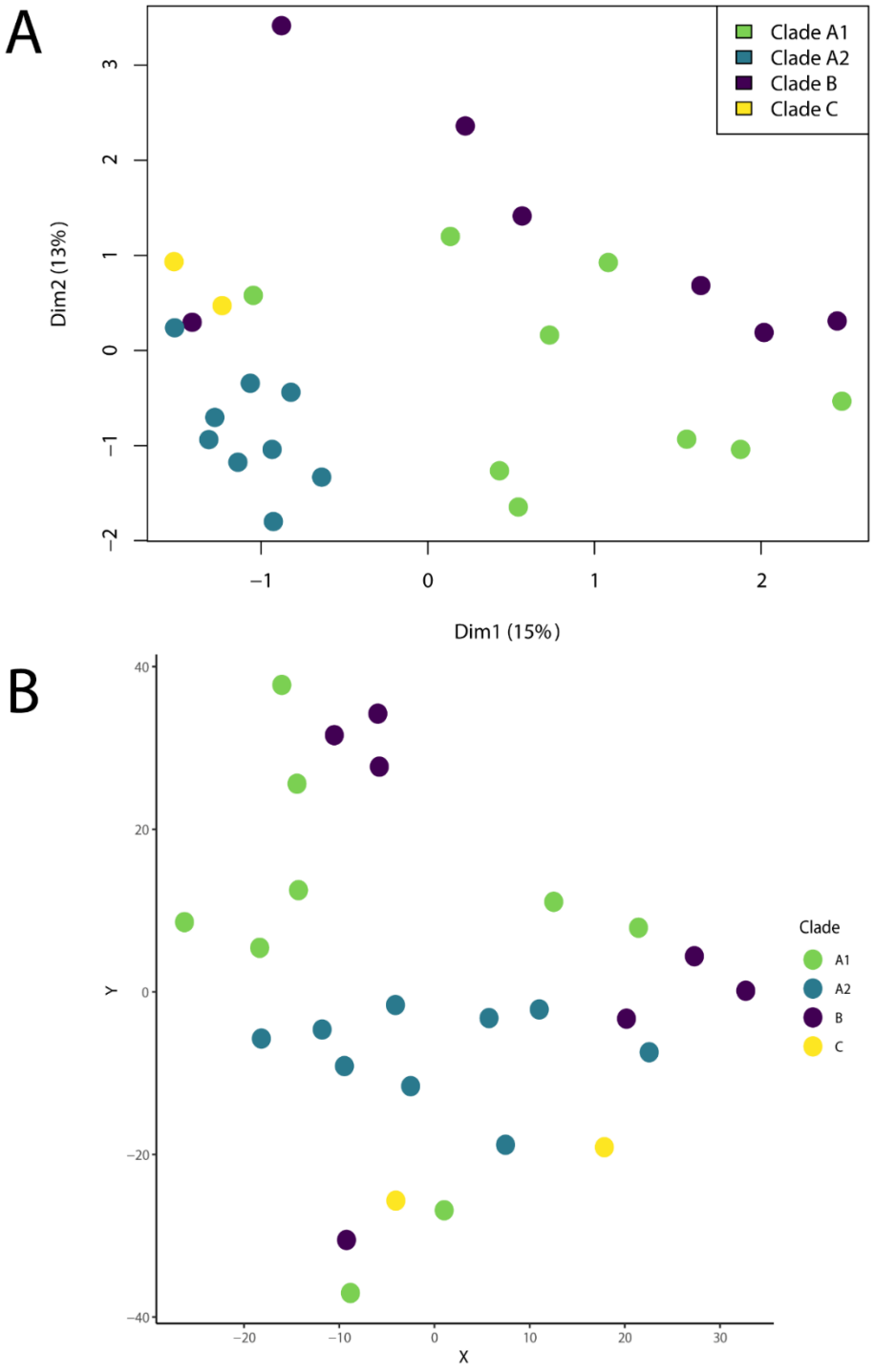


Figure 4-1. (A) MDS plot and (B) t-SNE plot of distances between TMM normalised gene counts for each DFT1 sample. Neither method of dimension reduction indicated strong separation between the clusters. Clade A2 exhibited the tightest clustering of all groups, but this did not include all samples in A2. Samples are coloured by genotypic clade.

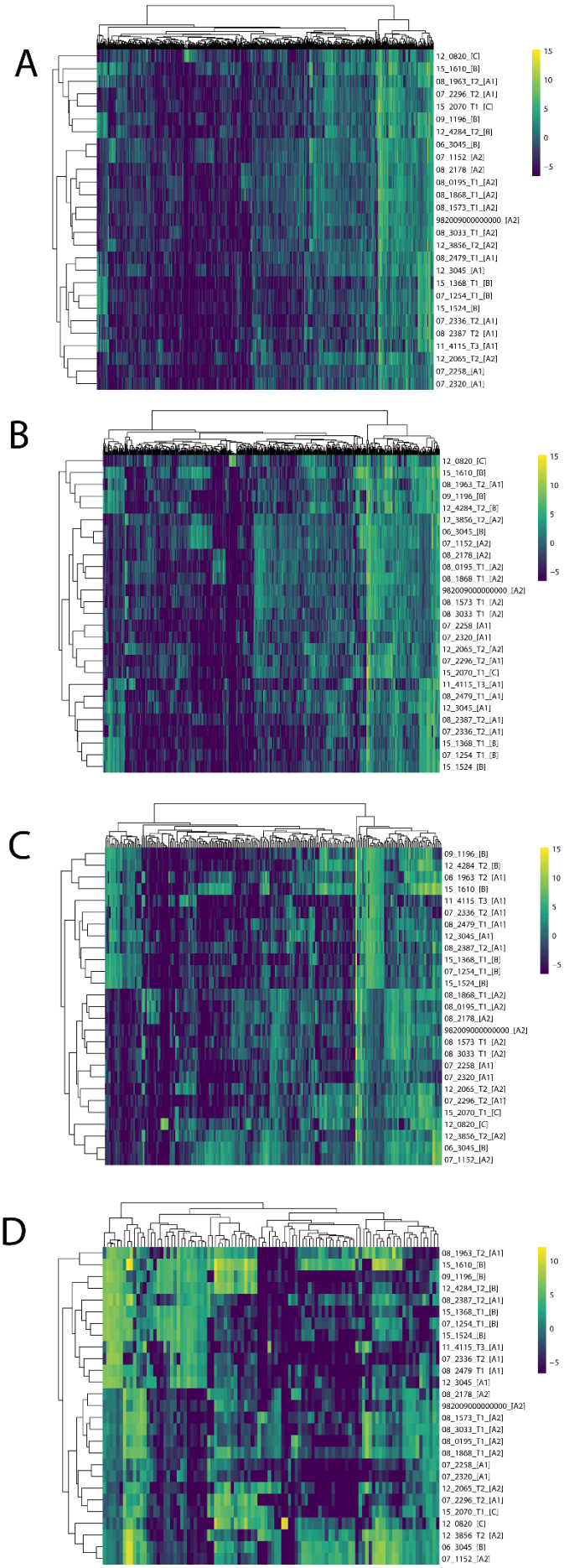

A
12_0820_[C]
15_1610_[B]
08_1963_T2_[A1]
07_2296_T2_[A1]
15_2070_T1_[C]
09_1196_[B]
12_4284_T2_[B]
06_3045_[B]
07_1152_[A2]
08_2178_[A2]
08_0195_T1_[A2]
08_1868_T1_[A2]
08_1573_T1_[A2]
982009000000000_[A2]
08_3033_T1_[A2]
12_3856_T2_[A2]
08_2479_T1_[A1]
12_3045_[A1]
15_1368_T1_[B]
07_1254_T1_[B]
15_1524_[B]
07_2336_T2_[A1]
08_2387_T2_[A1]
11_4115_T3_[A1]
12_2065_T2_[A2]
07_2258_[A1]
07_2320_[A1]
15
10
5
0
−5
B
12_0820_[C]
15_1610_[B]
08_1963_T2_[A1]
09_1196_[B]
12_4284_T2_[B]
12_3856_T2_[A2]
06_3045_[B]
07_1152_[A2]
08_2178_[A2]
08_0195_T1_[A2]
08_1868_T1_[A2]
982009000000000_[A2]
08_1573_T1_[A2]
08_3033_T1_[A2]
07_2258_[A1]
07_2320_[A1]
12_2065_T2_[A2]
07_2296_T2_[A1]
15_2070_T1_[C]
11_4115_T3_[A1]
08_2479_T1_[A1]
12_3045_[A1]
08_2387_T2_[A1]
07_2336_T2_[A1]
15_1368_T1_[B]
07_1254_T1_[B]
15_1524_[B]
15
10
5
0
−5
C
09_1196_[B]
12_4284_T2_[B]
08_1963_T2_[A1]
15_1610_[B]
11_4115_T3_[A1]
07_2336_T2_[A1]
08_2479_T1_[A1]
12_3045_[A1]
08_2387_T2_[A1]
15_1368_T1_[B]
07_1254_T1_[B]
15_1524_[B]
08_1868_T1_[A2]
08_0195_T1_[A2]
08_2178_[A2]
982009000000000_[A2]
08_1573_T1_[A2]
08_3033_T1_[A2]
07_2258_[A1]
07_2320_[A1]
12_2065_T2_[A2]
07_2296_T2_[A1]
15_2070_T1_[C]
12_0820_[C]
12_3856_T2_[A2]
06_3045_[B]
07_1152_[A2]
15
10
5
0
−5
D
08_1963_T2_[A1]
15_1610_[B]
09_1196_[B]
12_4284_T2_[B]
08_2387_T2_[A1]
15_1368_T1_[B]
07_1254_T1_[B]
15_1524_[B]
11_4115_T3_[A1]
07_2336_T2_[A1]
08_2479_T1_[A1]
12_3045_[A1]
08_2178_[A2]
982009000000000_[A2]
08_1573_T1_[A2]
08_3033_T1_[A2]
08_0195_T1_[A2]
08_1868_T1_[A2]
07_2258_[A1]
07_2320_[A1]
12_2065_T2_[A2]
07_2296_T2_[A1]
15_2070_T1_[C]
12_0820_[C]
12_3856_T2_[A2]
06_3045_[B]
07_1152_[A2]
10
5
0
−5

Figure 4-2. Heatmaps displaying (A) 1000 most variable genes (B) 500 most variable genes (C) 250 most variable genes (D) 100 most variable genes. None showed a clear mosaic pattern that would be expected if distinct clusters were present in the dataset. Yellow represents a higher value (indicating genes are upregulated in that sample) and dark blue represents a lower value (indicating that genes are downregulated in that sample). The name of the clade is in square brackets after the sample name.

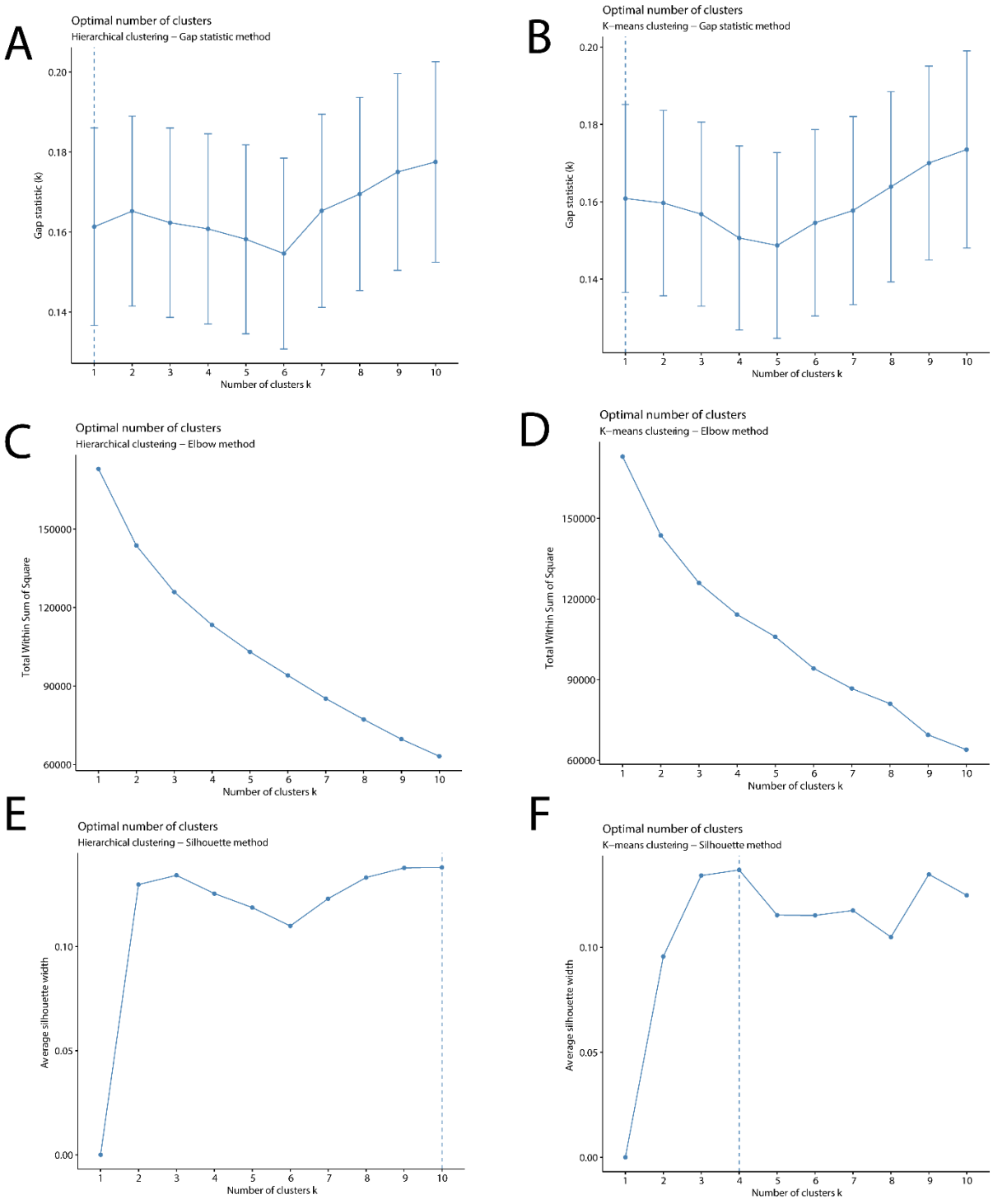


Figure 4-3. Clustering validation indices. The Gap statistic method indicated the optimum number of clusters was one for both (A) hierarchical clustering and (B) k-means clustering; the Elbow method was ambiguous for both (C) hierarchical clustering and (D) k-means clustering; and the Silhouette method indicated the optimum number of clusters was 10 (the maximum) for (E) hierarchical clustering and 4 for (F) k-means clustering.

The dotted line represents the highest value i.e. what the test deems to be the optimum number of clusters in the dataset.

Bootstrap analysis and Monte Carlo simulation of the hierarchical clustering proved similarly contradictory, with neither method completely supporting each other or any of the validation indices. The dendrogram generated from bootstrapping analysis did not yield high values, with 11 out of 27 samples unable to be assigned to any clade with AU > 95 (a common cutoff value for determining significance) (**Figure A3-8**). The AU values are approximately unbiased probability values, which are obtained by multiscale bootstrap resampling and are more reliable than ordinary bootstrap resampling. Of the seven clusters that did have AU > 95%, six consisted of only two samples. The Monte Carlo based approach using sigclust2 identified four different significant nodes across the dendrogram (**Figure A3-9**).

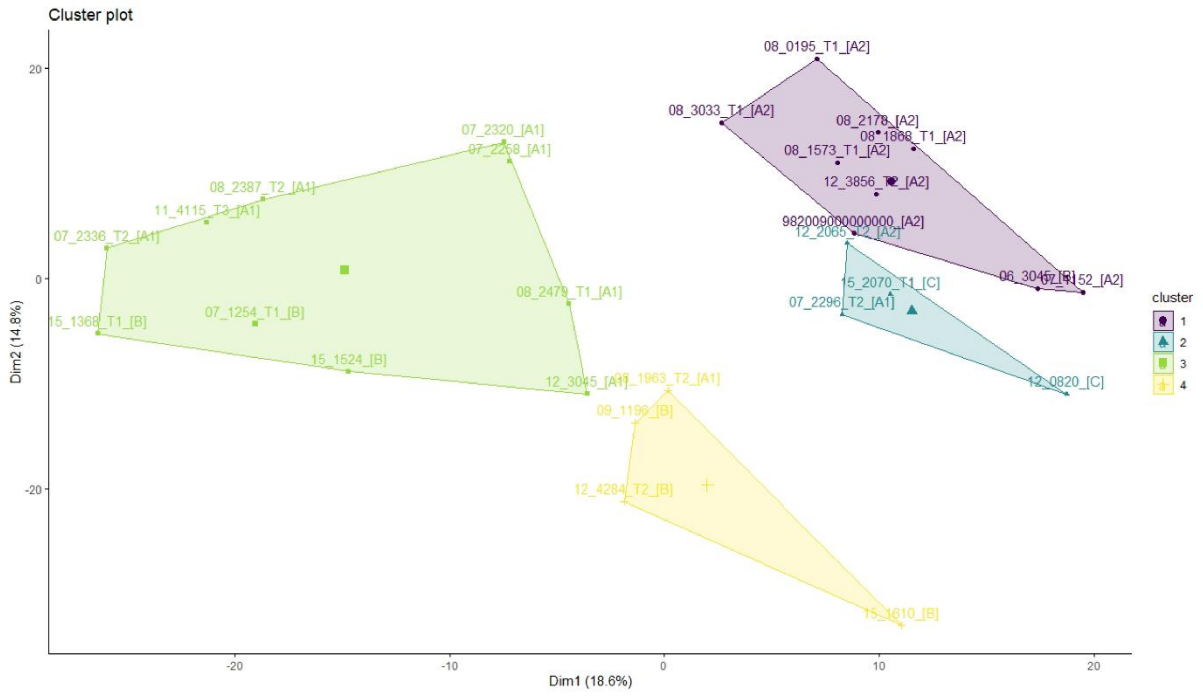


Figure 4-4. K means clustering results with k=4. Each cluster contained samples from at least two different genotypic clades.

The results of the clustering analysis on the full dataset (when not filtered for purity) are reported in **Figure A3-10 – Figure A3-13**.

Discussion

In this study we demonstrate that there are no significant differences in gene expression patterns between Devil Facial Tumours from different genotypic subgroups. The lack of consensus between clustering methods, and the lack of natural groups in the datasets using the visualisation techniques (MDS, t-SNE, heatmaps) indicated a single cluster for our dataset. This was further supported by the only validation index that constitutes a formal statistical test (the Gap method) (Patel et al., 2022). These results were consistent when applied to both the dataset filtered by >80% purity (**Figure 4-1**, **Figure 4-2**, **Figure 4-3**, **Figure 4-4**) and the full dataset (**Figure A3-8 – Figure A3-13**). Collectively, this shows that despite the genetic (Kwon et al., 2020) and karyotypic (Pearse et al., 2012) variation that exists between individuals, gene expression remains largely consistent across a broad temporal (2006 to 2015) and geographic range (~21 120 km^2).

Vaccination has been successful in managing other wildlife diseases, such as rabies (MacInnes et al., 2001; Slate et al., 2009), but never for cancers. Cancer is highly heterogenous, often consisting of multiple molecular subtypes. A common method of defining these subtypes of cancer is through transcriptional profiling (S. Gao et al., 2017; Kunz et al., 2018; Laurell et al., 2009; Lehmann et al., 2011; D. Wang et al., 2022) – identifying groups with similar patterns of gene expression to guide treatment. Our results indicate that such transcriptomic subtypes are not present in DFT1 and so a single vaccine does have the potential to be effective for a large cross section of tumours.

Typically, we would expect to see variation among tumour biopsies in such a clustering analysis. This is because most tumours arise in different individuals and have a different underlying genetic profile. DFT1 is different, being a clonal cell line that has only arisen once in a single female Tasmanian devil (Murchison et al., 2010). As a transmissible cancer,

we were not sure what we would see as this single cell has passed from animal to animal and undergone many cell divisions since it evolved at least 30 years ago. As such, we were surprised to see such a stable phenotype. However, this does appear to be consistent with previous studies that have implied DFTD has an unusual genetic stability for a cancer.

In addition to variation between individuals, non-contagious cancers also exhibit a high level of heterogeneity within individuals. Many tumours are composed of subclonal populations that may be considered analogous to the clades of DFT1 as both have arisen from a single cell (L. G. Morris et al., 2016). Unlike DFT1, these subclones often exhibit phenotypic diversity. High levels of heterogeneity are associated with poor prognosis due to the selective pressure of chemotherapy and other treatments (R. Gao et al., 2017; Greaves & Maley, 2012).

However, DFT1 is also unusual as it is not treated with chemotherapy and devils rarely mount an immune response against the disease (Pye, Hamede, et al., 2016). DFT1 cells do not express MHC-I molecules so are not recognised by the host immune system (Siddle et al., 2013). This susceptibility to disease is likely exacerbated by low immune gene diversity in the Tasmanian devil (Morris et al., 2013). We suggest that this lack of immune response weakens the co-evolutionary arms race between host and pathogen, resulting in a more stable phenotype.

DFT1 also occupies an extremely narrow ecological niche for a contagious disease, being one of the few pathogens that is simultaneously an infectious agent and a mammalian cell (Metzger & Goff, 2016). If deviations beyond a particular phenotype result in unviable or non-transmissible cells (Pearse et al., 2012), there is likely a strong selective pressure to evolve slowly. Such a conserved phenotype has also been observed in the only other contagious cancer to afflict vertebrates (Canine Transmissible Venereal Tumour, CTVT) with histopathological screening showing no significant histopathological differences between clades (Strakova, 2017). Although transcriptomics has been used in CTVT

(Frampton et al., 2018), this study focused on the differential response to treatment rather than natural phenotypic variation between samples.

One previous study has suggested that transcriptomic differences exist in DFT1 tumours from different geographic locations (Kozakiewicz et al., 2021). However, we also suggest that pair-wise comparisons between locations are not informative, as they assume each location only contains a single subtype, with no migration occurring between sites. Neutral population genetic data contradicts this however, showing movement across the landscape in Tasmania (Farquharson et al., 2022).

We acknowledge that our results should be interpreted in the context of the study's limitations, namely that methods for estimating tumour purity were restricted due to the nature of the data. However, this study and dataset both represent a substantial contribution to the field of DFTD research and improve on previous studies that did not estimate tumour purity (Kozakiewicz et al., 2021).

In conclusion, our results show that DFT1 does not consist of multiple transcriptomic subtypes and that a one-shot vaccine that will work across all clades should have potential to manage the disease. As one of the largest studies to use transcriptomics to investigate phenotypic variation within DFT1, it also provides novel insights into this unique group of diseases.

Conflict of Interest Statement

The authors have no conflicts to declare.

Acknowledgements

Funding for this project was provided by the ARC Centre of Excellence for Innovations in Peptide and Protein Science (CE200100012) and ARC Linkage Project (LP180100244).

Data Availability Statement

The raw sequencing reads generated and analysed during this study are available in the National Centre for Biotechnology Information (NCBI) short read archive under BioProject number PRJNA1067341.

Benefit-Sharing Statement

This study complies with benefit sharing under the Convention on Biological Diversity.

Author Contributions

CJH and KB conceptualised and sourced funding for the study. CP, EP, KCB, CJH and KB designed the study. SF either collected, or coordinated collection, of the samples and provided access to individual metadata. CP and KCB performed RNA extractions. CP performed analyses and wrote the main manuscript. All authors reviewed drafts of the manuscript.

CHAPTER FIVE:

GENERAL DISCUSSION

GENERAL DISCUSSION

5.1 SUMMARY OF RESULTS

Wildlife diseases are increasing and present a major concern for conservation, agriculture and public health (DeCandia et al., 2018; Leifels et al., 2022). This thesis shows how genetics and bioinformatics can be used to further understand and manage these diseases. Using a combination of genomic (Chapter Two) and transcriptomic (Chapter Four) resources that I generated in conjunction with public data (Chapter Three), I was able to address a range of questions about the genetic interplay between marsupials and their diseases. These results collectively have implications for conservation management (for example translocations) but also offer a workflow that can be applied to other species and diseases. My main findings were:

1. Identification of a novel branch of marsupial-specific Ras genes (MgRas) almost exclusively expressed in reproductive organs (ovary, testes, yolk sac), indicating they have a reproductive function (Chapter Two).
2. The Dasyuromorphia and Peramelemorphia lineages have undergone order specific expansions in the Ras genes that is linked to their reproductive life histories but is also likely making them more susceptible to cancers (Chapter Two).
3. Identification of intraspecific diversity at the nucleotide, amino acid and copy number level in koala antimicrobial peptides. This is the first intraspecific comparison of AMPs in a marsupial species. I showed that non-synonymous SNPs in the active peptide region are predicted to change antimicrobial function (Chapter Three).
4. Identification of duplications in the PhciCATH5 AMP, a cathelicidin with activity against chlamydia, that were more common in southern populations where clinical symptoms of chlamydia are less severe (Chapter Three).

5. That there are no distinct transcriptomic differences between different genetic variants of devil facial tumour disease, taking into account tumour purity and genotypic clade (Chapter Four).

5.2 IMPORTANCE OF THESIS FINDINGS AND FUTURE DIRECTIONS

5.2.1 Cancer and Reproduction

The field of comparative oncology has identified molecular mechanisms for cancer resistance in many species but has not yet explored any genetic basis for cancer susceptibility. I addressed this gap in Chapter Two by generating two marsupial reference genomes and using comparative genomics techniques to characterize the evolution of cancer related genes in marsupials. I showed a statistically significant expansion of cancer related genes in the dasyurid ancestor and identified a novel branch of Ras subfamily genes. This branch was present in all marsupials, with expansions evident in the Dasyuromorphia and Peremelemorphia lineages. As these genes are almost exclusively expressed in reproductive organs, I hypothesise that they have a reproductive function. Both the Dasyuromorphia and Peramelemorphia have unique reproductive biology (supernumerary young and chorioallantoic placentae, respectively). The gene expansion of these genes may assist in this reproductive strategy but also may become oncogenic when mutated, and making the dasyurids more predisposed to neoplasia.

The expression of MgRas genes in reproductive organs evokes a long-hypothesised link between cancer and reproduction. Antagonistic pleiotropy has been observed in many genes that simultaneously increase reproductive output and render an individual more susceptible to cancer. For example, the *Xmrk* allele in *Xiphophorous* fish that leads to increased size and colouration that improve reproductive success yet result in higher rates of melanoma (Fernandez & Bowser, 2010; Fernandez & Morris, 2008); mutations in the

BRCA1 and *BRCA2* genes predispose humans to breast and ovarian cancer yet also increase fertility (Easton et al., 1995; Smith et al., 2012).

Another major example of antagonistic pleiotropy includes genes involved in gametogenesis. There are many shared features between cancer cells and germ cells, including immortality, invasion and migration (Simpson et al., 2005). This is exemplified by cancer-testis antigens (CTAs), genes that are almost exclusively expressed in testes, ovaries, trophoblasts and tumour cells (Li et al., 2020). Due to the unique expression patterns of MgRas genes, it is possible some may also function as CTAs. Future studies should investigate this by investigating gene expression in embryonic and cancer biopsies – firstly, whether any of these genes are upregulated and secondly, whether this is due to mutations in either coding or promoter regions.

The parallels between placentation and neoplasia have been discussed since the early 20th century (Ross, 2015). Both processes involve the same mechanisms and molecular pathways (rapid cell proliferation, immune evasion, cell migration and angiogenesis) (Costanzo et al., 2018). It has been hypothesised that tumours may develop by co-opting these pathways (Costanzo et al., 2018), or that species with more invasive placentas (haemochorial) are at higher risk of lethal cancers (Dujon et al., 2023). Several studies have investigated the relationship between degree of placentation and cancer mortality across species and either found no relationship (Boddy, Abegglen, et al., 2020) or contradictory results (D'Souza & Wagner, 2014; Dujon et al., 2023). These studies, however, included only small numbers of marsupial species. Future studies should incorporate data from species exhibiting the four different types of placentation in marsupials (no allantochorion, temporary allantochorion, allantachorion without villi and allantochorion with villi and fusion) (Renfree & Shaw, 2021) when modelling the relationship between placentation and malignancy. This will help elucidate whether degree of placenta invasion increases cancer risk in marsupials. In addition, future studies should

compare gene expression between healthy and tumour biopsies in marsupials to identify drivers of carcinogenesis.

5.2.2 AMPs and Drug Development

I investigated immunogenetic variation on an intraspecific level using the koala (Chapter Three). Although immune gene diversity is known to influence individual susceptibility to disease, many studies on non-model organisms have only focused on one or two well-known gene families (e.g. MHC and TLR). Using 418 koala resequenced genomes, I was characterised intraspecific functional (AMP) diversity across the species' geographic range. I selected this gene family as cathelicidins and defensins have been characterized in many marsupial species, with the former showing many species-specific expansions and potent activity against diseases of conservation concern (e.g. chlamydia in koalas, DFTD in Tasmanian devils). Yet before my study it was unknown whether any level of intraspecific diversity existed. I showed that AMPs are more conserved than other immune genes, but diversity does exist at the nucleotide, amino acid and copy number level, including non-synonymous SNPs. Similar to neutral diversity, nucleotide diversity decreased along a geographic gradient. Although interestingly, CNVs were more common in the south. I hypothesise that the duplications in *PhciCATH5* result in higher resistance to chlamydia, as this cathelicidin has displayed activity against the bacteria *in vitro*.

Future studies should further investigate the activity of these different alleles (PhciCATH3_Hap1-3, PhciDEFA1_Hap1 - 4, PhciDEFB7_Hap1 - 3, PhciDEFB9_Hap1 - 4, PhciDEFB10_Hap1 - 2, PhciDEFB12_Hap1 - 5, PhciDEFB19_Hap1 - 4, PhciDEFB27_Hap1 - 2, and PhciDEFB28_Hap1 - 2) *in vitro*. AMPs have long been recognised as potent targets for drug development, due to their rapid activity against a broad range of microbes, anti-biofilm activity and the fact that they are unlikely to induce resistance (Luo & Song, 2021; Zasloff, 2002). Although many are in clinical and preclinical trials (Koo & Seo, 2019), there remain limitations such as low bioavailability and selectivity (de Oliveira et al., 2023). For this reason, rational design has been used to modify peptide

properties such as enhanced activity (Higgs et al., 2007; Molhoek et al., 2010), reduction in toxicity and improved *in vivo* effectiveness (Han et al., 2021). More recently, AMPs have been optimised for therapeutic development through directed evolution (Liu et al., 2025; Zhang et al., 2025). The identification of different alleles (eg non-synonymous SNPs with predicted changes in function in PhciCATH3, PhciDEFA1, PhciDEFB7, PhciDEFB9, PhciDEFB10, PhciDEFB28) maximises the amount of template sequences upon which rational design can be performed for therapeutic development.

Another exciting avenue for future research is the characterisation of novel AMPs in marsupials. Traditionally, AMPs have been identified through peptidomic methods (performing high performance liquid chromatography on biological tissues or secretions) (Conlon & Sonnevend, 2009; Lauth et al., 2002) or homology based bioinformatic methods (Peel et al., 2016; Peel et al., 2021; Tang et al., 2024). However, these methods are limited in that they are reliant on obtaining available samples or the presence on the conserved amino acid motifs. More recently, machine learning methods have enabled the identification of novel AMPs in a range of different species (Fingerhut et al., 2020; G. Wang et al., 2022). I recommend future studies characterise and investigate what level of intraspecific variation exists in more novel peptides beyond cathelicidins and defensins.

5.2.3 Wildlife Disease Management

Although genomic data can provide valuable insights into genetic diversity and resolve phylogenetic relationships, it is important to note that genomic differences do not necessarily equate to functional differences. I explored this question in Chapter Four using another disease of major conservation concern, DFT1. DFT1 consists of five different phylogenetic clades (Clades A – E), but it was unknown whether these clades exhibited phenotypic differences. This led to concerns about performing wild to wild translocation due to the risk of spreading different disease strains.

I generated 35 DFT1 transcriptomes that were representative of the disease's four major phylogenetic clades and applied two types of unsupervised clustering to show that they all had similar gene expression profiles. This is one of the largest DFT1 transcriptome datasets generated to date, and the only study to assess both tumour purity and genotypic clade in analysing gene expression differences. I provide insight into contagious cancers that has implication for vaccine development, as these results show a single treatment approach may be effective against a large range of tumours.

Managing wildlife diseases presents many challenges and involves trade-offs, uncertainties and competing concerns from stakeholders (McEachran et al., 2024). Strategies include culling, removing infected individuals, vaccination and establishment of disease-free insurance populations, although none of these are without their limitations. In the case of DFT1, culling was ineffective (Lachish et al., 2010) and the removal of infected animals inadvertently selected for slower growing strains of the disease animals (Ujvari et al., 2014).

Although not a direct method of managing disease, translocation of devils between populations is a powerful conservation strategy to increase genetic diversity and adaptive potential of populations (Frankham, 2015). My results from Chapter Five indicate that all DFT1 strains have largely similar gene expression profiles, mitigating concerns that translocating devils between different genotypic clades may lead to more virulent strains of the disease.

Currently the most promising strategy for managing DFT1 is through the development of a vaccine. My results support the development of a single vaccine to treat all strains of the disease and also provides a valuable dataset that can be used for future research and vaccine development. For example, identification and assessment of neoantigen expression compared to wild type transcripts. Neoantigens are unique antigens that are expressed by

tumours but not healthy cells, and so an ideal target for vaccines. They can be identified using a differential expression analysis between DFTD and healthy cells.

5.3 LIMITATIONS

One major limitation in this study was the lack of koala metadata – for example, disease status, whether individuals were symptomatic or not, and whether they were able to resolve infection. This information would have allowed analyses such as genome-wide association studies to identify potential correlations between SNPs in AMPs and disease susceptibility.

Another limitation was the inability to assess tumour purity of DFTD biopsies prior to sequencing, which resulted in a smaller sample size (35 biopsies were sequenced but only 27 were retained in the final analysis). Further, the study was reliant on bulk RNA sequencing due to the logistics of collecting samples from wild animals. Ideally, single cell RNA sequencing would have provided higher resolution transcriptomic analysis, allowing a more accurate characterisation of the gene expression profiles of the different clades.

In Chapter Two, limitations included the inability to measure cancer prevalence in species (eg as a percentage) and instead only count incidents. Another limitation was the restricted number of transcriptomes for some species (for example, only two for the kowari and bandicoot).

5.4 CONCLUSIONS

The work presented in this thesis has made a significant contribution to our understanding of wildlife diseases, which are increasingly being recognised as a major concern worldwide. I have developed genomic and transcriptomic resources for marsupials providing further insights into their immunogenetics and disease. I documented previously unknown diversity at both inter- and intra-specific level in lesser studied groups of marsupial genes,

such as AMPs and oncogenes, and provided insight into gene expression variability in a contagious cancer. My thesis has identified the following new questions:

- What is the role of MgRas genes in both reproduction and cancer?
- What are the functional effects of AMP diversity in koalas?
- Why does DFTD exhibit such a conserved phenotype?

Collectively, this lays the groundwork for future studies into marsupial immunogenetics and inform conservation management decisions. The methods employed here may also be used for other threatened wildlife species, particularly those suffering from disease.

REFERENCES

AbdelRahman, Y. M., & Belland, R. J. (2005). The chlamydial developmental cycle. *FEMS microbiology reviews*, *29*(5), 949-959.

Abegglen, L. M., Caulin, A. F., Chan, A., Lee, K., Robinson, R., Campbell, M. S., Kiso, W. K., Schmitt, D. L., Waddell, P. J., & Bhaskara, S. (2015). Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. *Jama*, *314*(17), 1850-1860. https://doi.org/DOI:10.3390/genes14030546

Abercrombie, M. (1979). Contact inhibition and malignancy. *Nature*, *281*(5729), 259-262.

Adolfsson, A., Ackerman, M., & Brownstein, N. C. (2019). To cluster, or not to cluster: An analysis of clusterability methods. *Pattern Recognition*, *88*, 13-26.

Agerberth, B., Gunne, H., Odeberg, J., Kogner, P., Boman, H. G., & Gudmundsson, G. H. (1995). FALL-39, a putative human peptide antibiotic, is cysteine-free and expressed in bone marrow and testis. *Proceedings of the National Academy of Sciences*, *92*(1), 195-199.

Ahn, B., Jeon, H., Cho, H.-s., Nagasundarapandian, S., & Park, C. (2022). Sequence polymorphisms of PR39 cathelicidins and extensive copy variations in commercial pig breeds. *Gene*, *822*, 146323. https://doi.org/https://doi.org/10.1016/j.gene.2022.146323

Aiello, C. M., Nussear, K. E., Walde, A. D., Esque, T. C., Emblidge, P. G., Sah, P., Bansal, S., & Hudson, P. J. (2014). Disease dynamics during wildlife translocations: disruptions to the host population and potential consequences for transmission in desert tortoise contact networks. *Animal Conservation*, *17*, 27-39.

Aktipis, C. A., Boddy, A. M., Jansen, G., Hibner, U., Hochberg, M. E., Maley, C. C., & Wilkinson, G. S. (2015). Cancer across the tree of life: cooperation and cheating in multicellularity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1673), 20140219.

Albuquerque, T. A., Drummond do Val, L., Doherty, A., & de Magalhaes, J. P. (2018). From humans to hydra: patterns of cancer across the tree of life. *Biological Reviews*, *93*(3), 1715-1734.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.

Anderson, M. W., Reynolds, S. H., You, M., & Maronpot, R. M. (1992). Role of proto-oncogene activation in carcinogenesis. *Environmental health perspectives*, *98*, 13-24.

Anderson, W. I., Peters, D. N., Stoffregen, D. A., Steinberg, H., & Wallace, C. (1990). Glossal squamous cell carcinoma with pulmonary metastasis in a kowari (Dasyuroides byrnei). *Australian veterinary journal*, *67*(1), 29. https://doi.org/10.1111/j.1751-0813.1990.tb07389.x

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. In: Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Antonio, L.-D.-F., & Widegren, B. (2005). The non-random location of human oncogenes and tumour suppressor genes. *Caryologia*, *58*(1), 1-14.

Arasu, A., Murugan, S., Essa, M. M., Velusamy, T., & Guillemin, G. J. (2018). PAX3: A molecule with oncogenic or tumor suppressor function is involved in cancer. *BioMed Research International*, *2018*(1), 1095459.

Arnal, A., Tissot, T., Ujvari, B., Nunney, L., Solary, E., Laplane, L., Bonhomme, F., Vittecoq, M., Tasiemski, A., & Renaud, F. (2016). The guardians of inherited oncogenic vulnerabilities. *Evolution*, *70*(1), 1-6.

Ashman, R., Keast, D., Stanley, N., & Waring, H. (1975). The immunological responses of marsupials. *American Zoologist*, *15*(1), 155-166.

Attwood, H., & Woolley, P. (1973). Spontaneous malignant neoplasms in dasyurid marsupials. *Journal of Comparative Pathology*, *83*(4), 569-581.

Austad, S. N. (1997). Comparative aging and life histories in mammals. *Experimental gerontology*, *32*(1-2), 23-38.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., & Schaeffer, R. D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, *373*(6557), 871-876.

Bagheri, M., & Zahmatkesh, A. (2018). Evolution and species-specific conservation of toll-like receptors in terrestrial vertebrates. *International Reviews of Immunology*, *37*(5), 217-228. https://doi.org/https://doi.org/10.1080/08830185.2018.1506780

Bakar, S. A., Hollox, E. J., & Armour, J. A. (2009). Allelic recombination between distinct genomic locations generates copy number diversity in human β-defensins. *Proceedings of the National Academy of Sciences*, *106*(3), 853-858. https://doi.org/https://doi.org/10.1073/pnas.0809073106

Baker, A. M., Eldridge, M. D., Fisher, D. O., Frankham, G., Helgen, K., Jackson, S. M., Potter, S., Travouillon, K. J., & Umbrello, L. S. (2023). Taxonomy and diversity of living Australasian marsupials. In *American and Australasian Marsupials: an evolutionary, biogeographical, and ecological approach* (pp. 1-85). Springer.

Bals, R., & Wilson, J. (2003). Cathelicidins-a family of multifunctional antimicrobial peptides. *Cellular and Molecular Life Sciences CMLS*, *60*, 711-720.

Barba, I., Carrillo-Bosch, L., & Seoane, J. (2024). Targeting the Warburg Effect in Cancer: Where Do We Stand? *International Journal of Molecular Sciences*, *25*(6), 3142.

Barreiro, L. B., & Quintana-Murci, L. (2010). From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nature reviews genetics*, *11*(1), 17-30.

Basden, K., Cooper, D., & Deane, E. (1997). Development of the lymphoid tissues of the tammar wallaby Macropus eugenii. *Reproduction, Fertility and Development*, *9*(2), 243-254.

Batley, K. C., Sandoval-Castillo, J., Kemper, C. M., Attard, C. R., Zanardo, N., Tomo, I., Beheregaray, L. B., & Möller, L. M. (2019). Genome-wide association study of an unusual dolphin mortality event reveals candidate genes for susceptibility and resistance to cetacean morbillivirus. *Evolutionary applications*, *12*(4), 718-732.

Beck, R. M. (2023). Diversity and phylogeny of marsupials and their stem relatives (Metatheria). In *American and Australasian Marsupials: An Evolutionary, Biogeographical, and Ecological Approach* (pp. 1-66). Springer.

Begley, U., Sosa, M. S., Avivar-Valderas, A., Patil, A., Endres, L., Estrada, Y., Chan, C. T., Su, D., Dedon, P. C., & Aguirre-Ghiso, J. A. (2013). A human tRNA methyltransferase 9-like protein prevents tumour growth by regulating LIN9 and HIF1-α. *EMBO molecular medicine*, *5*(3), 366-383.

Belov, K., Harrison, G. A., & Cooper, D. W. (1998). Molecular cloning of the cDNA encoding the constant region of the immunoglobulin A heavy chain (Cα) from a marsupial: Trichosurus vulpecula (common brushtail possum). *Immunology Letters*, *60*(2-3), 165-170.

Belov, K., Harrison, G. A., Miller, R. D., & Cooper, D. W. (1999a). Isolation and sequence of a cDNA coding for the heavy chain constant region of IgG from the Australian brushtail possum, Trichosurus vulpecula. *Molecular immunology*, *36*(8), 535-541.

Belov, K., Harrison, G. A., Miller, R. D., & Cooper, D. W. (1999b). Molecular cloning of the brushtail possum (Trichosurus vulpecula) immunglobulin E heavy chain constant region. *Molecular immunology*, *36*(18), 1255-1261.

Belov, K., Harrison, G. A., Miller, R. D., & Cooper, D. W. (2001). Characterisation of the κ light chain of the brushtail possum (Trichosurus vulpecula). *Veterinary immunology and immunopathology*, *78*(3-4), 317-324.

Belov, K., Harrison, G. A., Rosenberg, G. H., Miller, R. D., & Cooper, D. W. (1999). Isolation and comparison of the IgM heavy chain constant regions from Australian (Trichosurus vulpecula) and American (Monodelphis domestica) marsupials. *Developmental & Comparative Immunology*, *23*(7-8), 649-656.

Belov, K., Lam, M.-P., & Colgan, D. (2004). Marsupial MHC class II β genes are not orthologous to the eutherian β gene families. *Journal of Heredity*, *95*(4), 338-345.

Belov, K., Miller, R. D., Old, J. M., & Young, L. J. (2013). Marsupial immunology bounding ahead. *Australian Journal of Zoology*, *61*(1), 24-40.

Belov, K., Sanderson, C. E., Deakin, J. E., Wong, E. S., Assange, D., McColl, K. A., Gout, A., de Bono, B., Barrow, A. D., & Speed, T. P. (2007). Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system. *Genome research*, *17*(7), 982-991.

Belton, J.-M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., & Dekker, J. (2012). Hi–C: a comprehensive technique to capture the conformation of genomes. *Methods*, *58*(3), 268-276.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research*, *27*(2), 573-580.

Bera, A., Singh, S., Nagaraj, R., & Vaidya, T. (2003). Induction of autophagic cell death in Leishmania donovani by antimicrobial peptides. *Molecular and biochemical parasitology*, *127*(1), 23-35. https://doi.org/https://doi.org/10.1016/S0166-6851(02)00300-6

Berger, A., Imielinski, M., Duke, F., Wala, J., Kaplan, N., Shi, G., Andres, D., & Meyerson, M. (2014). Oncogenic RIT1 mutations in lung adenocarcinoma. *Oncogene*, *33*(35), 4418-4423.

Berman, A. (2011). Invited review: Are adaptations present to support dairy cattle productivity in warm climates? *Journal of dairy science*, *94*(5), 2147-2158. https://doi.org/https://doi.org/10.3168/jds.2010-3962

Bi, S., Zheng, X., Wang, X., Cignetti, N. E., Yang, S., & Wible, J. R. (2018). An Early Cretaceous eutherian and the placental–marsupial dichotomy. *Nature*, *558*(7710), 390-395.

Bickhart, D. M., Hou, Y., Schroeder, S. G., Alkan, C., Cardone, M. F., Matukumalli, L. K., Song, J., Schnabel, R. D., Ventura, M., & Taylor, J. F. (2012). Copy number variation of individual cattle genomes using next-generation sequencing. *Genome research*, *22*(4), 778-790. https://doi.org/http://www.genome.org/cgi/doi/10.1101/gr.133967.111

Bildik, G., Liang, X., Sutton, M. N., Bast Jr, R. C., & Lu, Z. (2022). DIRAS3: an imprinted tumor suppressor gene that regulates RAS and PI3K-driven cancer growth, motility, autophagy, and tumor dormancy. *Molecular cancer therapeutics*, *21*(1), 25-37.

Black, L. A., Landersdorfer, C. B., Bulitta, J. B., Griffith, J. E., & Govendir, M. (2014). Evaluation of enrofloxacin use in koalas (Phascolarctos cinereus) via population pharmacokinetics and M onte C arlo simulation. *Journal of veterinary pharmacology and therapeutics*, *37*(3), 301-311.

Blanchong, J. A., Robinson, S. J., Samuel, M. D., & Foster, J. T. (2016). Application of genetics and genomics to wildlife epidemiology. *The Journal of Wildlife Management*, *80*(4), 593-608.

Blanshard, W., & Bodley, K. (2008). 8. Koalas. In *Medicine of Australian mammals* (pp. 227-329). BioOne.

Boddy, A. M., Abegglen, L. M., Pessier, A. P., Aktipis, A., Schiffman, J. D., Maley, C. C., & Witte, C. (2020). Lifetime cancer prevalence and life history traits in mammals. *Evolution, medicine, and public health*, *2020*(1), 187-195.

Boddy, A. M., Harrison, T. M., & Abegglen, L. M. (2020). Comparative oncology: New insights into an ancient disease. *Iscience*, *23*(8), 101373.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120.

Booth, R., & Nyari, S. (2020). Clinical comparison of five anti-chlamydial antibiotics in koalas (Phascolarctos cinereus). *PloS one*, *15*(7), e0236758.

Borthwick, C. R., Young, L. J., & Old, J. M. (2014). The development of the immune tissues in marsupial pouch young. *Journal of Morphology*, *275*(7), 822-839.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology*, *10*(4), e1003537.

Boudjadi, S., Chatterjee, B., Sun, W., Vemu, P., & Barr, F. G. (2018). The expression and function of PAX3 in development and disease. *Gene*, *666*, 145-157.

Brahma, B., Patra, M. C., Karri, S., Chopra, M., Mishra, P., De, B. C., Kumar, S., Mahanty, S., Thakur, K., & Poluri, K. M. (2015). Diversity, antimicrobial action and structure-activity relationship of buffalo cathelicidins. *PloS one*, *10*(12), e0144741. https://doi.org/DOI:10.1371/journal.pone.0144741

Brandies, P. A., Tang, S., Johnson, R. S., Hogg, C. J., & Belov, K. (2020). The first Antechinus reference genome provides a resource for investigating the genetic basis of semelparity and age-related neuropathologies. *Gigabyte*, *2020*.

Briscoe, N. J., Krockenberger, A., Handasyde, K. A., & Kearney, M. R. (2015). Bergmann meets Scholander: geographical variation in body size and insulation in the koala is related to climate. *Journal of Biogeography*, *42*(4), 791-802. https://doi.org/https://doi.org/10.1111/jbi.12445

Brouwer, L., Barr, I., Van De POL, M., Burke, T., Komdeur, J., & Richardson, D. S. (2010). MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. *Molecular Ecology*, *19*(16), 3444-3455. https://doi.org/https://doi.org/10.1111/j.1365-294X.2010.04750.x

Brown, O. J. (2006). Tasmanian devil (Sarcophilus harrisii) extinction on the Australian mainland in the mid-Holocene: multicausality and ENSO intensification. *Alcheringa: An Australasian Journal of Palaeontology*, *30*(S1), 49-57.

Brüniche-Olsen, A., Jones, M. E., Austin, J. J., Burridge, C. P., & Holland, B. R. (2014). Extensive population decline in the Tasmanian devil predates European settlement and devil facial tumour disease. *Biology letters*, *10*(11), 20140619.

Bruzos, A. L., Santamarina, M., Garcia-Souto, D., Díaz, S., Rocha, S., Zamora, J., Lee, Y., Viña-Feás, A., Quail, M. A., & Otero, I. (2023). Somatic evolution of marine transmissible leukemias in the common cockle, Cerastoderma edule. *Nature Cancer*, *4*(11), 1575-1591.

Bulls, S. E., Platner, L., Ayub, W., Moreno, N., Arditi, J.-P., Dreyer, S., McCain, S., Wagner, P., Burgstaller, S., & Davis, L. R. (2022). Cancer prevalence is related to body mass and lifespan in tetrapods and remarkably low in turtles. *bioRxiv*, 2022.2007. 2012.499088.

Burach, F., Pospischil, A., Hanger, J., Loader, J., Pillonel, T., Greub, G., & Borel, N. (2014). Chlamydiaceae and Chlamydia-like organisms in the koala (Phascolarctos cinereus)—organ distribution and histopathological findings. *Veterinary microbiology*, *172*(1-2), 230-240.

Burnard, D., Gillett, A., & Polkinghorne, A. (2018). Chlamydia pecorum in joint tissue and synovial fluid of a koala (Phascolarctos cinereus) with arthritis. *Journal of wildlife diseases*, *54*(3), 646-649.

Cáceres, N. C., & Dickman, C. R. (2023). American and Australasian Marsupials: An Introduction. In *American and Australasian Marsupials: An Evolutionary, Biogeographical, and Ecological Approach* (pp. 1-18). Springer.

Cagan, A., Baez-Ortega, A., Brzozowska, N., Abascal, F., Coorens, T. H., Sanders, M. A., Lawson, A. R., Harvey, L. M., Bhosle, S., & Jones, D. (2022). Somatic mutation rates scale with lifespan across mammals. *Nature*, *604*(7906), 517-524.

Caldwell, A., Coleby, R., Tovar, C., Stammnitz, M. R., Kwon, Y. M., Owen, R. S., Tringides, M., Murchison, E. P., Skjødt, K., & Thomas, G. J. (2018). The newly-arisen Devil facial tumour disease 2 (DFT2) reveals a mechanism for the emergence of a contagious cancer. *Elife*, *7*, e35314.

Candille, S. I., Kaelin, C. B., Cattanach, B. M., Yu, B., Thompson, D. A., Nix, M. A., Kerns, J. A., Schmutz, S. M., Millhauser, G. L., & Barsh, G. S. (2007). A β-defensin mutation causes black coat color in domestic dogs. *Science*, *318*(5855), 1418-1423.

Canfield, P., Hartley, W., & Reddacliff, G. (1990a). Spontaneous proliferations in Australian marsupials—a survey and review. 1. Macropods, koalas, wombats, possums and gliders. *Journal of Comparative Pathology*, *103*(2), 135-146.

Canfield, P., Hartley, W., & Reddacliff, G. (1990b). Spontaneous proliferations in Australian marsupials—a survey and review. 2. Dasyurids and bandicoots. *Journal of Comparative Pathology*, *103*(2), 147-158.

Carey, A. J., Timms, P., Rawlinson, G., Brumm, J., Nilsson, K., Harris, J. M., & Beagley, K. W. (2010). A Multi-Subunit Chlamydial Vaccine Induces Antibody and Cell-Mediated Immunity in Immunized Koalas (Phascolarctos cinereus): Comparison of Three Different Adjuvants. *American journal of reproductive immunology*, *63*(2), 161-172.

Caulin, A. F., Graham, T. A., Wang, L.-S., & Maley, C. C. (2015). Solutions to Peto's paradox revealed by mathematical modelling and cross-species cancer gene analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1673), 20140222.

Caulin, A. F., & Maley, C. C. (2011). Peto's Paradox: evolution's prescription for cancer prevention. *Trends in ecology & evolution*, *26*(4), 175-182.

Challis, R., Kumar, S., Sotero-Caio, C., Brown, M., & Blaxter, M. (2023). Genomes on a Tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life. *Wellcome open research*, *8*.

Chan, R. J., & Feng, G.-S. (2007). PTPN11 is the first identified proto-oncogene that encodes a tyrosine phosphatase. *Blood*, *109*(3), 862-867.

Chapman, J. R., Hellgren, O., Helin, A. S., Kraus, R. H., Cromie, R. L., & Waldenström, J. (2016). The evolution of innate immune genes: purifying and balancing selection on β-defensins in waterfowl. *Molecular Biology and Evolution*,

33(12), 3075-3087. https://doi.org/https://doi.org/10.1093/molbev/msw167

Chapman, J. R., Hill, T., & Unckless, R. L. (2019). Balancing selection drives the maintenance of genetic variation in Drosophila antimicrobial peptides. *Genome biology and evolution*, *11*(9), 2691-2701.

Charlesworth, B. (2000). Fisher, Medawar, Hamilton and the evolution of aging. *Genetics*, *156*(3), 927-931.

Chaudhuri, S., Thomas, S., & Munster, P. (2021). Immunotherapy in breast cancer: A clinician's perspective. *Journal of the National Cancer Center*, *1*(2), 47-57.

Chen, Y., Cao, L., Zhong, M., Zhang, Y., Han, C., Li, Q., Yang, J., Zhou, D., Shi, W., & He, B. (2012). Anti-HIV-1 activity of a new scorpion venom peptide derivative Kn2-7. *PloS one*, *7*(4), e34947. https://doi.org/DOI:10.1371/journal.pone.0034947

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature methods*, *18*(2), 170-175.

Cheng, Y., Grueber, C., Hogg, C. J., & Belov, K. (2022). Improved high-throughput MHC typing for non-model species using long-read sequencing. *Molecular Ecology Resources*, *22*(3), 862-876.

Cheng, Y., Prickett, M. D., Gutowska, W., Kuo, R., Belov, K., & Burt, D. W. (2015). Evolution of the avian β -defensin and cathelicidin genes. *BMC evolutionary biology*, *15*, 1-17. https://doi.org/DOI: 10.1186/s12862-015-0465-3,PMID: 26373713

Cheng, Y., Sanderson, C., Jones, M., & Belov, K. (2012). Low MHC class II diversity in the Tasmanian devil (Sarcophilus harrisii). *Immunogenetics*, *64*(7), 525-533.

Chesnokova, L. S., Slepenkov, S. V., & Witt, S. N. (2004). The insect antimicrobial peptide, L-pyrrhocoricin, binds to and stimulates the ATPase activity of both wild-type and lidless DnaK. *FEBS letters*, *565*(1-3), 65-69.

Cho, H.-s., Yum, J., Larivière, A., Lévêque, N., Le, Q. V. C., Ahn, B., Jeon, H., Hong, K., Soundrarajan, N., & Kim, J.-H. (2020). Opossum Cathelicidins Exhibit Antimicrobial Activity Against a Broad Spectrum of Pathogens Including West Nile Virus. *Frontiers in Immunology*, *11*, 347. https://doi.org/https://doi.org/10.3389/fimmu.2020.00347

Clark, A. G., & Wang, L. (1997). Molecular population genetics of Drosophila immune system genes. *Genetics*, *147*(2), 713-724.

Colicelli, J. (2004). Human RAS superfamily proteins and related GTPases. *Science's STKE*, *2004*(250), re13-re13.

Collisson, E. A., Bailey, P., Chang, D. K., & Biankin, A. V. (2019). Molecular subtypes of pancreatic cancer. *Nature reviews Gastroenterology & hepatology*, *16*(4), 207-220.

Compton, Z. T., Mellon, W., Harris, V. K., Rupp, S., Mallo, D., Kapsetaki, S. E., Wilmot, M., Kennington, R., Noble, K., & Baciu, C. (2025). Cancer prevalence across vertebrates. *Cancer discovery*, *15*(1), 227-244.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., & Zhang, X. (2016). A survey of best practices for RNA-seq data analysis. *Genome biology*, *17*, 1-19.

Conlon, J. M., & Sonnevend, A. (2009). Antimicrobial peptides in frog skin secretions. In *Antimicrobial peptides: methods and protocols* (pp. 3-14). Springer.

Consortium, U. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic acids research*, *47*(D1), D506-D515.

Costanzo, V., Bardelli, A., Siena, S., & Abrignani, S. (2018). Exploring the links between cancer and placenta development. *Open biology*, *8*(6), 180081.

Cui, J., Batley, K. C., Silver, L. W., McLennan, E. A., Hogg, C. J., & Belov, K. (2025). Spatial variation in toll-like receptor diversity in koala populations across their geographic distribution. *Immunogenetics*, *77*(1), 5.

Cunningham, A. A. (1996). Disease risks of wildlife translocations. *Conservation Biology*, *10*(2), 349-353.

Cutts, J. H., & Krause, W. J. (1982). Postnatal development of the spleen in Didelphis virginiana. *Journal of Anatomy*, *135*(Pt 3), 601.

D'Souza, A. W., & Wagner, G. P. (2014). Malignant cancer and invasive placentation: a case for positive pleiotropy between endometrial and malignancy phenotypes. *Evolution, medicine, and public health*, *2014*(1), 136-145.

da Silva, J. M., Giachetto, P. F., da Silva, L. O., Cintra, L. C., Paiva, S. R., Yamagishi, M. E. B., & Caetano, A. R. (2016). Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTLs affecting production traits. *BMC genomics*, *17*, 1-14. https://doi.org/DOI:10.1186/s12864-016-2752-9

Daher, K. A., Selsted, M. E., & Lehrer, R. I. (1986). Direct inactivation of viruses by human granulocyte defensins. *Journal of virology*, *60*(3), 1068-1074.

Dalmaijer, E. S., Nord, C. L., & Astle, D. E. (2022). Statistical power for cluster analysis. *BMC bioinformatics*, *23*(1), 1-28.

Damas, J., Corbo, M., Kim, J., Turner-Maier, J., Farré, M., Larkin, D. M., Ryder, O. A., Steiner, C., Houck, M. L., & Hall, S. (2022). Evolution of the ancestral mammalian karyotype and syntenic regions. *Proceedings of the National Academy of Sciences*, *119*(40), e2209139119.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008.

Danis, T., & Rokas, A. (2024). The evolution of gestation length in eutherian mammals. *Proceedings B*, *291*(2033), 20241412.

De Falco, F., Perillo, A., Del Piero, F., Del Prete, C., Zizzo, N., Marcus, I., & Roperto, S. (2022). ERAS is constitutively expressed in the tissues of adult horses and may be a key player in basal autophagy. *Frontiers in veterinary science*, *9*, 818294.

de Oliveira, K. B. S., Leite, M. L., Cunha, V. A., da Cunha, N. B., & Franco, O. L. (2023). Challenges and advances in antimicrobial peptide development. *Drug Discovery Today*, *28*(8), 103629.

De Smet, K., & Contreras, R. (2005). Human antimicrobial peptides: defensins, cathelicidins and histatins. *Biotechnology letters*, *27*, 1337-1347.

De Smith, A., Walters, R., Froguel, P., & Blakemore, A. (2009). Human genes involved in copy number variation: mechanisms of origin, functional effects and implications for disease. *Cytogenetic and Genome Research*, *123*(1-4), 17-26. https://doi.org/DOI: 10.1159/000184688

Deakin, J. E., & Belov, K. (2012). A comparative genomics approach to understanding transmissible cancer in Tasmanian devils. *Annual Review of Genomics and Human Genetics*, *13*, 207-222.

Deakin, J. E., Bender, H. S., Pearse, A.-M., Rens, W., O'Brien, P. C., Ferguson-Smith, M. A., Cheng, Y., Morris, K., Taylor, R., & Stuart, A. (2012). Genomic restructuring in the Tasmanian devil facial tumour: chromosome painting and gene mapping provide clues to evolution of a transmissible tumour. *PLoS genetics*, *8*(2), e1002483.

DeCandia, A. L., Dobson, A. P., & vonHoldt, B. M. (2018). Toward an integrative molecular approach to wildlife disease. *Conservation Biology*, *32*(4), 798-807.

Delport, W., Poon, A. F., Frost, S. D., & Kosakovsky Pond, S. L. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*, *26*(19), 2455-2457. https://doi.org/doi:10.1093/bioinformatics/btq429

Dempsey, S., Pye, R. J., Gilbert, A. T., Fountain-Jones, N. M., Moffat, J. M., Benson-Amram, S., Smyser, T. J., & Flies, A. S. (2022). Evaluation of oral baits and distribution methods for Tasmanian devils (Sarcophilus harrisii). *Wildlife Research*.

Dentro, S. C., Leshchiner, I., Haase, K., Tarabichi, M., Wintersinger, J., Deshwar, A. G., Yu, K., Rubanova, Y., Macintyre, G., & Demeulemeester, J. (2021). Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. *Cell*, *184*(8), 2239-2254. e2239.

DePinho, R. A. (2000). The age of cancer. *Nature*, *408*(6809), 248-254.

Dhople, V., Krukemeyer, A., & Ramamoorthy, A. (2006). The human beta-defensin-3, an antibacterial peptide with multiple biological functions. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1758*(9), 1499-1512.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21.

Domazet-Lošo, T., & Tautz, D. (2010). Phylostratigraphic tracking of cancer genes suggests a link to the emergence of multicellularity in metazoa. *BMC biology*, *8*(1), 1-10.

Doolan, D. L., Apte, S. H., & Proietti, C. (2014). Genome-based vaccine design: the promise for malaria and other infectious diseases. *International journal for parasitology*, *44*(12), 901-913.

Dorin, J. R., & Barratt, C. L. (2014). Importance of β-defensins in sperm function. *Molecular human reproduction*, *20*(9), 821-826.

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, *7*, 1-8.

Duchêne, D. A., Bragg, J. G., Duchêne, S., Neaves, L. E., Potter, S., Moritz, C., Johnson, R. N., Ho, S. Y., & Eldridge, M. D. (2018). Analysis of phylogenomic tree space resolves relationships among marsupial families. *Systematic Biology*, *67*(3), 400-412.

Dujon, A. M., Bramwell, G., Roche, B., Thomas, F., & Ujvari, B. (2021). Transmissible cancers in mammals and bivalves: How many examples are there? Predictions indicate widespread occurrence. *Bioessays*, *43*(3), 2000222.

Dujon, A. M., Gatenby, R. A., Bramwell, G., MacDonald, N., Dohrmann, E., Raven, N., Schultz, A., Hamede, R., Gérard, A.-L., & Giraudeau, M. (2020). Transmissible cancers in an evolutionary perspective. *Iscience*, *23*(7), 101269.

Dujon, A. M., Schofield, G., Bramwell, G., Raven, N., Hamede, R., Thomas, F., & Ujvari, B. (2020). Global meta-analysis of over 50 years of multidisciplinary and international collaborations on transmissible cancers. *Evolutionary applications*, *13*(7), 1745-1755.

Dujon, A. M., Vincze, O., Lemaitre, J.-F., Alix-Panabières, C., Pujol, P., Giraudeau, M., Ujvari, B., & Thomas, F. (2023). The effect of placentation type, litter size, lactation and gestation length on cancer risk in mammals. *Proceedings of the Royal Society B*, *290*(2001), 20230940.

Easton, D. F., Ford, D., & Bishop, D. T. (1995). Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *American journal of human genetics*, *56*(1), 265.

Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. In *Genome Informatics 2009: Genome Informatics Series Vol. 23* (pp. 205-211). World Scientific.

Edwards, M., Hinds, L. A., Deane, E., & Deakin, J. (2012). A review of complementary mechanisms which protect the developing marsupial pouch young. *Developmental &Comparative Immunology*, *37*(2), 213-220. https://doi.org/https://doi.org/10.1016/j.dci.2012.03.013

Efeyan, A., & Serrano, M. (2007). p53: guardian of the genome and policeman of the oncogenes. *Cell cycle*, *6*(9), 1006-1010.

Effron, M., Griner, L., & Benirschke, K. (1977). Nature and rate of neoplasia found in captive wild mammals, birds, and reptiles at necropsy. *Journal of the National Cancer Institute*, *59*(1), 185-198.

El-Sayed, A., & Kamel, M. (2020). Climatic changes and their role in emergence and re-emergence of diseases. *Environmental Science and Pollution Research*, *27*(18), 22336-22352.

Elbers, J. P., Brown, M. B., & Taylor, S. S. (2018). Identifying genome-wide immune gene variation underlying infectious disease in wildlife populations–a next

generation sequencing approach in the gopher tortoise. *BMC genomics*, *19*, 1-10. https://doi.org/DOI 10.1186/s12864-018-4452-0

Eldridge, M. D., Deakin, J. E., MacDonald, A. J., Byrne, M., Fitzgerald, A., Johnson, R. N., Moritz, C., Palmer, S., & Young, A. (2020). The Oz Mammals Genomics (OMG) initiative: developing genomic resources for mammal conservation at a continental scale. *Australian Zoologist*, *40*(3), 505-509.

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in ecology &evolution*, *29*(1), 51-63.

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology*, *20*, 1-14.

Environment, D. o. t. (2025). *Phascolarctos cinereus (combined populations of Qld, NSW and the ACT)* https://www.environment.gov.au/cgi-bin/sprat/public/publicspecies.pl?taxon_id=85104

Evdokimov, A., Kutuzov, M., Petruseva, I., Lukjanchikova, N., Kashina, E., Kolova, E., Zemerova, T., Romanenko, S., Perelman, P., & Prokopov, D. (2018). Naked mole rat cells display more efficient excision repair than mouse cells. *Aging (Albany NY)*, *10*(6), 1454.

Farquharson, K. A., McLennan, E. A., Cheng, Y., Alexander, L., Fox, S., Lee, A. V., Belov, K., & Hogg, C. J. (2022). Restoring faith in conservation action: Maintaining wild genetic diversity through the Tasmanian devil insurance program. *Iscience*, *25*(7).

Fernandez-Rojo, M. A., Deplazes, E., Pineda, S. S., Brust, A., Marth, T., Wilhelm, P., Martel, N., Ramm, G. A., Mancera, R. L., & Alewood, P. F. (2018). Gomesin peptides prevent proliferation and lead to the cell death of devil facial tumour disease cells. *Cell death discovery*, *4*(1), 19.

Fernandez, A. A., & Bowser, P. R. (2010). Selection for a dominant oncogene and large male size as a risk factor for melanoma in the Xiphophorus animal model. *Molecular Ecology*, *19*(15), 3114-3123.

Fernandez, A. A., & Morris, M. R. (2008). Mate choice for more melanin as a mechanism to maintain a functional oncogene. *Proceedings of the National Academy of Sciences*, *105*(36), 13503-13507.

Ferraiuolo, M., Verduci, L., Blandino, G., & Strano, S. (2017). Mutant p53 protein and the hippo transducers YAP and TAZ: a critical oncogenic node in human cancers. *International Journal of Molecular Sciences*, *18*(5), 961.

Fingerhut, L. C., Miller, D. J., Strugnell, J. M., Daly, N. L., & Cooke, I. R. (2020). ampir: an R package for fast genome-wide prediction of antimicrobial peptides. *Bioinformatics*, *36*(21), 5262-5263.

Flies, A. S., Flies, E. J., Fox, S., Gilbert, A., Johnson, S. R., Liu, G.-S., Lyons, A. B., Patchett, A. L., Pemberton, D., & Pye, R. J. (2020). An oral bait vaccination approach for the Tasmanian devil facial tumor diseases. *Expert review of vaccines*, *19*(1), 1-10.

Flot, J. F. (2010). SeqPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. *Molecular Ecology Resources*, *10*(1), 162-166. https://doi.org/https://doi.org/10.1111/j.1755-0998.2009.02732.x

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, *117*(17), 9451-9457.

Formenti, G., Abueg, L., Brajuka, A., Brajuka, N., Gallardo-Alba, C., Giani, A., Fedrigo, O., & Jarvis, E. D. (2022). Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics*, *38*(17), 4214-4216.

Frampton, D., Schwenzer, H., Marino, G., Butcher, L. M., Pollara, G., Kriston-Vizi, J., Venturini, C., Austin, R., De Castro, K. F., & Ketteler, R. (2018). Molecular signatures of regression of the canine transmissible venereal tumor. *Cancer Cell*, *33*(4), 620-633. e626.

Frankham, R. (2015). Genetic rescue of small inbred populations: Meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology*, *24*(11), 2610-2618.

Fuchs, R. P. (2002). How DNA lesions are turned into mutations within cells? *Oncogene*, *21*(58), 8957-8966.

Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology*, *26*(20), 5369-5406. https://doi.org/https://doi.org/10.1111/mec.14264

Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, *31*(22), 3718-3720.

Galli, L., Pereira, A., Márquez, A., & Mazzoni, R. (2006). Ranavirus detection by PCR in cultured tadpoles (Rana catesbeiana Shaw, 1802) from South America. *Aquaculture*, *257*(1-4), 78-82.

Gallo, R. L., Kim, K. J., Bernfield, M., Kozak, C. A., Zanetti, M., Merluzzi, L., & Gennaro, R. (1997). Identification of CRAMP, a cathelin-related antimicrobial peptide expressed in the embryonic and adult mouse. *Journal of Biological Chemistry*, *272*(20), 13088-13093.

Gallo, R. L., Murakami, M., Ohtake, T., & Zaiou, M. (2002). Biology and clinical relevance of naturally occurring antimicrobial peptides. *Journal of Allergy and Clinical Immunology*, *110*(6), 823-831. https://doi.org/https://doi.org/10.1067/mai.2002.129801

Ganz, T. (2003). Defensins: antimicrobial peptides of innate immunity. *Nature reviews immunology*, *3*(9), 710-720.

Ganz, T., & Lehrer, R. I. (1994). Defensins. *Current opinion in immunology*, *6*(4), 584-589.

Gao, R., Kim, C., Sei, E., Foukakis, T., Crosetto, N., Chan, L.-K., Srinivasan, M., Zhang, H., Meric-Bernstam, F., & Navin, N. (2017). Nanogrid single-nucleus

RNA sequencing reveals phenotypic diversity in breast cancer. *Nature communications*, *8*(1), 228.

Gao, S., Qiu, Z., Song, Y., Mo, C., Tan, W., Chen, Q., Liu, D., Chen, M., & Zhou, H. (2017). Unsupervised clustering reveals new prostate cancer subtypes. *Translational Cancer Research*, *6*(3), 561-572.

García-España, A., & Philips, M. R. (2023). Origin and evolution of RAS membrane targeting. *Oncogene*, *42*(21), 1741-1750.

García-Navas, V., Kear, B. P., & Westerman, M. (2020). The geography of speciation in dasyurid marsupials. *Journal of Biogeography*, *47*(9), 2042-2053.

Gemmell, R. T., Veitch, C., & Nelson, J. (2002). Birth in marsupials. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *131*(4), 621-630.

Giamarellos-Bourboulis, E. J., Platzer, M., Karagiannidis, I., Kanni, T., Nikolakis, G., Ulrich, J., Bellutti, M., Gollnick, H., Bauer, M., & Zouboulis, C. C. (2016). High copy numbers of β-defensin cluster on 8p23. 1, confer genetic susceptibility, and modulate the physical course of hidradenitis suppurativa/acne inversa. *Journal of Investigative Dermatology*, *136*(8), 1592-1598. https://doi.org/https://doi.org/10.1016/j.jid.2016.04.021

Gillenwaters, E. N., Seabury, C. M., Elliott, J. S., & Womack, J. E. (2009). Sequence analysis and polymorphism discovery in 4 members of the bovine cathelicidin gene family. *Journal of Heredity*, *100*(2), 241-245. https://doi.org/https://doi.org/10.1093/jhered/esn112

Gilroy, D., Van Oosterhout, C., Komdeur, J., & Richardson, D. S. (2016). Avian β-defensin variation in bottlenecked populations: the Seychelles warbler and other congeners. *Conservation Genetics*, *17*, 661-674. https://doi.org/DOI:10.1007/s10592-016-0813-x

Godbout, E. J., Madaline, T., Casadevall, A., Bearman, G., & Pirofski, L.-a. (2020). The damage response framework and infection prevention: From concept to bedside. *Infection Control & Hospital Epidemiology*, *41*(3), 337-341.

Goin, F. J. (2023). Cenozoic metatherian evolution in the Americas. In *American and Australasian marsupials: an evolutionary, biogeographical, and ecological approach* (pp. 249-267). Springer.

Goitre, L., Trapani, E., Trabalzini, L., & Retta, S. F. (2014). The Ras superfamily of small GTPases: the unlocked secrets. *Ras signaling: methods and protocols*, 1-18.

Govendir, M., Hanger, J., Loader, J., Kimble, B., Griffith, J., Black, L., Krockenberger, M., & Higgins, D. (2012). Plasma concentrations of chloramphenicol after subcutaneous administration to koalas (Phascolarctos cinereus) with chlamydiosis. *Journal of veterinary pharmacology and therapeutics*, *35*(2), 147-154.

Greaves, M., & Maley, C. C. (2012). Clonal evolution in cancer. *Nature*, *481*(7381), 306-313.

Griner, L. (1979). Neoplasms in Tasmanian devils (Sarcophilus harrisii). *Journal of the National Cancer Institute*, *62*(3), 589-595.

Grueber, C. E., Wallis, G. P., & Jamieson, I. G. (2014). Episodic positive selection in the evolution of avian toll-like receptor innate immunity genes. *PloS one*, *9*(3), e89632. https://doi.org/doi:10.1371/journal.pone.0089632

Gu, Y., Mohammad, I. S., & Liu, Z. (2020). Overview of the STAT-3 signaling pathway in cancer and the development of specific inhibitors. *Oncology letters*, *19*(4), 2585-2594.

Guiler, E. (1970). Observations on the Tasmanian devil, Sarcophilus harrisii (Marsupialia: Dasyuridae) II. Reproduction, breeding and growth of pouch young. *Australian Journal of Zoology*, *18*(1), 63-70.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & 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.

Guinney, J., Dienstmann, R., Wang, X., De Reynies, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., & Angelino, P. (2015). The consensus molecular subtypes of colorectal cancer. *Nature medicine*, *21*(11), 1350-1356.

Guthmiller, J. M., Vargas, K. G., Srikantha, R., Schomberg, L. L., Weistroffer, P. L., McCray Jr, P. B., & Tack, B. F. (2001). Susceptibilities of oral bacteria and yeast to mammalian cathelicidins. *Antimicrobial agents and chemotherapy*, *45*(11), 3216-3219.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., & Lieber, M. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, *8*(8), 1494-1512.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic acids symposium series,

Hallmann, A., Michnowska, A., Chomiczewska, A., Lipiński, M., & Smolarz, K. (2022). Bivalves transmissible neoplasia: biochemical aspects of contagious cancer in a clam Macoma balthica. *Cellular Physiology and Biochemistry*, *56*(6).

Hamede, R. K., McCallum, H., & Jones, M. (2013). Biting injuries and transmission of T asmanian devil facial tumour disease. *Journal of Animal Ecology*, *82*(1), 182-190.

Hammel, M., Simon, A., Arbiol, C., Villalba, A., Burioli, E. A., Pépin, J. f., Lamy, J. b., Benabdelmouna, A., Bernard, I., & Houssin, M. (2022). Prevalence and polymorphism of a mussel transmissible cancer in Europe. *Molecular Ecology*, *31*(3), 736-751.

Han, Y. (2019). Analysis of the role of the Hippo pathway in cancer. *Journal of translational medicine*, *17*(1), 116.

Han, Y., Zhang, M., Lai, R., & Zhang, Z. (2021). Chemical modifications to increase the therapeutic potential of antimicrobial peptides. *Peptides*, *146*, 170666.

Hanahan, D. (2022). Hallmarks of cancer: new dimensions. *Cancer discovery*, *12*(1), 31-46.

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, *100*(1), 57-70.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, *144*(5), 646-674.

Hancock, R., & Patrzykat, A. (2002). Clinical development of cationic antimicrobial peptides: from natural to novel antibiotics. *Current drug targets-Infectious disorders*, *2*(1), 79-83.

Haney, E. F., Mansour, S. C., & Hancock, R. E. (2017). Antimicrobial peptides: an introduction. *Antimicrobial peptides: methods and protocols*, 3-22.

Hao, X., Yang, H., Wei, L., Yang, S., Zhu, W., Ma, D., Yu, H., & Lai, R. (2012). Amphibian cathelicidin fills the evolutionary gap of cathelicidin in vertebrate. *Amino acids*, *43*(2), 677-685.

Harrenstien, L. A., Munson, L., Seal, U. S., Zoo, A., & Group, A. A. M. C. S. (1996). Mammary cancer in captive wild felids and risk factors for its development: a retrospective study of the clinical behavior of 31 cases. *Journal of Zoo and Wildlife Medicine*, 468-476.

Harry, E. *Paired REad TEXTure Viewer. OpenGL Powered Pretext Contact Map Viewer.* In https://github.com/sanger-tol/PretextView

Hart, S. F., Yonemitsu, M. A., Giersch, R. M., Garrett, F. E., Beal, B. F., Arriagada, G., Davis, B. W., Ostrander, E. A., Goff, S. P., & Metzger, M. J. (2023). Centuries of genome instability and evolution in soft-shell clam, Mya arenaria, bivalve transmissible neoplasia. *Nature Cancer*, *4*(11), 1561-1574.

Hartley, G. A., Frankenberg, S. R., Robinson, N. M., MacDonald, A. J., Hamede, R. K., Burridge, C. P., Jones, M. E., Faulkner, T., Shute, H., & Rose, K. (2024). Genome of the endangered eastern quoll (Dasyurus viverrinus) reveals signatures of historical decline and pelage color evolution. *Communications Biology*, *7*(1), 636.

Hartmann, K. (2012). Clinical aspects of feline retroviruses: a review. *Viruses*, *4*(11), 2684-2710.

Harvey, J. (1964). An unidentified virus which causes the rapid production of tumours in mice. *Nature*, *204*(4963), 1104-1105.

Harvey, K. F., Zhang, X., & Thomas, D. M. (2013). The Hippo pathway and human cancer. *Nature reviews cancer*, *13*(4), 246-257.

Hawkins, C. E., Baars, C., Hesterman, H., Hocking, G., Jones, M. E., Lazenby, B., Mann, D., Mooney, N., Pemberton, D., & Pyecroft, S. (2006). Emerging disease and population decline of an island endemic, the Tasmanian devil Sarcophilus harrisii. *Biological Conservation*, *131*(2), 307-324.

Hawkins, C. E., McCallum, H., Mooney, N., Jones, M., & Holdsworth, M. (2008). *Sarcophilus harrisii. The IUCN Red List of Threatened Species*. Retrieved 05/09/2022 from

Hayden, E. C. (2014). The $1,000 genome. *Nature*, *507*(7492), 294.

Hayes, D. A., Kunde, D. A., Taylor, R. L., Pyecroft, S. B., Sohal, S. S., & Snow, E. T. (2017). ERBB3: A potential serum biomarker for early detection and therapeutic target for devil facial tumour 1 (DFT1). *PloS one*, *12*(6), e0177919.

Hellgren, O. (2015). Allelic variation at innate immune genes (avian β-defensins), within a natural population of great tits. *Journal of Avian Biology*, *46*(1), 113-118. https://doi.org/https://doi.org/10.1111/jav.00370

Hellgren, O., Sheldon, B. C., & Buckling, A. (2010). In vitro tests of natural allelic variation of innate immune genes (avian β-defensins) reveal functional differences in microbial inhibition. *Journal of evolutionary biology*, *23*(12), 2726-2730. https://doi.org/https://doi.org/10.1111/j.1420-9101.2010.02115.x

Higgs, R., Lynn, D. J., Cahalane, S., Alaña, I., Hewage, C. M., James, T., Lloyd, A. T., & O'Farrelly, C. (2007). Modification of chicken avian β-defensin-8 at positively selected amino acid sites enhances specific antimicrobial activity. *Immunogenetics*, *59*, 573-580. https://doi.org/DOI:10.1007/s00251-007-0219-5

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518-522.

Höffken, V., Hermann, A., Pavenstädt, H., & Kremerskothen, J. (2021). WWC proteins: important regulators of hippo signaling in cancer. *Cancers*, *13*(2), 306.

Hogg, C. J., Edwards, R. J., Farquharson, K. A., Silver, L. W., Brandies, P., Peel, E., Escalona, M., Jaya, F. R., Thavornkanlapachai, R., & Batley, K. (2024). Extant and extinct bilby genomes combined with Indigenous knowledge improve conservation of a unique Australian marsupial. *Nature Ecology & Evolution*, *8*(7), 1311-1326.

Hogg, C. J., Silver, L., McLennan, E. A., & Belov, K. (2023). Koala Genome Survey: An Open Data Resource to Improve Conservation Planning. *Genes*, *14*(3), 546.

Holdo, R. M., Sinclair, A. R., Dobson, A. P., Metzger, K. L., Bolker, B. M., Ritchie, M. E., & Holt, R. D. (2009). A disease-mediated trophic cascade in the Serengeti and its implications for ecosystem C. *PLoS biology*, *7*(9), e1000210.

Holliday, R. (1994). Longevity and fecundity in eutherian mammals. In *Genetics and evolution of aging* (pp. 217-225). Springer.

Holliday, R. (2006). Aging is no longer an unsolved problem in biology. *Annals of the New York Academy of Sciences*, *1067*(1), 1-9.

Hollings, T., Jones, M., Mooney, N., & Mccallum, H. (2014). Trophic cascades following the disease-induced decline of an apex predator, the Tasmanian devil. *Conservation Biology*, *28*(1), 63-75.

Hollox, E. (2009). Copy number variation of beta-defensins and relevance to disease. *Cytogenetic and Genome Research*, *123*(1-4), 148-155. https://doi.org/DOI: 10.1159/000184702

Hollox, E., Armour, J., & Barber, J. (2003). Extensive normal copy number variation of a β-defensin antimicrobial-gene cluster. *The American Journal of Human Genetics*, *73*(3), 591-600.

Hollox, E. J., & Armour, J. A. (2008). Directional and balancing selection in human beta-defensins. *BMC evolutionary biology*, *8*, 1-14.

Hollox, E. J., Huffmeier, U., Zeeuwen, P. L., Palla, R., Lascorz, J., Rodijk-Olthuis, D., Van De Kerkhof, P. C., Traupe, H., De Jongh, G., & Heijer, M. d. (2008). Psoriasis is associated with increased β-defensin genomic copy number. *Nature genetics*, *40*(1), 23-25. https://doi.org/doi:10.1038/ng.2007.48

Holmes, E. C. (2022). COVID-19—lessons for zoonotic disease. In: American Association for the Advancement of Science.

Hon, T., Mars, K., Young, G., Tsai, Y.-C., Karalius, J. W., Landolin, J. M., Maurer, N., Kudrna, D., Hardigan, M. A., & Steiner, C. C. (2020). Highly accurate long-read HiFi sequencing data for five complex genomes. *Scientific data*, *7*(1), 399.

Hopkins, D., & Gaynor, B. (1985). Schwannoma in a kowari (Dasyuroides byrnei). *Australian veterinary journal*, *62*(10), 340-341. https://doi.org/10.1111/j.1751-0813.1985.tb07655.x

Hoyt, J. R., Kilpatrick, A. M., & Langwig, K. E. (2021). Ecology and impacts of white-nose syndrome on bats. *Nature Reviews Microbiology*, *19*(3), 196-210.

Hsu, C.-H., Chen, C., Jou, M.-L., Lee, A. Y.-L., Lin, Y.-C., Yu, Y.-P., Huang, W.-T., & Wu, S.-H. (2005). Structural and DNA-binding studies on the bovine antimicrobial peptide, indolicidin: evidence for multiple conformations involved in binding to membranes and DNA. *Nucleic acids research*, *33*(13), 4053-4064.

Hua, R., Ma, Y.-S., Yang, L., Hao, J.-J., Hua, Q.-Y., Shi, L.-Y., Yao, X.-Q., Zhi, H.-Y., & Liu, Z. (2024). Experimental evidence for cancer resistance in a bat species. *Nature communications*, *15*(1), 1401.

Hussain, S. P., Hofseth, L. J., & Harris, C. C. (2003). Radical causes of cancer. *Nature reviews cancer*, *3*(4), 276-285.

Ibeh, N., Feigin, C. Y., Frankenberg, S. R., McCarthy, D. J., Pask, A. J., & Romero, I. G. (2024). De novo transcriptome assembly and genome annotation of the fat-tailed dunnart (Sminthopsis crassicaudata). *Gigabyte*, *2024*.

Ikonomopoulou, M. P., Lopez-Mancheno, Y., Novelle, M. G., Martinez-Una, M., Gangoda, L., Pal, M., Costa-Machado, L. F., Fernandez-Marcos, P. J., Ramm, G. A., & Fernandez-Rojo, M. A. (2021). LXR stimulates a metabolic switch and reveals cholesterol homeostasis as a statin target in Tasmanian devil facial tumor disease. *Cell reports*, *34*(11).

Ingles, E. D., & Deakin, J. E. (2015). Global DNA Methylation patterns on marsupial and devil facial tumour chromosomes. *Molecular cytogenetics*, *8*, 1-11.

Ishige, T., Hara, H., Hirano, T., Kono, T., & Hanzawa, K. (2021). Genetic diversity of Japanese quail cathelicidins. *Poultry Science*, *100*(5), 101046. https://doi.org/https://doi.org/10.1016/j.psj.2021.101046

Itano, N., Atsumi, F., Sawai, T., Yamada, Y., Miyaishi, O., Senga, T., Hamaguchi, M., & Kimata, K. (2002). Abnormal accumulation of hyaluronan matrix diminishes contact inhibition of cell growth and promotes cell migration. *Proceedings of the National Academy of Sciences*, *99*(6), 3609-3614.

Jackson, S. (2007). *Australian mammals: biology and captive management*. Csiro Publishing.

James, S., Jennings, G., Kwon, Y. M., Stammnitz, M., Fraik, A., Storfer, A., Comte, S., Pemberton, D., Fox, S., & Brown, B. (2019). Tracing the rise of malignant

cell lines: Distribution, epidemiology and evolutionary interactions of two transmissible cancers in Tasmanian devils. *Evolutionary applications*, *12*(9), 1772-1780.

Jeggo, P. A., Pearl, L. H., & Carr, A. M. (2016). DNA repair, genome stability and cancer: a historical perspective. *Nature reviews cancer*, *16*(1), 35-42.

Jenssen, H., Hamill, P., & Hancock, R. E. (2006). Peptide antimicrobial agents. *Clinical microbiology reviews*, *19*(3), 491-511.

Jeon, H., Le, M. T., Ahn, B., Cho, H.-s., Yum, J., Hong, K., Kim, J.-H., Song, H., & Park, C. (2019). Copy number variation of PR-39 cathelicidin, and identification of PR-35, a natural variant of PR-39 with reduced mammalian cytotoxicity. *Gene*, *692*, 88-93. https://doi.org/https://doi.org/10.1016/j.gene.2018.12.065

John, K., Alla, V., Meier, C., & Pützer, B. M. (2011). GRAMD4 mimics p53 and mediates the apoptotic function of p73 at mitochondria. *Cell Death & Differentiation*, *18*(5), 874-886.

Johnson, R. N., O'Meally, D., Chen, Z., Etherington, G. J., Ho, S. Y., Nash, W. J., Grueber, C. E., Cheng, Y., Whittington, C. M., & Dennison, S. (2018). Adaptation and conservation insights from the koala genome. *Nature genetics*, *50*(8), 1102-1111. https://doi.org/https://doi.org/10.1038/s41588-018-0153-5

Jones, E. A., Cheng, Y., O'Meally, D., & Belov, K. (2017). Characterization of the antimicrobial peptide family defensins in the Tasmanian devil (Sarcophilus harrisii), koala (Phascolarctos cinereus), and tammar wallaby (Macropus eugenii). *Immunogenetics*, *69*, 133-143. https://doi.org/DOI:10.1007/s00251-016-0959-1

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, *451*(7181), 990-993.

Jones, M. E., Paetkau, D., Geffen, E., & Moritz, C. (2004). Genetic diversity and population structure of Tasmanian devils, the largest marsupial carnivore. *Molecular Ecology*, *13*(8), 2197-2209.

Jurd, R. D. (1994). "Not proper mammals": immunity in monotremes and marsupials. *Comparative immunology, microbiology and infectious diseases*, *17*(1), 41-52.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods*, *14*(6), 587-589.

Kassambara, A., & Mundt, F. (2017). Package factoextra: Extract and visualize the Results of Multivariate Data Analyses. R Package Version 1.0. 7. In: Foundation for Statistical Computing, Vienna, Austria.

Katoh, K., Misawa, K., Kuma, K. i., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, *30*(14), 3059-3066.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772-780.

Kayigwe, A. N., M. Darby, J., Lyons, A. B., L. Patchett, A., Lisowski, L., Liu, G.-S., & S. Flies, A. (2022). A human adenovirus encoding IFN-γ can transduce Tasmanian devil facial tumour cells and upregulate MHC-I. *Journal of General Virology*, *103*(11), 001812.

Keane, M., Semeiks, J., Webb, A. E., Li, Y. I., Quesada, V., Craig, T., Madsen, L. B., van Dam, S., Brawand, D., & Marques, P. I. (2015). Insights into the evolution of longevity from the bowhead whale genome. *Cell reports*, *10*(1), 112-122.

Khalil, A., & Nemer, G. (2020). The potential oncogenic role of the RAS-like GTP-binding gene RIT1 in glioblastoma. *Cancer Biomarkers*, *29*(4), 509-519.

Killian, J. K., Buckley, T. R., Stewart, N., Munday, B. L., & Jirtle, R. L. (2001). Marsupials and Eutherians reunited: genetic evidence for the Theria hypothesis of mammalian evolution. *Mammalian Genome*, *12*, 513-517.

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology*, *37*(8), 907-915.

Kimes, P. K., Liu, Y., Neil Hayes, D., & Marron, J. S. (2017). Statistical significance for hierarchical clustering. *Biometrics*, *73*(3), 811-821.

Kirkwood, T. B. (1977). Evolution of ageing. *Nature*, *270*(5635), 301-304.

Kirkwood, T. B., & Holliday, R. (1979). The evolution of ageing and longevity. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *205*(1161), 531-546.

Kloch, A., Wenzel, M. A., Laetsch, D. R., Michalski, O., Bajer, A., Behnke, J. M., Welc-Falęciak, R., & Piertney, S. B. (2018). Signatures of balancing selection in toll-like receptor (TLRs) genes–novel insights from a free-living rodent. *Scientific Reports*, *8*(1), 8361.

Koczulla, R., Von Degenfeld, G., Kupatt, C., Krötz, F., Zahler, S., Gloe, T., Issbrücker, K., Unterberger, P., Zaiou, M., & Lebherz, C. (2003). An angiogenic role for the human peptide antibiotic LL-37/hCAP-18. *The Journal of clinical investigation*, *111*(11), 1665-1672.

Koh, J., Itahana, Y., Mendenhall, I. H., Low, D., Soh, E. X. Y., Guo, A. K., Chionh, Y. T., Wang, L.-F., & Itahana, K. (2019). ABCB1 protects bat cells from DNA damage induced by genotoxic compounds. *Nature communications*, *10*(1), 1-14.

Kollipara, A., George, C., Hanger, J., Loader, J., Polkinghorne, A., Beagley, K., & Timms, P. (2012). Vaccination of healthy and diseased koalas (Phascolarctos cinereus) with a Chlamydia pecorum multi-subunit vaccine: evaluation of immunity and pathology. *Vaccine*, *30*(10), 1875-1885.

Koo, H. B., & Seo, J. (2019). Antimicrobial peptides under clinical investigation. *Peptide Science*, *111*(5), e24122.

Kosack, L., Wingelhofer, B., Popa, A., Orlova, A., Agerer, B., Vilagos, B., Majek, P., Parapatics, K., Lercher, A., & Ringler, A. (2019). The ERBB-STAT3 axis drives Tasmanian devil facial tumor disease. *Cancer Cell*, *35*(1), 125-139. e129.

Kosakovsky Pond, S. L., & Frost, S. D. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and*

Evolution, *22*(5), 1208-1222. https://doi.org/https://doi.org/10.1093/molbev/msi105

Kościuczuk, E. M., Lisowski, P., Jarczak, J., Strzałkowska, N., Jóźwik, A., Horbańczuk, J., Krzyżewski, J., Zwierzchowski, L., & Bagnicka, E. (2012). Cathelicidins: family of antimicrobial peptides. A review. *Molecular Biology Reports*, *39*, 10957-10970.

Kozakiewicz, C. P., Fraik, A. K., Patton, A. H., Ruiz-Aravena, M., Hamilton, D. G., Hamede, R., McCallum, H., Hohenlohe, P. A., Margres, M. J., & Jones, M. E. (2021). Spatial variation in gene expression of Tasmanian devil facial tumors despite minimal host transcriptomic response to infection. *BMC genomics*, *22*, 1-19.

Kreiss, A., Brown, G., Tovar, C., Lyons, A., & Woods, G. (2015). Evidence for induction of humoral and cytotoxic immune responses against devil facial tumor disease cells in Tasmanian devils (Sarcophilus harrisii) immunized with killed cell preparations. *Vaccine*, *33*(26), 3016-3025.

Kreiss, A., Cheng, Y., Kimble, F., Wells, B., Donovan, S., Belov, K., & Woods, G. M. (2011). Allorecognition in the Tasmanian devil (Sarcophilus harrisii), an endangered marsupial species with limited genetic diversity. *PloS one*, *6*(7), e22402.

Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., Li, M., Stecher, G., & Hedges, S. B. (2022). TimeTree 5: an expanded resource for species divergence times. *Molecular Biology and Evolution*, *39*(8), msac174.

Kunz, M., Löffler-Wirth, H., Dannemann, M., Willscher, E., Doose, G., Kelso, J., Kottek, T., Nickel, B., Hopp, L., & Landsberg, J. (2018). RNA-seq analysis identifies different transcriptomic types and developmental trajectories of primary melanomas. *Oncogene*, *37*(47), 6136-6151.

Kwon, Y. M., Gori, K., Park, N., Potts, N., Swift, K., Wang, J., Stammnitz, M. R., Cannell, N., Baez-Ortega, A., & Comte, S. (2020). Evolution and lineage dynamics of a transmissible cancer in Tasmanian devils. *PLoS biology*, *18*(11), e3000926.

Lachish, S., McCALLUM, H., Mann, D., Pukk, C. E., & Jones, M. E. (2010). Evaluation of selective culling of infected individuals to control Tasmanian devil facial tumor disease. *Conservation Biology*, *24*(3), 841-851.

Ladds, P. (2009). 37. Neoplasia and Related Proliferations in Terrestrial Mammals. In *Pathology of Australian native wildlife* (pp. 429-457). BioOne.

Langwig, K. E., Voyles, J., Wilber, M. Q., Frick, W. F., Murray, K. A., Bolker, B. M., Collins, J. P., Cheng, T. L., Fisher, M. C., & Hoyt, J. R. (2015). Context-dependent conservation responses to emerging wildlife diseases. *Frontiers in Ecology and the Environment*, *13*(4), 195-202.

Lapointe, J., Li, C., Higgins, J. P., Van De Rijn, M., Bair, E., Montgomery, K., Ferrari, M., Egevad, L., Rayford, W., & Bergerheim, U. (2004). Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proceedings of the National Academy of Sciences*, *101*(3), 811-816.

Lariviere, D., Ostrovsky, A., Gallardo, C., Syme, A., Abueg, L., Pickett, B., Formenti, G., Sozzoni, M., & Nekrutenko, A. (2024). Vertebrate genome assembly using HiFi, Bionano and Hi-C data-Step by Step.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., & Lopez, R. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947-2948.

Lau, Q., Griffith, J. E., & Higgins, D. P. (2014). Identification of MHCII variants associated with chlamydial disease in the koala (Phascolarctos cinereus). *PeerJ*, *2*, e443.

Laurell, C., Velázquez-Fernández, D., Lindsten, K., Juhlin, C., Enberg, U., Geli, J., Höög, A., Kjellman, M., Lundeberg, J., & Hamberger, B. (2009). Transcriptional profiling enables molecular classification of adrenocortical tumours. *European Journal of Endocrinology*, *161*(1), 141-152.

Laurie, E. (1946). The reproduction of the house-mouse (Mus musculus) living in different environments. *Proceedings of the Royal Society of London. Series B-Biological Sciences*, *133*(872), 248-281.

Lauth, X., Shike, H., Burns, J. C., Westerman, M. E., Ostland, V. E., Carlberg, J. M., Van Olst, J. C., Nizet, V., Taylor, S. W., & Shimizu, C. (2002). Discovery and characterization of two isoforms of moronecidin, a novel antimicrobial peptide from hybrid striped bass. *Journal of Biological Chemistry*, *277*(7), 5030-5039.

Lazenby, B. T., Tobler, M. W., Brown, W. E., Hawkins, C. E., Hocking, G. J., Hume, F., Huxtable, S., Iles, P., Jones, M. E., & Lawrence, C. (2018). Density trends and demographic signals uncover the long-term impact of transmissible cancer in Tasmanian devils. *Journal of Applied Ecology*, *55*(3), 1368-1379.

Lazzaro, B. P., Zasloff, M., & Rolff, J. (2020). Antimicrobial peptides: Application informed by evolution. *Science*, *368*(6490), eaau5480. https://doi.org/https://doi.org/10.1126/science.aau5480

Lee, T., Zenger, K. R., Close, R. L., & Phalen, D. N. (2011). Genetic analysis reveals a distinct and highly diverse koala (Phascolarctos cinereus) population in South Gippsland, Victoria, Australia. *Australian Mammalogy*, *34*(1), 68-74. https://doi.org/https://doi.org/10.1071/AM10035

Legge, S., Rumpff, L., Garnett, S. T., & Woinarski, J. C. (2023). Loss of terrestrial biodiversity in Australia: Magnitude, causation, and response. *Science*, *381*(6658), 622-631.

Legione, A. R., Patterson, J. L., Whiteley, P. L., Amery-Gale, J., Lynch, M., Haynes, L., Gilkerson, J. R., Polkinghorne, A., Devlin, J. M., & Sansom, F. M. (2016). Identification of unusual Chlamydia pecorum genotypes in Victorian koalas (Phascolarctos cinereus) and clinical variables associated with infection. *Journal of Medical Microbiology*, *65*(5), 420-428. https://doi.org/http://doi.org/doi:https://doi.org/10.1099/jmm.0.000241

Lehmann, B. D., Bauer, J. A., Chen, X., Sanders, M. E., Chakravarthy, A. B., Shyr, Y., & Pietenpol, J. A. (2011). Identification of human triple-negative breast cancer

subtypes and preclinical models for selection of targeted therapies. *The Journal of clinical investigation*, *121*(7), 2750-2767.

Lehrer, R. I., & Ganz, T. (2002). Defensins of vertebrate animals. *Current opinion in immunology*, *14*(1), 96-102.

Lei, Z., Tan, I. B., Das, K., Deng, N., Zouridis, H., Pattison, S., Chua, C., Feng, Z., Guan, Y. K., & Ooi, C. H. (2013). Identification of molecular subtypes of gastric cancer with different responses to PI3-kinase inhibitors and 5-fluorouracil. *Gastroenterology*, *145*(3), 554-565.

Leifels, M., Khalilur Rahman, O., Sam, I.-C., Cheng, D., Chua, F. J. D., Nainani, D., Kim, S. Y., Ng, W. J., Kwok, W. C., & Sirikanchana, K. (2022). The one health perspective to improve environmental surveillance of zoonotic viruses: lessons from COVID-19 and outlook beyond. *ISME communications*, *2*(1), 107.

Letunic, I., Khedkar, S., & Bork, P. (2021). SMART: recent updates, new developments and status in 2020. *Nucleic acids research*, *49*(D1), D458-D460.

Levine, A. J., Momand, J., & Finlay, C. A. (1991). The p53 tumour suppressor gene. *Nature*, *351*(6326), 453-456.

Levitt, N. C., & Hickson, I. D. (2002). Caretaker tumour suppressor genes that defend genome integrity. *Trends in molecular medicine*, *8*(4), 179-186.

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., Durbin, R., Edwards, S. V., Forest, F., & Gilbert, M. T. P. (2018). Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences*, *115*(17), 4325-4333.

Li, D., Zhang, L., Yin, H., Xu, H., Trask, J. S., Smith, D. G., Li, Y., Yang, M., & Zhu, Q. (2014). Evolution of primate α and θ defensins revealed by analysis of genomes. *Molecular Biology Reports*, *41*, 3859-3866.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094-3100.

Li, N., Qin, H., Zhu, F., Ding, H., Chen, Y., Lin, Y., Deng, R., Ma, T., Lv, Y., & Xiong, C. (2024). Potent prophylactic cancer vaccines harnessing surface antigens shared by tumour cells and induced pluripotent stem cells. *Nature Biomedical Engineering*, 1-19.

Li, X.-F., Ren, P., Shen, W.-Z., Jin, X., & Zhang, J. (2020). The expression, modulation and use of cancer-testis antigens as potential biomarkers for cancer immunotherapy. *American journal of translational research*, *12*(11), 7002.

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general-purpose read summarization program. *Bioinformatics*, *30*(7), 923-930.

Lightbody, G., Haberland, V., Browne, F., Taggart, L., Zheng, H., Parkes, E., & Blayney, J. K. (2019). Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application. *Briefings in bioinformatics*, *20*(5), 1795-1811.

Lima-de-Faria, A., Mitelman, F., Blomberg, J., & Pfeifer-Ohlsson, S. (1991). Telomeric location of retroviral oncogenes in humans. *Hereditas*, *114*(3), 207-211.

Linzmeier, R. M., & Ganz, T. (2005). Human defensin gene copy number polymorphisms: comprehensive analysis of independent variation in α-and β-defensin regions at 8p22–p23. *Genomics*, *86*(4), 423-430. https://doi.org/https://doi.org/10.1016/j.ygeno.2005.06.003

Liu, J., Fu, M., Wang, M., Wan, D., Wei, Y., & Wei, X. (2022). Cancer vaccines as promising immuno-therapeutics: platforms and current progress. *Journal of Hematology & Oncology*, *15*(1), 28.

Liu, X., Luo, J., Wang, X., Zhang, Y., & Chen, J. (2025). Directed evolution of antimicrobial peptides using multi-objective zeroth-order optimization. *Briefings in bioinformatics*, *26*(1), bbae715.

Ljubuncic, P., & Reznick, A. Z. (2009). The evolutionary theories of aging revisited–a mini-review. *Gerontology*, *55*(2), 205-216.

Loh, R., Bergfeld, J., Hayes, D., O'hara, A., Pyecroft, S., Raidal, S., & Sharpe, R. (2006). The pathology of devil facial tumor disease (DFTD) in Tasmanian devils (Sarcophilus harrisii). *Veterinary Pathology*, *43*(6), 890-895.

Lombard, L. S., & Witte, E. J. (1959). Frequency and types of tumors in mammals and birds of the Philadelphia Zoological Garden. *Cancer research*, *19*(2), 127-141.

López-García, B., Lee, P. H., Yamasaki, K., & Gallo, R. L. (2005). Anti-fungal activity of cathelicidins and their potential role in Candida albicans skin infection. *Journal of Investigative Dermatology*, *125*(1), 108-115.

Lorch, J. M., Gargas, A., Meteyer, C. U., Berlowski-Zier, B. M., Green, D. E., Shearn-Bochsler, V., Thomas, N. J., & Blehert, D. S. (2010). Rapid polymerase chain reaction diagnosis of white-nose syndrome in bats. *Journal of Veterinary Diagnostic Investigation*, *22*(2), 224-230.

Lott, M. J., Wright, B. R., Neaves, L. E., Frankham, G. J., Dennison, S., Eldridge, M. D., Potter, S., Alquezar-Planas, D. E., Hogg, C. J., & Belov, K. (2022). Future-proofing the koala: Synergising genomic and environmental data for effective species management. *Molecular Ecology*, *31*(11), 3035-3055. https://doi.org/https://doi.org/10.1111/mec.16446

Louro, R., Nakaya, H. I., Paquola, A. C., Martins, E. A., Da Silva, A. M., Verjovski-Almeida, S., & Reis, E. M. (2004). RASL11A, member of a novel small monomeric GTPase gene family, is down-regulated in prostate tumors. *Biochemical and biophysical research communications*, *316*(3), 618-627.

Lovell, J. T., Sreedasyam, A., Schranz, M. E., Wilson, M., Carlson, J. W., Harkess, A., Emms, D., Goodstein, D. M., & Schmutz, J. (2022). GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *Elife*, *11*, e78526.

Luo, Y., & Song, Y. (2021). Mechanism of antimicrobial peptides: antimicrobial, anti-inflammatory and antibiofilm activities. *International Journal of Molecular Sciences*, *22*(21), 11401.

Lynch, M. (2008). 13. Bandicoots and bilbies. *Medicine of Australian mammals*, 439.

Lynn, D. J., & Bradley, D. G. (2007). Discovery of α-defensins in basal mammals. *Developmental & Comparative Immunology*, *31*(10), 963-967.

Machado, L. R., & Ottolini, B. (2015). An evolutionary history of defensins: a role for copy number variation in maximizing host innate and adaptive immune responses. *Frontiers in Immunology*, *6*, 132185. https://doi.org/doi:10.3389/fimmu.2015.00115

Machalaba, C., Feferholtz, Y., Uhart, M., & Karesh, W. (2020). Wildlife conservation status and disease trends: ten years of reports to the Worldwide Monitoring System for Wild Animal Diseases. *Rev. Sci. Tech. Jan*, *39*(3), 991-1001.

MacInnes, C. D., Smith, S. M., Tinline, R. R., Ayers, N. R., Bachmann, P., Ball, D. G., Calder, L. A., Crosgrey, S. J., Fielding, C., & Hauschildt, P. (2001). Elimination of rabies from red foxes in eastern Ontario. *Journal of wildlife diseases*, *37*(1), 119-132.

Mackie, J., Gillett, A., Palmieri, C., Feng, T., & Higgins, D. (2016). Pneumonia due to Chlamydia pecorum in a koala (Phascolarctos cinereus). *Journal of Comparative Pathology*, *155*(4), 356-360.

MacRae, S. L., Zhang, Q., Lemetre, C., Seim, I., Calder, R. B., Hoeijmakers, J., Suh, Y., Gladyshev, V. N., Seluanov, A., & Gorbunova, V. (2015). Comparative analysis of genome maintenance genes in naked mole rat, mouse, and human. *Aging cell*, *14*(2), 288-291.

Madsen, T., Arnal, A., Vittecoq, M., Bernex, F., Abadie, J., Labrut, S., Garcia, D., Faugère, D., Lemberger, K., & Beckmann, C. (2017). Cancer prevalence and etiology in wild and captive animals. In *Ecology and evolution of cancer* (pp. 11-46). Elsevier.

Mahmoud, M. M., Alenezi, M., Al-Hejin, A. M., Abujamel, T. S., Aljoud, F., Noorwali, A., Awad, I. A., Alkhaled, M., & Yacoub, H. A. (2022). Anticancer activity of chicken cathelicidin peptides against different types of cancer. *Molecular Biology Reports*, *49*(6), 4321-4339.

Maidment, T. I., Bryan, E. R., Pyne, M., Barnes, M., Eccleston, S., Cunningham, S., Whitlock, E., Redman, K., Nicolson, V., & Beagley, K. W. (2023). Characterisation of the koala (Phascolarctos cinereus) pouch microbiota in a captive population reveals a dysbiotic compositional profile associated with neonatal mortality. *Microbiome*, *11*(1), 75. https://doi.org/DOI:10.1186/s40168-023-01527-9

Maier, V. H., Dorn, K. V., Gudmundsdottir, B. K., & Gudmundsson, G. H. (2008). Characterisation of cathelicidin gene family members in divergent fish species. *Molecular immunology*, *45*(14), 3723-3730.

Mangoni, M. L., McDermott, A. M., & Zasloff, M. (2016). Antimicrobial peptides and wound healing: biological and therapeutic considerations. *Experimental dermatology*, *25*(3), 167-173.

Margres, M. J., Ruiz-Aravena, M., Hamede, R., Chawla, K., Patton, A. H., Lawrance, M. F., Fraik, A. K., Stahlke, A. R., Davis, B. W., & Ostrander, E. A. (2020). Spontaneous tumor regression in Tasmanian devils associated with RASL11A activation. *Genetics*, *215*(4), 1143-1152.

Margres, M. J., Ruiz-Aravena, M., Hamede, R., Jones, M. E., Lawrance, M. F., Hendricks, S. A., Patton, A., Davis, B. W., Ostrander, E. A., & McCallum, H. (2018). The genomic basis of tumor regression in Tasmanian devils (Sarcophilus harrisii). *Genome biology and evolution*, *10*(11), 3012-3025.

Marnett, L. J., & Plastaras, J. P. (2001). Endogenous DNA damage and mutation. *Trends in Genetics*, *17*(4), 214-221.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, *17*(1), 10-12.

Martin, M. L., & Weisbecker, V. (2023). Function and constraint in the marsupial postcranium. In *American and Australasian Marsupials: An Evolutionary, Biogeographical, and Ecological Approach* (pp. 403-429). Springer.

Mason, G. J. (2010). Species differences in responses to captivity: stress, welfare and the comparative method. *Trends in ecology & evolution*, *25*(12), 713-721.

Mathew, M., Beagley, K. W., Timms, P., & Polkinghorne, A. (2013). Preliminary characterisation of tumor necrosis factor alpha and interleukin-10 responses to Chlamydia pecorum infection in the koala (Phascolarctos cinereus). *PloS one*, *8*(3), e59958.

Mathew, M., Pavasovic, A., Prentis, P. J., Beagley, K. W., Timms, P., & Polkinghorne, A. (2013). Molecular characterisation and expression analysis of Interferon gamma in response to natural Chlamydia infection in the koala, Phascolarctos cinereus. *Gene*, *527*(2), 570-577.

McAloose, D., Munson, L., & Naydan, D. (2007). Histologic features of mammary carcinomas in zoo felids treated with melengestrol acetate (MGA) contraceptives. *Veterinary Pathology*, *44*(3), 320-326.

McAloose, D., & Newton, A. L. (2009). Wildlife cancer: a conservation perspective. *Nature reviews cancer*, *9*(7), 517-526.

McCarthy, D. J., Campbell, K. R., Lun, A. T., & Wills, Q. F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics*, *33*(8), 1179-1186.

McClatchey, A. I., & Yap, A. S. (2012). Contact inhibition (of proliferation) redux. *Current opinion in cell biology*, *24*(5), 685-694.

McEachran, M. C., Harvey, J. A., Mummah, R. O., Bletz, M. C., Teitelbaum, C. S., Rosenblatt, E., Rudolph, F. J., Arce, F., Yin, S., & Prosser, D. J. (2024). Reframing wildlife disease management problems with decision analysis. *Conservation Biology*, *38*(4), e14284.

McEwen, G. K., Alquezar-Planas, D. E., Dayaram, A., Gillett, A., Tarlinton, R., Mongan, N., Chappell, K. J., Henning, J., Tan, M., & Timms, P. (2021). Retroviral integrations contribute to elevated host cancer rates during germline invasion. *Nature communications*, *12*(1), 1316.

Meade, K. G., Cahalane, S., Narciandi, F., Cormican, P., Lloyd, A. T., & O'Farrelly, C. (2008). Directed alteration of a novel bovine β-defensin to improve antimicrobial efficacy against methicillin-resistant Staphylococcus aureus (MRSA).

International journal of antimicrobial agents, *32*(5), 392-397. https://doi.org/https://doi.org/10.1016/j.ijantimicag.2008.05.005

Melzer, A., Carrick, F., Menkhorst, P., Lunney, D., & John, B. S. (2000). Overview, critical assessment, and conservation implications of koala distribution and abundance. *Conservation Biology*, *14*(3), 619-628.

Mendes, F. K., Vanderpool, D., Fulton, B., & Hahn, M. W. (2020). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics*, *36*(22-23), 5516-5518.

Metzger, M. J., & Goff, S. P. (2016). A sixth modality of infectious disease: contagious cancer from devils to clams and beyond. *PLoS Pathogens*, *12*(10), e1005904.

Metzger, M. J., Villalba, A., Carballal, M. J., Iglesias, D., Sherry, J., Reinisch, C., Muttray, A. F., Baldwin, S. A., & Goff, S. P. (2016). Widespread transmission of independent cancer lineages within multiple bivalve species. *Nature*, *534*(7609), 705-709.

Mikkelsen, T. S., Wakefield, M. J., Aken, B., Amemiya, C. T., Chang, J. L., Duke, S., Garber, M., Gentles, A. J., Goodstadt, L., & Heger, A. (2007). Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. *Nature*, *447*(7141), 167-177.

Milanese, M., Segat, L., Arraes, L. C., Garzino-Demo, A., & Crovella, S. (2009). Copy number variation of defensin genes and HIV infection in Brazilian children. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *50*(3), 331-333.

Miller, W., Hayes, V. M., Ratan, A., Petersen, D. C., Wittekindt, N. E., Miller, J., Walenz, B., Knight, J., Qi, J., & Zhao, F. (2011). Genetic diversity and population structure of the endangered marsupial Sarcophilus harrisii (Tasmanian devil). *Proceedings of the National Academy of Sciences*, *108*(30), 12348-12353.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, *37*(5), 1530-1534.

Minias, P., Pikus, E., Whittingham, L. A., & Dunn, P. O. (2019). Evolution of copy number at the MHC varies across the avian tree of life. *Genome biology and evolution*, *11*(1), 17-28. https://doi.org/https://doi.org/10.1093/gbe/evy253

Minias, P., & Vinkler, M. (2022). Selection balancing at innate immune genes: Adaptive polymorphism maintenance in toll-like receptors. *Molecular Biology and Evolution*, *39*(5), msac102. https://doi.org/https://doi.org/10.1093/molbev/msac102

Molhoek, E. M., Van Dijk, A., Veldhuizen, E. J., Dijk-Knijnenburg, H., Mars-Groenendijk, R. H., Boele, L. C., Kaman-van Zanten, W. E., Haagsman, H. P., & Bikker, F. J. (2010). Chicken cathelicidin-2-derived peptides with enhanced immunomodulatory and antibacterial activities against biological warfare agents. *International journal of antimicrobial agents*, *36*(3), 271-274. https://doi.org/https://doi.org/10.1016/j.ijantimicag.2010.06.001

Monteleone, G., Calascibetta, D., Scaturro, M., Galluzzo, P., Palmeri, M., Riggio, V., & Portolano, B. (2011). Polymorphisms of β -defensin genes in Valle del Belice

dairy sheep. *Molecular Biology Reports*, *38*, 5405-5412. https://doi.org/DOI:10.1007/s11033-011-0694-5

Morris, K., Austin, J. J., & Belov, K. (2013). Low major histocompatibility complex diversity in the Tasmanian devil predates European settlement and may explain susceptibility to disease epidemics. *Biology letters*, *9*(1), 20120900.

Morris, K. M., O'Meally, D., Zaw, T., Song, X., Gillett, A., Molloy, M. P., Polkinghorne, A., & Belov, K. (2016). Characterisation of the immune compounds in koala milk using a combined transcriptomic and proteomic approach. *Scientific Reports*, *6*(1), 35011.

Morris, K. M., Wright, B., Grueber, C. E., Hogg, C., & Belov, K. (2015). Lack of genetic diversity across diverse immune genes in an endangered mammal, the Tasmanian devil (S arcophilus harrisii). *Molecular Ecology*, *24*(15), 3860-3872.

Morris, L. G., Riaz, N., Desrichard, A., Şenbabaoğlu, Y., Hakimi, A. A., Makarov, V., Reis-Filho, J. S., & Chan, T. A. (2016). Pan-cancer analysis of intratumor heterogeneity as a prognostic determinant of survival. *Oncotarget*, *7*(9), 10051.

Morrison, C. E., Hogg, C. J., Gales, R., Johnson, R. N., & Grueber, C. E. (2020). Low innate immune-gene diversity in the critically endangered orange-bellied parrot (Neophema chrysogaster). *Emu-Austral Ornithology*, *120*(1), 56-64. https://doi.org/https://doi.org/10.1080/01584197.2019.1686994

Morton, S., & Fletcher, T. (1989). Fauna of Australia.

Mu, D., Tursun, M., Duckett, D. R., Drummond, J. T., Modrich, P., & Sancar, A. (1997). Recognition and repair of compound DNA lesions (base damage and mismatch) by human mismatch repair and excision repair systems. *Molecular and cellular biology*.

Mukherjee, S., Sarkar-Roy, N., Wagener, D. K., & Majumder, P. P. (2009). Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. *Proceedings of the National Academy of Sciences*, *106*(17), 7073-7078. https://doi.org/https://doi.org/10.1073/pnas.0811357106

Munson, L., & Moresco, A. (2007). Comparative pathology of mammary gland cancers in domestic and wild animals. *Breast disease*, *28*(1), 7-21.

Murchison, E. P., Schulz-Trieglaff, O. B., Ning, Z., Alexandrov, L. B., Bauer, M. J., Fu, B., Hims, M., Ding, Z., Ivakhno, S., & Stewart, C. (2012). Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. *Cell*, *148*(4), 780-791.

Murchison, E. P., Tovar, C., Hsu, A., Bender, H. S., Kheradpour, P., Rebbeck, C. A., Obendorf, D., Conlan, C., Bahlo, M., & Blizzard, C. A. (2010). The Tasmanian devil transcriptome reveals Schwann cell origins of a clonally transmissible cancer. *Science*, *327*(5961), 84-87.

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution*, *30*(5), 1196-1205. https://doi.org/https://doi.org/10.1093/molbev/mst030

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS genetics*, *8*(7), e1002764. https://doi.org/DOI:10.1371/journal.pgen.1002764

Nei, M., Gu, X., & Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences*, *94*(15), 7799-7806.

Nguyen, L. T., Haney, E. F., & Vogel, H. J. (2011). The expanding scope of antimicrobial peptide structures and their modes of action. *Trends in biotechnology*, *29*(9), 464-472.

Nirenberg, M., & Leder, P. (1964). RNA Codewords and Protein Synthesis: The Effect of Trinucleotides upon the Binding of sRNA to Ribosomes. *Science*, *145*(3639), 1399-1407.

Niyonsaba, F., Ushio, H., Nakano, N., Ng, W., Sayama, K., Hashimoto, K., Nagaoka, I., Okumura, K., & Ogawa, H. (2007). Antimicrobial peptides human β-defensins stimulate epidermal keratinocyte migration, proliferation and production of proinflammatory cytokines and chemokines. *Journal of Investigative Dermatology*, *127*(3), 594-604.

Nunney, L. (2013). The real war on cancer: the evolutionary dynamics of cancer suppression. *Evolutionary applications*, *6*(1), 11-19.

Nunney, L., Maley, C. C., Breen, M., Hochberg, M. E., & Schiffman, J. D. (2015). Peto's paradox and the promise of comparative oncology. In (Vol. 370, pp. 20140177): The Royal Society.

Nyari, S., Waugh, C. A., Dong, J., Quigley, B. L., Hanger, J., Loader, J., Polkinghorne, A., & Timms, P. (2017). Epidemiology of chlamydial infection and disease in a free-ranging koala (Phascolarctos cinereus) population. *PloS one*, *12*(12), e0190114.

Obendorf, D. L. (1993). Diseases of dasyurid marsupials. *The biology and management of Australasian carnivorous marsupials (M. Roberts, J. Carnio, G. Crawshaw, and M. Hutchins, eds.). Metropolitan Toronto Zoo and the Monotreme and Marsupial Advisory Group of the American Association of Zoological Parks and Aquariums, Toronto, Ontario, Canada*, 39-46.

Ockert, L. E., McLennan, E. A., Fox, S., Belov, K., & Hogg, C. J. (2024). Characterising the Tasmanian devil (Sarcophilus harrisii) pouch microbiome in lactating and non-lactating females. *Scientific Reports*, *14*(1), 15188.

Oh, S. C., Sohn, B. H., Cheong, J.-H., Kim, S.-B., Lee, J. E., Park, K. C., Lee, S. H., Park, J.-L., Park, Y.-Y., & Lee, H.-S. (2018). Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. *Nature communications*, *9*(1), 1777.

Oikkonen, L., & Lise, S. (2017). Making the most of RNA-seq: Pre-processing sequencing data with Opossum for reliable SNP variant detection. *Wellcome open research*, *2*.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2018). Package 'vegan'. *Community ecology package, version*, *2*(3).

Old, J. M., & Deane, E. M. (2000). Development of the immune system and immunological protection in marsupial pouch young. *Developmental & Comparative Immunology*, *24*(5), 445-454. https://doi.org/https://doi.org/10.1016/S0145-305X(00)00008-2

Old, J. M., & Deane, E. M. (2003). The detection of mature T-and B-cells during development of the lymphoid tissues of the tammar wallaby (Macropus eugenii). *Journal of Anatomy*, *203*(1), 123-131.

Old, J. M., Selwood, L., & Deane, E. M. (2004). The appearance and distribution of mature T and B cells in the developing immune tissues of the stripe-faced dunnart (Sminthopsis macroura). *Journal of Anatomy*, *205*(1), 25-33.

Old, J. M., & Stannard, H. J. (2020). Conservation of quolls (Dasyurus spp.) in captivity–a review. *Australian Mammalogy*, *43*(3), 277-289.

Pan, D., & Zhang, L. (2008). Tandemly arrayed genes in vertebrate genomes. *International Journal of Genomics*, *2008*(1), 545269.

Park, C. B., Kim, H. S., & Kim, S. C. (1998). Mechanism of action of the antimicrobial peptide buforin II: buforin II kills microorganisms by penetrating the cell membrane and inhibiting cellular functions. *Biochemical and biophysical research communications*, *244*(1), 253-257.

Parra, Z. E., Baker, M. L., Schwarz, R. S., Deakin, J. E., Lindblad-Toh, K., & Miller, R. D. (2007). A unique T cell receptor discovered in marsupials. *Proceedings of the National Academy of Sciences*, *104*(23), 9776-9781.

Parrott, M. L., & Edwards, A. M. (2023). Reproductive strategies and biology of the Australasian marsupials. In *American and Australasian Marsupials: An Evolutionary, Biogeographical, and Ecological Approach* (pp. 1-49). Springer.

Patchett, A. L., Coorens, T. H., Darby, J., Wilson, R., McKay, M. J., Kamath, K. S., Rubin, A., Wakefield, M., Mcintosh, L., & Mangiola, S. (2020). Two of a kind: transmissible Schwann cell cancers in the endangered Tasmanian devil (Sarcophilus harrisii). *Cellular and Molecular Life Sciences*, 77, 1847-1858.

Patchett, A. L., Darby, J. M., Tovar, C., Lyons, A. B., & Woods, G. M. (2016). The immunomodulatory small molecule imiquimod induces apoptosis in devil facial tumour cell lines. *PloS one*, *11*(12), e0168068.

Patchett, A. L., Tovar, C., Corcoran, L. M., Lyons, A. B., & Woods, G. M. (2017). The toll-like receptor ligands Hiltonol®(polyICLC) and imiquimod effectively activate antigen-specific immune responses in Tasmanian devils (Sarcophilus harrisii). *Developmental & Comparative Immunology*, *76*, 352-360.

Patchett, A. L., Wilson, R., Charlesworth, J. C., Corcoran, L. M., Papenfuss, A. T., Lyons, B. A., Woods, G. M., & Tovar, C. (2018). Transcriptome and proteome profiling reveals stress-induced expression signatures of imiquimod-treated Tasmanian devil facial tumor disease (DFTD) cells. *Oncotarget*, *9*(22), 15895.

Patel, P., Sivaiah, B., & Patel, R. (2022). Approaches for finding optimal number of clusters using k-means and agglomerative hierarchical clustering techniques. 2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP),

Patil, A., Hughes, A. L., & Zhang, G. (2004). Rapid evolution and diversification of mammalian α-defensins as revealed by comparative analysis of rodent and primate genes. *Physiological genomics*, *20*(1), 1-11.

Patil, A. A., Cai, Y., Sang, Y., Blecha, F., & Zhang, G. (2005). Cross-species analysis of the mammalian β-defensin gene family: presence of syntenic gene clusters and preferential expression in the male reproductive tract. *Physiological genomics*, *23*(1), 5-17. https://doi.org/https://doi.org/10.1152/physiolgenomics.00104.2005

Patterson, J. L., Lynch, M., Anderson, G. A., Noormohammadi, A. H., Legione, A., Gilkerson, J. R., & Devlin, J. M. (2015). The prevalence and clinical significance of Chlamydia infection in island and mainland populations of Victorian koalas (Phascolarctos cinereus). *Journal of wildlife diseases*, *51*(2), 309-317. https://doi.org/https://doi.org/10.7589/2014-07-176

Paudel, Y., Madsen, O., Megens, H.-J., Frantz, L. A., Bosse, M., Bastiaansen, J. W., Crooijmans, R. P., & Groenen, M. A. (2013). Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. *BMC genomics*, *14*, 1-13.

Pavel, M., Renna, M., Park, S. J., Menzies, F. M., Ricketts, T., Füllgrabe, J., Ashkenazi, A., Frake, R. A., Lombarte, A. C., & Bento, C. F. (2018). Contact inhibition controls cell survival and proliferation via YAP/TAZ-autophagy axis. *Nature communications*, *9*(1), 1-18.

Pearse, A.-M., & Swift, K. (2006). Transmission of devil facial-tumour disease. *Nature*, *439*(7076), 549-549.

Pearse, A.-M., Swift, K., Hodson, P., Hua, B., McCallum, H., Pyecroft, S., Taylor, R., Eldridge, M. D., & Belov, K. (2012). Evolution in a transmissible cancer: a study of the chromosomal changes in devil facial tumor (DFT) as it spreads through the wild Tasmanian devil population. *Cancer genetics*, *205*(3), 101-112.

Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. *Current protocols in bioinformatics*, *42*(1), 3.1. 1-3.1. 8.

Peck, S., Michael, S., Knowles, G., Davis, A., & Pemberton, D. (2019). Causes of mortality and severe morbidity requiring euthanasia in captive Tasmanian devils (Sarcophilus harrisii) in Tasmania. *Australian veterinary journal*, *97*(4), 89-92.

Peel, E., Cheng, Y., Djordjevic, J., Fox, S., Sorrell, T., & Belov, K. (2016). Cathelicidins in the Tasmanian devil (Sarcophilus harrisii). *Scientific Reports*, *6*(1), 1-9. https://doi.org/DOI: 10.1038/srep35019

Peel, E., Cheng, Y., Djordjevic, J. T., Kuhn, M., Sorrell, T., & Belov, K. (2017). Marsupial and monotreme cathelicidins display antimicrobial activity, including against methicillin-resistant Staphylococcus aureus. *Microbiology*, *163*(10), 1457-1465. https://doi.org/DOI:10.1099/mic.0.000536

Peel, E., Cheng, Y., Djordjevic, J. T., O'Meally, D., Thomas, M., Kuhn, M., Sorrell, T. C., Huston, W. M., & Belov, K. (2021). Koala cathelicidin PhciCath5 has antimicrobial activity, including against Chlamydia pecorum. *PloS one*, *16*(4), e0249658. https://doi.org/https://doi.org/10.1371/journal.pone.0249658

Peel, E., Gonsalvez, A., Hogg, C. J., & Belov, K. (2025). Marsupial cathelicidins: characterization, antimicrobial activity and evolution in this unique mammalian lineage. *Frontiers in Immunology*, *16*, 1524092.

Peel, E., Hogg, C., & Belov, K. (2024). Characterisation of defensins across the marsupial family tree. *Developmental &Comparative Immunology*, 105207. https://doi.org/https://doi.org/10.1016/j.dci.2024.105207

Peel, E., Ng, H. J. J., & Belov, K. (2019). Monotreme, marsupial and bat immunology. *Current therapy in medicine of Australian mammals. Collingwood, Victoria, Australia: CSIRO Publishing*, 97-10.

Peel, E., Silver, L., Brandies, P., Zhu, Y., Cheng, Y., Hogg, C. J., & Belov, K. (2022). Best genome sequencing strategies for annotation of complex immune gene families in wildlife. *GigaScience*, *11*, giac100.

Perry, J. S. (1953). The reproduction of the African elephant, Loxodonta africana. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 93-149.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology*, *33*(3), 290-295.

Pessier, A., Stalis, I., Sutherland-Smith, M., Spelman, L., & Montali, R. (1999). Soft tissue sarcomas associated with identification microchip implants in two small zoo mammals. Proc. Am. Assoc. Zoo Vet. Annu. Meet,

Peto, R. (2016). Epidemiology, multistage models, and short-term mutagenicity tests. *International Journal of Epidemiology*, *45*(3), 621-637.

Petrohilos, C., Patchett, A., Hogg, C. J., Belov, K., & Peel, E. (2023). Tasmanian devil cathelicidins exhibit anticancer activity against Devil Facial Tumour Disease (DFTD) cells. *Scientific Reports*, *13*(1), 12698. https://doi.org/https://doi.org/10.1038/s41598-023-39901-0

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, *25*(13), 1605-1612.

Phillips, S., Quigley, B. L., & Timms, P. (2019). Seventy years of Chlamydia vaccine research–limitations of the past and directions for the future. *Frontiers in Microbiology*, *10*, 433459.

Phillips, S., Robbins, A., Loader, J., Hanger, J., Booth, R., Jelocnik, M., Polkinghorne, A., & Timms, P. (2018). Chlamydia pecorum gastrointestinal tract infection associations with urogenital tract infections in the koala (Phascolarctos cinereus). *PloS one*, *13*(11), e0206471.

Podlaszczuk, P., Indykiewicz, P., Markowski, J., & Minias, P. (2020). Relaxation of selective constraints shapes variation of toll-like receptors in a colonial waterbird, the black-headed gull. *Immunogenetics*, *72*, 251-262.

Polkinghorne, A., Hanger, J., & Timms, P. (2013). Recent advances in understanding the biology, epidemiology and control of chlamydial infections in koalas. *Veterinary microbiology*, *165*(3-4), 214-223.

Pond, S. L. K., & Frost, S. D. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics*, *21*(10), 2531-2533. https://doi.org/https://doi.org/10.1093/bioinformatics/bti320

Porto, W. F., & Alencar, S. A. (2023). In silico assessment of missense point mutations on human cathelicidin LL-37. *Journal of Molecular Graphics and Modelling*, *118*, 108368. https://doi.org/https://doi.org/10.1016/j.jmgm.2022.108368

Postler, T. S., Wang, A., Brundu, F. G., Wang, P., Wu, Z., Butler, K. E., Grinberg-Bleyer, Y., Krishnareddy, S., Lagana, S. M., & Saqi, A. (2023). A pan-cancer analysis implicates human NKIRAS1 as a tumor-suppressor gene. *Proceedings of the National Academy of Sciences*, *120*(46), e2312595120.

Potter, S., Eldridge, M. D., & Ho, S. Y. (2023). Molecular evolution in Australasian marsupials. In *American and Australasian Marsupials: An Evolutionary, Biogeographical, and Ecological Approach* (pp. 1-31). Springer.

Price, G., & Farquharson, K. . (2022). *PacBio HiFi genome assembly using hifiasm v2.1. WorkflowHub.* .

Prior, I. A., Lewis, P. D., & Mattos, C. (2012). A comprehensive survey of Ras mutations in cancer. *Cancer research*, *72*(10), 2457-2467.

Promislow, D. E., & Harvey, P. H. (1990). Living fast and dying young: A comparative analysis of life-history variation among mammals. *Journal of Zoology*, *220*(3), 417-437.

Puritz, J. B., Guo, X., Hare, M., He, Y., Hillier, L. W., Jin, S., Liu, M., Lotterhos, K. E., Minx, P., & Modak, T. (2024). A second unveiling: Haplotig masking of the eastern oyster genome improves population-level inference. *Molecular Ecology Resources*, *24*(1), e13801.

Pye, R. J., Darby, J., Flies, A. S., Fox, S., Carver, S., Elmer, J., Swift, K., Hogg, C., Pemberton, D., & Woods, G. (2021). Post-release immune responses of Tasmanian devils vaccinated with an experimental devil facial tumour disease vaccine. *Wildlife Research*, *48*(8), 701-712.

Pye, R. J., Hamede, R., Siddle, H. V., Caldwell, A., Knowles, G. W., Swift, K., Kreiss, A., Jones, M. E., Lyons, A. B., & Woods, G. M. (2016). Demonstration of immune responses against devil facial tumour disease in wild Tasmanian devils. *Biology letters*, *12*(10), 20160553.

Pye, R. J., Patchett, A., McLennan, E., Thomson, R., Carver, S., Fox, S., Pemberton, D., Kreiss, A., Baz Morelli, A., & Silva, A. (2018). Immunization strategies producing a humoral IgG immune response against devil facial tumor disease in the majority of Tasmanian devils destined for wild release. *Frontiers in Immunology*, *9*, 259.

Pye, R. J., Pemberton, D., Tovar, C., Tubio, J. M., Dun, K. A., Fox, S., Darby, J., Hayes, D., Knowles, G. W., & Kreiss, A. (2016). A second transmissible cancer in Tasmanian devils. *Proceedings of the National Academy of Sciences*, *113*(2), 374-379.

Pye, R. J., Woods, G. M., & Kreiss, A. (2016). Devil facial tumor disease. *Veterinary Pathology*, *53*(4), 726-736.

Queirós, J., Alves, P. C., Vicente, J., Gortázar, C., & de la Fuente, J. (2018). Genome-wide associations identify novel candidate loci associated with genetic susceptibility to tuberculosis in wild boar. *Scientific Reports*, *8*(1), 1980.

Quigley, B. L., Carver, S., Hanger, J., Vidgen, M. E., & Timms, P. (2018). The relative contribution of causal factors in the transition from infection to clinical chlamydial disease. *Scientific Reports*, *8*(1), 8893.

Quigley, B. L., & Timms, P. (2020). Helping koalas battle disease–Recent advances in Chlamydia and koala retrovirus (KoRV) disease understanding and treatment in koalas. *FEMS microbiology reviews*, *44*(5), 583-605.

Quigley, B. L., Tzipori, G., Nilsson, K., & Timms, P. (2020). High-throughput immunogenetic typing of koalas suggests possible link between MHC alleles and cancers. *Immunogenetics*, *72*(9), 499-506.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841-842.

Quinlan, M. P., & Settleman, J. (2009). Isoform-specific ras functions in development and cancer. *Future oncology*, *5*(1), 105-116.

Rahman, M. T., Sobur, M. A., Islam, M. S., Ievy, S., Hossain, M. J., El Zowalaty, M. E., Rahman, A. T., & Ashour, H. M. (2020). Zoonotic diseases: etiology, impact, and control. *Microorganisms*, *8*(9), 1405.

Ramanathan, B., Davis, E. G., Ross, C. R., & Blecha, F. (2002). Cathelicidins: microbicidal activity, mechanisms of action, and roles in innate immunity. *Microbes and infection*, *4*(3), 361-372. https://doi.org/https://doi.org/10.1016/S1286-4579(02)01549-6

Rangwala, S. H., Kuznetsov, A., Ananiev, V., Asztalos, A., Borodin, E., Evgeniev, V., Joukov, V., Lotov, V., Pannu, R., & Rudnev, D. (2021). Accessing NCBI data using the NCBI sequence viewer and genome data viewer (GDV). *Genome research*, *31*(1), 159-169.

Ratcliffe, H. L. (1933). Incidence and nature of tumors in captive wild mammals and birds. *The American Journal of Cancer*, *17*(1), 116-135.

Renfree, M. B. (2010). Marsupials: placental mammals with a difference. *Placenta*, *31*, S21-S26.

Renfree, M. B., Papenfuss, A. T., Deakin, J. E., Lindsay, J., Heider, T., Belov, K., Rens, W., Waters, P. D., Pharo, E. A., & Shaw, G. (2011). Genome sequence of an Australian kangaroo, Macropus eugenii, provides insight into the evolution of mammalian reproduction and development. *Genome biology*, *12*, 1-26.

Renfree, M. B., & Shaw, G. (2021). Placentation in marsupials. *Placentation in Mammals: Tribute to EC Amoroso's Lifetime Contributions to Viviparity*, 41-60.

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., & Kim, J. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, *592*(7856), 737-746.

Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S. R., Consortium, W., Wilkie, A. O., McVean, G., & Lunter, G. (2014). Integrating mapping-, assembly-and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature genetics*, *46*(8), 912-918.

Risso, D., Schwartz, K., Sherlock, G., & Dudoit, S. (2011). GC-content normalization for RNA-Seq data. *BMC bioinformatics*, *12*, 1-17.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, *43*(7), e47-e47.

Robbins, A., Hanger, J., Jelocnik, M., Quigley, B. L., & Timms, P. (2020). Koala immunogenetics and chlamydial strain type are more directly involved in chlamydial disease progression in koalas from two south east Queensland koala populations than koala retrovirus subtypes. *Scientific Reports*, *10*(1), 15013.

Robbins, A., Loader, J., Timms, P., & Hanger, J. (2018). Optimising the short and long-term clinical outcomes for koalas (Phascolarctos cinereus) during treatment for chlamydial infection and disease. *PloS one*, *13*(12), e0209679.

Robertson, A. G., Groeneveld, C. S., Jordan, B., Lin, X., McLaughlin, K. A., Das, A., Fall, L. A., Fantini, D., Taxter, T. J., & Mogil, L. S. (2020). Identification of differential tumor subtypes of T1 bladder cancer. *European urology*, *78*(4), 533-537.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature biotechnology*, *29*(1), 24-26.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139-140.

Rodrigues, C. H., Garg, A., Keizer, D., Pires, D. E., & Ascher, D. B. (2022). CSM-peptides: A computational approach to rapid identification of therapeutic peptides. *Protein Science*, *31*(10), e4442.

Roperto, S., Russo, V., Urraro, C., Restucci, B., Corrado, F., De Falco, F., & Roperto, F. (2017). ERas is constitutively expressed in full term placenta of pregnant cows. *Theriogenology*, *103*, 162-168.

Ross, C. A. (2015). The trophoblast model of cancer. *Nutrition and Cancer*, *67*(1), 61-67.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53-65.

Roycroft, E., MacDonald, A. J., Moritz, C., Moussalli, A., Portela Miguez, R., & Rowe, K. C. (2021). Museum genomics reveals the rapid decline and extinction of Australian rodents since European settlement. *Proceedings of the National Academy of Sciences*, *118*(27), e2021390118.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, *34*(12), 3299-3302.

Ruprecht, K., Mayer, J., Sauter, M., Roemer, K., & Mueller-Lantzsch, N. (2008). Endogenous retroviruses: endogenous retroviruses and cancer. *Cellular and Molecular Life Sciences*, *65*, 3366-3382.

Russell, R. E., DiRenzo, G. V., Szymanski, J. A., Alger, K. E., & Grant, E. H. (2020). Principles and mechanisms of wildlife population persistence in the face of disease. *Frontiers in Ecology and Evolution*, *8*, 569016.

Sanderson, M. J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, *19*(2), 301-302.

Santamarina, M., Bruzos, A. L., Pequeño, A., Rodríguez-Castro, J., Díaz, S., & Tubio, J. (2024). Novel PCR assay for the identification of two transmissible cancers in Cerastoderma edule. *Journal of Invertebrate Pathology*, 108232.

Santra, M. K., Wajapeyee, N., & Green, M. R. (2009). F-box protein FBXO31 mediates cyclin D1 degradation to induce G1 arrest after DNA damage. *Nature*, *459*(7247), 722-725.

Sathoff, A. E., & Samac, D. A. (2019). Antibacterial activity of plant defensins. *Molecular Plant-Microbe Interactions*, *32*(5), 507-514.

Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I. (2020). GenBank. *Nucleic acids research*, *48*(D1), D84-D86.

Scheele, B. C., Pasmans, F., Skerratt, L. F., Berger, L., Martel, A., Beukema, W., Acevedo, A. A., Burrowes, P. A., Carvalho, T., & Catenazzi, A. (2019). Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity. *Science*, *363*(6434), 1459-1463.

Schmitt, C., Garant, D., Doyon, K., Bousquet, N., Gaudreau, L., Bélisle, M., & Pelletier, F. (2017). Patterns of diversity and spatial variability of β -defensin innate immune genes in a declining wild population of tree swallows. *Journal of Heredity*, *108*(3), 262-269.

Schneider, S., Vincek, V., Tichy, H., Figueroa, F., & Klein, J. (1991). MHC class II genes of a marsupial, the red-necked wallaby (Macropus rufogriseus): identification of new gene families. *Molecular Biology and Evolution*, *8*(6), 753-766.

Schneider, V. A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.-C., Kitts, P. A., Murphy, T. D., Pruitt, K. D., Thibaud-Nissen, F., & Albracht, D. (2017). Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome research*, *27*(5), 849-864.

Schultz, J. A., Cloutier, R. N., & Côté, I. M. (2016). Evidence for a trophic cascade on rocky reefs following sea star mass mortality in British Columbia. *PeerJ*, *4*, e1980.

Scocchi, M., Skerlavaj, B., Romeo, D., & Gennaro, R. (1992). Proteolytic cleavage by neutrophil elastase converts inactive storage proforms to antibacterial bactenecins. *European journal of biochemistry*, *209*(2), 589-595.

Seim, I., Ma, S., Zhou, X., Gerashchenko, M. V., Lee, S.-G., Suydam, R., George, J. C., Bickham, J. W., & Gladyshev, V. N. (2014). The transcriptome of the bowhead whale Balaena mysticetus reveals adaptations of the longest-lived mammal. *Aging (Albany NY)*, *6*(10), 879.

Sekiguchi, M., Seki, M., Kawai, T., Yoshida, K., Yoshida, M., Isobe, T., Hoshino, N., Shirai, R., Tanaka, M., & Souzaki, R. (2020). Integrated multiomics analysis of hepatoblastoma unravels its heterogeneity and provides novel druggable targets. *NPJ precision oncology*, *4*(1), 20.

Selsted, M. E., & Ouellette, A. J. (2005). Mammalian defensins in the antimicrobial immune response. *Nature immunology*, *6*(6), 551-557. https://doi.org/doi:10.1038/ni1206

Seluanov, A., Gladyshev, V. N., Vijg, J., & Gorbunova, V. (2018). Mechanisms of cancer resistance in long-lived mammals. *Nature reviews cancer*, *18*(7), 433-441.

Seluanov, A., Hine, C., Azpurua, J., Feigenson, M., Bozzella, M., Mao, Z., Catania, K. C., & Gorbunova, V. (2009). Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat. *Proceedings of the National Academy of Sciences*, *106*(46), 19352-19357.

Selwood, L., & Woolley, P. (1991). A timetable of embryonic development, and ovarian and uterine changes during pregnancy, in the stripe-faced dunnart, Sminthopsis macroura (Marsupialia: Dasyuridae). *Reproduction*, *91*(1), 213-227.

Semple, C. A., Gautier, P., Taylor, K., & Dorin, J. R. (2006). The changing of the guard: Molecular diversity and rapid evolution of β -defensins. *Molecular diversity*, *10*, 575-584.

Semple, F., & Dorin, J. R. (2012). β -Defensins: multifunctional modulators of infection, inflammation and more? *Journal of innate immunity*, *4*(4), 337-348.

Shafee, T. M., Lay, F. T., Phan, T. K., Anderson, M. A., & Hulett, M. D. (2017). Convergent evolution of defensin sequence, structure and function. *Cellular and Molecular Life Sciences*, *74*, 663-682.

Shipley, L. A., Forbey, J. S., & Moore, B. D. (2009). Revisiting the dietary niche: when is a mammalian herbivore a specialist? *Integrative and Comparative Biology*, *49*(3), 274-290.

Siddle, H. V. (2017). Cancer as a contagious disease. *HLA*, *89*(4), 209-214.

Siddle, H. V., Kreiss, A., Eldridge, M. D., Noonan, E., Clarke, C. J., Pyecroft, S., Woods, G. M., & Belov, K. (2007). Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. *Proceedings of the National Academy of Sciences*, *104*(41), 16221-16226.

Siddle, H. V., Kreiss, A., Tovar, C., Yuen, C. K., Cheng, Y., Belov, K., Swift, K., Pearse, A.-M., Hamede, R., & Jones, M. E. (2013). Reversible epigenetic down-regulation of MHC molecules by devil facial tumour disease illustrates immune escape by a contagious cancer. *Proceedings of the National Academy of Sciences*, *110*(13), 5103-5108.

Siddle, H. V., Marzec, J., Cheng, Y., Jones, M., & Belov, K. (2010). MHC gene copy number variation in Tasmanian devils: implications for the spread of a contagious cancer. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1690), 2001-2006.

Siegal-Willott, J., Heard, D., Sliess, N., Naydan, D., & Roberts, J. (2007). Microchip-associated leiomyosarcoma in an Egyptian fruit bat (Rousettus aegyptiacus). *Journal of Zoo and Wildlife Medicine*, *38*(2), 352-356.

Silva, P. M., Gonçalves, S., & Santos, N. C. (2014). Defensins: antifungal lessons from eukaryotes. *Frontiers in Microbiology*, *5*, 97.

Silver, L. W., Cheng, Y., Quigley, B. L., Robbins, A., Timms, P., Hogg, C. J., & Belov, K. (2022). A targeted approach to investigating immune genes of an iconic Australian marsupial. *Molecular Ecology*, *31*(12), 3286-3303.

Silver, L. W., Hogg, C. J., & Belov, K. (2024). Plethora of new marsupial genomes informs our knowledge of marsupial MHC class II. *Genome biology and evolution*, *16*(8), evae156.

Silver, L. W., McLennan, E. A., Beaman, J., da Silva, K. B., Timms, P., Hogg, C. J., & Belov, K. (2024). Using bioinformatics to investigate functional diversity: a case study of MHC diversity in koalas. *Immunogenetics*, 1-15.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210-3212.

Simmons, G., Young, P., Hanger, J., Jones, K., Clarke, D., McKee, J., & Meers, J. (2012). Prevalence of koala retrovirus in geographically diverse populations in Australia. *Australian veterinary journal*, *90*(10), 404-409. https://doi.org/https://doi.org/10.1111/j.1751-0813.2012.00964.x

Simpson, A. J., Caballero, O. L., Jungbluth, A., Chen, Y.-T., & Old, L. J. (2005). Cancer/testis antigens, gametogenesis and cancer. *Nature reviews cancer*, *5*(8), 615-625.

Slate, D., Algeo, T. P., Nelson, K. M., Chipman, R. B., Donovan, D., Blanton, J. D., Niezgoda, M., & Rupprecht, C. E. (2009). Oral rabies vaccination in North America: opportunities, complexities, and challenges. *PLoS neglected tropical diseases*, *3*(12), e549.

Smit, A., Hubley, R., & Green, P. (2013-2015). *RepeatMasker Open-4.0*. In

Smith, K. R., Hanson, H. A., Mineau, G. P., & Buys, S. S. (2012). Effects of BRCA1 and BRCA2 mutations on female fertility. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1732), 1389-1395.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B., & Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature*, *321*(6071), 674-679.

Solovyev, V., Kosarev, P., Seledsov, I., & Vorobyev, D. (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome biology*, *7*, 1-12.

Sondka, Z., Dhir, N. B., Carvalho-Silva, D., Jupe, S., Madhumita, n., McLaren, K., Starkey, M., Ward, S., Wilding, J., & Ahmed, M. (2024). COSMIC: a curated

database of somatic variants and clinical data for cancer. *Nucleic acids research*, *52*(D1), D1210-D1217.

Sørensen, O. E., Follin, P., Johnsen, A. H., Calafat, J., Tjabringa, G. S., Hiemstra, P. S., & Borregaard, N. (2001). Human cathelicidin, hCAP-18, is processed to the antimicrobial peptide LL-37 by extracellular cleavage with proteinase 3. *Blood, The Journal of the American Society of Hematology*, *97*(12), 3951-3959.

Spandidos, D. A., & Anderson, M. L. (1989). Oncogenes and onco-suppressor genes: their involvement in cancer. *The Journal of pathology*, *157*(1), 1-10.

Spurgin, L. G., & Richardson, D. S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1684), 979-988. https://doi.org/doi:10.1098/rspb.2009.2084

Stammnitz, M. R., Coorens, T. H., Gori, K. C., Hayes, D., Fu, B., Wang, J., Martin-Herranz, D. E., Alexandrov, L. B., Baez-Ortega, A., & Barthorpe, S. (2018). The origins and vulnerabilities of two transmissible cancers in Tasmanian devils. *Cancer Cell*, *33*(4), 607-619. e615.

Stammnitz, M. R., Gori, K., Kwon, Y. M., Harry, E., Martin, F. J., Billis, K., Cheng, Y., Baez-Ortega, A., Chow, W., & Comte, S. (2022). The evolution of two transmissible cancers in Tasmanian devils. *bioRxiv*.

Stammnitz, M. R., Gori, K., Kwon, Y. M., Harry, E., Martin, F. J., Billis, K., Cheng, Y., Baez-Ortega, A., Chow, W., & Comte, S. (2023). The evolution of two transmissible cancers in Tasmanian devils. *Science*, *380*(6642), 283-293.

Stammnitz, M. R., Gori, K., & Murchison, E. P. (2024). No evidence that a transmissible cancer has shifted from emergence to endemism in Tasmanian devils. *Royal Society Open Science*, *11*(4), 231875.

Stearns, S. C. (1976). Life-history tactics: a review of the ideas. *The Quarterly review of biology*, *51*(1), 3-47.

Strakova, A. (2017). *Genome Diversity and Evolution in Canine Transmissible Venereal Tumour* University of Cambridge].

Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature*, *458*(7239), 719-724.

Straube, E. F., & Callinan, R. B. (1980). Cutaneous squamous cell carcinoma associated with mammary adenocarcinoma in an eastern quoll Dasyurus viverrinus. *Journal of Comparative Pathology*, *90*(3), 495-497. https://doi.org/10.1016/0021-9975(80)90020-1

Strom, S. P. (2016). Current practices and guidelines for clinical next-generation sequencing oncology testing. *Cancer biology &medicine*, *13*(1), 3.

Suarez-Cabrera, C., Ojeda-Perez, I., Sanchez-Baltasar, R., Page, A., Bravo, A., Navarro, M., & Ramirez, A. (2021). ERAS, a Member of the Ras Superfamily, Acts as an Oncoprotein in the Mammary Gland. *Cancers*, *13*(21), 5588.

Sulak, M., Fong, L., Mika, K., Chigurupati, S., Yon, L., Mongan, N. P., Emes, R. D., & Lynch, V. J. (2016). TP53 copy number expansion is associated with the

evolution of increased body size and an enhanced DNA damage response in elephants. *Elife*, *5*, e11994.

Sura, R., French, R., Goldman, B., & Schwartz, D. (2011). Neoplasia and granulomas surrounding microchip transponders in Damaraland mole rats (Cryptomys damarensis). *Veterinary Pathology*, *48*(4), 896-902.

Suzuki, R., & Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, *22*(12), 1540-1542.

Takahashi, K., Mitsui, K., & Yamanaka, S. (2003). Role of ERas in promoting tumour-like properties in mouse embryonic stem cells. *Nature*, *423*(6939), 541-545.

Tan, I. B., Ivanova, T., Lim, K. H., Ong, C. W., Deng, N., Lee, J., Tan, S. H., Wu, J., Lee, M. H., & Ooi, C. H. (2011). Intrinsic subtypes of gastric cancer, based on gene expression pattern, predict survival and respond differently to chemotherapy. *Gastroenterology*, *141*(2), 476-485. e411.

Tanaka, Y., Ikeda, T., Kishi, Y., Masuda, S., Shibata, H., Takeuchi, K., Komura, M., Iwanaka, T., Muramatsu, S.-i., & Kondo, Y. (2009). ERas is expressed in primate embryonic stem cells but not related to tumorigenesis. *Cell Transplantation*, *18*(4), 381-389.

Tang, S., Peel, E., Belov, K., Hogg, C. J., & Farquharson, K. A. (2024). Multi-omics resources for the Australian southern stuttering frog (Mixophyes australis) reveal assorted antimicrobial peptides. *Scientific Reports*, *14*(1), 3991.

Tarlinton, R., & Greenwood, A. D. (2024). Koala retrovirus and neoplasia: correlation and underlying mechanisms. *Current Opinion in Virology*, *67*, 101427.

Tarlinton, R., Meers, J., Hanger, J., & Young, P. (2005). Real-time reverse transcriptase PCR for the endogenous koala retrovirus reveals an association between plasma viral load and neoplastic disease in koalas. *Journal of General Virology*, *86*(3), 783-787.

Taylor, R. L., Zhang, Y., Schöning, J. P., & Deakin, J. E. (2017). Identification of candidate genes for devil facial tumour disease tumourigenesis. *Scientific Reports*, *7*(1), 8761.

Team, R. C. (2022). *R: A language and environment for statistical computing*. In https://www.R-project.org/

Teeling, E. C., Vernes, S. C., Dávalos, L. M., Ray, D. A., Gilbert, M. T. P., Myers, E., & Consortium, B. K. (2018). Bat biology, genomes, and the Bat1K project: to generate chromosome-level genomes for all living bat species. *Annual review of animal biosciences*, *6*(1), 23-46.

Thakur, M., Buniello, A., Brooksbank, C., Gurwitz, K. T., Hall, M., Hartley, M., Hulcoop, D. G., Leach, A. R., Marques, D., & Martin, M. (2024). EMBL's European bioinformatics institute (EMBL-EBI) in 2023. *Nucleic acids research*, *52*(D1), D10-D17.

Thies, K. A., Cole, M. W., Schafer, R. E., Spehar, J. M., Richardson, D. S., Steck, S. A., Das, M., Lian, A. W., Ray, A., & Shakya, R. (2021). The small G-protein RalA promotes progression and metastasis of triple-negative breast cancer. *Breast cancer research*, *23*(1), 65.

Thorne, E. T., & Williams, E. S. (1988). Disease and endangered species: the black-footed ferret as a recent example. *Conservation Biology*, *2*(1), 66-74.

Tian, R., Han, K., Geng, Y., Yang, C., Shi, C., Thomas, P. B., Pearce, C., Moffatt, K., Ma, S., & Xu, S. (2022). A chromosome-level genome of Antechinus flavipes provides a reference for an Australian marsupial genus with male death after mating. *Molecular Ecology Resources*, *22*(2), 740-754.

Tian, X., Azpurua, J., Hine, C., Vaidya, A., Myakishev-Rempel, M., Ablaeva, J., Mao, Z., Nevo, E., Gorbunova, V., & Seluanov, A. (2013). High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Nature*, *499*(7458), 346-349.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411-423.

Tollis, M., Robbins, J., Webb, A. E., Kuderna, L. F., Caulin, A. F., Garcia, J. D., Bèrubè, M., Pourmand, N., Marques-Bonet, T., & O'Connell, M. J. (2019). Return to the sea, get huge, beat cancer: an analysis of cetacean genomes including an assembly for the humpback whale (Megaptera novaeangliae). *Molecular Biology and Evolution*, *36*(8), 1746-1763.

Torgovnick, A., & Schumacher, B. (2015). DNA repair mechanisms in cancer development and therapy. *Frontiers in genetics*, *6*, 157.

Torres, A. M., & Kuchel, P. W. (2004). The β -defensin-fold family of polypeptides. *Toxicon*, *44*(6), 581-588.

Tørresen, O. K., Star, B., Mier, P., Andrade-Navarro, M. A., Bateman, A., Jarnot, P., Gruca, A., Grynberg, M., Kajava, A. V., & Promponas, V. J. (2019). Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic acids research*, *47*(21), 10994-11006.

Tovar, C., Pye, R. J., Kreiss, A., Cheng, Y., Brown, G. K., Darby, J., Malley, R. C., Siddle, H. V., Skjødt, K., & Kaufman, J. (2017). Regression of devil facial tumour disease following immunotherapy in immunised Tasmanian devils. *Scientific Reports*, *7*(1), 43827.

Tripathi, S., Wang, G., White, M., Qi, L., Taubenberger, J., & Hartshorn, K. L. (2015). Antiviral activity of the human cathelicidin, LL-37, and derived peptides on seasonal and pandemic influenza A viruses. *PloS one*, *10*(4), e0124706.

Trowsdale, J., & Parham, P. (2004). Mini-review: defense strategies and immunity-related genes. *European journal of immunology*, *34*(1), 7-17. https://doi.org/https://doi.org/10.1002/eji.200324693

Twin, J. E., & Pearse, A. M. (1986). A malignant mixed salivary tumour and a mammary carcinoma in a young wild eastern spotted native cat Dasyurus viverrinus (marsupialia). *Journal of Comparative Pathology*, *96*(3), 301-306. https://doi.org/10.1016/0021-9975(86)90050-2

Tyndale-Biscoe, C. H., & Renfree, M. (1987). *Reproductive physiology of marsupials*. Cambridge University Press.

Ujvari, B., Pearse, A.-M., Peck, S., Harmsen, C., Taylor, R., Pyecroft, S., Madsen, T., Papenfuss, A. T., & Belov, K. (2013). Evolution of a contagious cancer: epigenetic variation in Devil Facial Tumour Disease. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1750), 20121720.

Ujvari, B., Pearse, A. M., Swift, K., Hodson, P., Hua, B., Pyecroft, S., Taylor, R., Hamede, R., Jones, M., & Belov, K. (2014). Anthropogenic selection enhances cancer evolution in T asmanian devil tumours. *Evolutionary applications*, *7*(2), 260-265.

Van Bressem, M.-F., Duignan, P. J., Banyard, A., Barbieri, M., Colegrove, K. M., De Guise, S., Di Guardo, G., Dobson, A., Domingo, M., & Fauquier, D. (2014). Cetacean morbillivirus: current knowledge and future directions. *Viruses*, *6*(12), 5145-5181.

van der Kraan, L. E., Wong, E. S., Lo, N., Ujvari, B., & Belov, K. (2013). Identification of natural killer cell receptor genes in the genome of the marsupial Tasmanian devil (Sarcophilus harrisii). *Immunogenetics*, *65*, 25-35.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, *9*(11).

Van Harten, R. M., Van Woudenbergh, E., Van Dijk, A., & Haagsman, H. P. (2018). Cathelicidins: immunomodulatory antimicrobials. *Vaccines*, *6*(3), 63.

Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science*, *324*(5930), 1029-1033.

VanderWaal, K. L., Atwill, E. R., Isbell, L. A., & McCowan, B. (2014). Quantifying microbe transmission networks for wild and domestic ungulates in Kenya. *Biological Conservation*, *169*, 136-146.

Venturutti, L., Romero, L. V., Urtreger, A. J., Chervo, M. F., Cordo Russo, R. I., Mercogliano, M. F., Inurrigarro, G., Pereyra, M., Proietti, C., & Izzo, F. (2016). Stat3 regulates ErbB-2 expression and co-opts ErbB-2 nuclear function to induce miR-21 expression, PDCD4 downregulation and breast cancer metastasis. *Oncogene*, *35*(17), 2208-2222.

Vincze, O., Colchero, F., Lemaître, J.-F., Conde, D. A., Pavard, S., Bieuville, M., Urrutia, A. O., Ujvari, B., Boddy, A. M., & Maley, C. C. (2022). Cancer risk across mammals. *Nature*, *601*(7892), 263-267.

Vinkler, M., Fiddaman, S. R., Těšický, M., O'Connor, E. A., Savage, A. E., Lenz, T. L., Smith, A. L., Kaufman, J., Bolnick, D. I., & Davies, C. S. (2023). Understanding the evolution of immune genes in jawed vertebrates. *Journal of evolutionary biology*, *36*(6), 847-873. https://doi.org/https://doi.org/10.1111/jeb.14181

Vittecoq, M., Roche, B., Daoust, S. P., Ducasse, H., Missé, D., Abadie, J., Labrut, S., Renaud, F., Gauthier-Clerc, M., & Thomas, F. (2013). Cancer: a missing link in ecosystem functioning? *Trends in ecology &evolution*, *28*(11), 628-635.

Wang, D., Dai, J., Suo, C., Wang, S., Zhang, Y., & Chen, X. (2022). Molecular subtyping of esophageal squamous cell carcinoma by large-scale transcriptional

profiling: Characterization, therapeutic targets, and prognostic value. *Frontiers in genetics*, *13*, 1033214.

Wang, G., Vaisman, I. I., & van Hoek, M. L. (2022). Machine learning prediction of antimicrobial peptides. In *Computational Peptide Science: Methods and Protocols* (pp. 1-37). Springer.

Wang, G., Watson, K. M., Peterkofsky, A., & Buckheit Jr, R. W. (2010). Identification of novel human immunodeficiency virus type 1-inhibitory peptides based on the antimicrobial peptide database. *Antimicrobial agents and chemotherapy*, *54*(3), 1343-1346. https://doi.org/https://doi.org/10.1128/aac.01448-09

Wang GuangShun, W. G. (2017). Discovery, classification and functional diversity of antimicrobial peptides. In *Antimicrobial peptides: discovery, design and novel therapeutic strategies* (pp. 1-19). CABI Wallingford UK.

Wang, J., Dou, X., Song, J., Lyu, Y., Zhu, X., Xu, L., Li, W., & Shan, A. (2019). Antimicrobial peptides: Promising alternatives in the post feeding antibiotic era. *Medicinal research reviews*, *39*(3), 831-859. https://doi.org/https://doi.org/10.1002/med.21542

Wang, J., Wong, E. S., Whitley, J. C., Li, J., Stringer, J. M., Short, K. R., Renfree, M. B., Belov, K., & Cocks, B. G. (2011). Ancient antimicrobial peptides kill antibiotic-resistant pathogens: Australian mammals provide new options. *PloS one*, *6*(8), e24030. https://doi.org/doi:10.1371/journal.pone.0024030

Wang, X., Douglas, K. C., VandeBerg, J. L., Clark, A. G., & Samollow, P. B. (2014). Chromosome-wide profiling of X-chromosome inactivation and epigenetic states in fetal brain and placenta of the opossum, Monodelphis domestica. *Genome research*, *24*(1), 70-83.

Wangsa, D., Braun, R., Stuelten, C. H., Brown, M., Bauer, K. M., Emons, G., Weston, L. A., Hu, Y., Yang, H. H., & Vila-Casadesús, M. (2019). Induced chromosomal aneuploidy results in global and consistent deregulation of the transcriptome of cancer cells. *Neoplasia*, *21*(7), 721-729.

Ward, M., Carwardine, J., Yong, C. J., Watson, J. E., Silcock, J., Taylor, G. S., Lintermans, M., Gillespie, G. R., Garnett, S. T., & Woinarski, J. (2021). A national-scale dataset for threats impacting Australia's imperiled flora and fauna. *Ecology and Evolution*, *11*(17), 11749-11761.

Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, *171*(4356), 737-738.

Waugh, C., Khan, S. A., Carver, S., Hanger, J., Loader, J., Polkinghorne, A., Beagley, K., & Timms, P. (2016). A prototype recombinant-protein based Chlamydia pecorum vaccine results in reduced chlamydial burden and less clinical disease in free-ranging koalas (Phascolarctos cinereus). *PloS one*, *11*(1), e0146934.

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Molecular Biology and Evolution*, *35*(3), 773-777. https://doi.org/https://doi.org/10.1093/molbev/msx335

Weiss, S., Taggart, D., Smith, I., Helgen, K. M., & Eisenhofer, R. (2021). Host reproductive cycle influences the pouch microbiota of wild southern hairy-nosed wombats (Lasiorhinus latifrons). *Animal Microbiome*, *3*, 1-14. https://doi.org/DOI:10.1186/s42523-021-00074-8

Westerman, M., Krajewski, C., Kear, B. P., Meehan, L., Meredith, R. W., Emerling, C. A., & Springer, M. S. (2016). Phylogenetic relationships of dasyuromorphian marsupials revisited. *Zoological Journal of the Linnean Society*, *176*(3), 686-701.

Wetterstrand, K. A. The Cost of Sequencing a Human Genome—NHGRI. 2020. *URl: https://www. genome. gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost*.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Williams, G. C. (2001). Pleiotropy, Natural Selection, and the Evolution of Senescence: Evolution 11, 398-411 (1957). *Science of Aging Knowledge Environment*, *2001*(1), cp13-cp13.

Wiseman, H., & Halliwell, B. (1996). Damage to DNA by reactive oxygen and nitrogen species: role in inflammatory disease and progression to cancer. *Biochemical Journal*, *313*(Pt 1), 17.

Wittbrodt, J., Adam, D., Malitschek, B., Mäueler, W., Raulf, F., Telling, A., Robertson, S. M., & Schartl, M. (1989). Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu locus in Xiphophorus. *Nature*, *341*(6241), 415-421.

Wong, E. S., Papenfuss, A. T., Heger, A., Hsu, A. L., Ponting, C. P., Miller, R. D., Fenelon, J. C., Renfree, M. B., Gibbs, R. A., & Belov, K. (2011). Transcriptomic analysis supports similar functional roles for the two thymuses of the tammar wallaby. *BMC genomics*, *12*, 1-12.

Wood, R. D., Mitchell, M., Sgouros, J., & Lindahl, T. (2001). Human DNA repair genes. *Science*, *291*(5507), 1284-1289.

Woolley, P. (1971). Maintenance and breeding of laboratory colonies. *International Zoo Yearbook*, *11*(1), 351-354.

Woolley, P. (1988). Reproduction in the Ningbing Antechinus (Marsupialia, Dasyuridae)-Field and Laboratory Observations. *Wildlife Research*, *15*(2), 149-156.

Wright, B., Willet, C. E., Hamede, R., Jones, M., Belov, K., & Wade, C. M. (2017). Variants in the host genome may inhibit tumour growth in devil facial tumours: evidence from genome-wide association. *Scientific Reports*, *7*(1), 423.

Wu, C., Duan, Y., Gong, S., Osterhoff, G., Kallendrusch, S., & Schopow, N. (2022). Identification of Tumor Antigens and Immune Subtypes for the Development of mRNA Vaccines and Individualized Immunotherapy in Soft Tissue Sarcoma. *Cancers*, *14*(2), 448.

Wu, C., Qin, C., Long, W., Wang, X., Xiao, K., & Liu, Q. (2022). Tumor antigens and immune subtypes of glioblastoma: the fundamentals of mRNA vaccine and individualized immunotherapy development. *Journal of big Data*, *9*(1), 1-25.

Xiao, Y., Hughes, A. L., Ando, J., Matsuda, Y., Cheng, J.-F., Skinner-Noble, D., & Zhang, G. (2004). A genome-wide screen identifies a single β-defensin gene cluster in the chicken: implications for the origin and evolution of mammalian defensins. *BMC genomics*, *5*, 1-11.

Xie, T.-x., Wei, D., Liu, M., Gao, A. C., Ali-Osman, F., Sawaya, R., & Huang, S. (2004). Stat3 activation regulates the expression of matrix metalloproteinase-2 and tumor invasion and metastasis. *Oncogene*, *23*(20), 3550-3560.

Xiong, Y., Wang, K., Zhou, H., Peng, L., You, W., & Fu, Z. (2018). Profiles of immune infiltration in colorectal cancer and their clinical significant: A gene expression-based study. *Cancer medicine*, *7*(9), 4496-4508.

Yadav, M. (1973). The presence of the cervical and thoracic thymus lobes in marsupials. *Australian Journal of Zoology*, *21*(3), 285-301.

Yadav, M., Stanley, N., & Waring, H. (1972). The thymus glands of a marsupial, Setonix brachyurus (quokka), and their role in immune responses: structure and growth of the thymus glands. *Australian Journal of Experimental Biology and Medical Science*, *50*(3), 347-356.

Yang, D., Chen, Q., Schmidt, A. P., Anderson, G. M., Wang, J. M., Wooters, J., Oppenheim, J. J., & Chertov, O. (2000). LL-37, the neutrophil granule–and epithelial cell–derived cathelicidin, utilizes formyl peptide receptor–like 1 (FPRL1) as a receptor to chemoattract human peripheral blood neutrophils, monocytes, and T cells. *The Journal of experimental medicine*, *192*(7), 1069-1074.

Yonemitsu, M. A., Sevigny, J. K., Vandepas, L. E., Dimond, J. L., Giersch, R. M., Gurney-Smith, H. J., Abbott, C. L., Supernault, J., Withler, R., & Smith, P. D. (2023). Multiple lineages of transmissible neoplasia in the basket cockle (C. nuttallii) with repeated horizontal transfer of mitochondrial DNA. *bioRxiv*, 2023.2010. 2011.561945.

Yue, X., Chang, T., DeJarnette, J., Marshall, C., Lei, C., & Liu, W.-S. (2013). Copy number variation of PRAMEY across breeds and its association with male fertility in Holstein sires. *Journal of dairy science*, *96*(12), 8024-8034. https://doi.org/https://doi.org/10.3168/jds.2013-7037

Zanetti, M. (2005). The role of cathelicidins in the innate host defenses of mammals. *Current issues in molecular biology*, *7*(2), 179-196.

Zanetti, M., Del Sal, G., Storici, P., Schneider, C., & Romeo, D. (1993). The cDNA of the neutrophil antibiotic Bac5 predicts a pro-sequence homologous to a cysteine proteinase inhibitor that is common to other neutrophil antibiotics. *Journal of Biological Chemistry*, *268*(1), 522-526.

Zanetti, M., Gennaro, R., & Romeo, D. (1995). Cathelicidins: a novel protein family with a common proregion and a variable C-terminal antimicrobial domain. *FEBS letters*, *374*(1), 1-5.

Zasloff, M. (2002). Antimicrobial peptides of multicellular organisms. *Nature*, *415*(6870), 389-395.

Zhang, F., Gu, W., Hurles, M. E., & Lupski, J. R. (2009). Copy number variation in human health, disease, and evolution. *Annual Review of Genomics and Human*

Genetics, *10*, 451-481. https://doi.org/https://doi.org/10.1146%2Fannurev.genom.9.081307.164217

Zhang, G., Cowled, C., Shi, Z., Huang, Z., Bishop-Lilly, K. A., Fang, X., Wynne, J. W., Xiong, Z., Baker, M. L., & Zhao, W. (2013). Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science*, *339*(6118), 456-460.

Zhang, H., Wang, Y., Zhu, Y., Huang, P., Gao, Q., Li, X., Chen, Z., Liu, Y., Jiang, J., & Gao, Y. (2025). Machine learning and genetic algorithm-guided directed evolution for the development of antimicrobial peptides. *Journal of Advanced Research*, *68*, 415-428.

Zhang, J., Yang, L., Tian, Z., Zhao, W., Sun, C., Zhu, L., Huang, M., Guo, G., & Liang, G. (2021). Large-scale screening of antifungal peptides based on quantitative structure–activity Relationship. *ACS Medicinal Chemistry Letters*, *13*(1), 99-104. https://doi.org/DOI: 10.1021/acsmedchemlett.1c00556

Zhang, L., Dong, X., Tian, X., Lee, M., Ablaeva, J., Firsanov, D., Lee, S.-G., Maslov, A. Y., Gladyshev, V. N., & Seluanov, A. (2021). Maintenance of genome sequence integrity in long-and short-lived rodent species. *Science advances*, *7*(44), eabj3284.

Zhang, X., Wang, K., Wang, L., Yang, Y., Ni, Z., Xie, X., Shao, X., Han, J., Wan, D., & Qiu, Q. (2016). Genome-wide patterns of copy number variation in the Chinese yak genome. *BMC genomics*, *17*, 1-12.

Zhang, Y., & Li, G. (2020). A tumor suppressor DLC1: The functions and signal pathways. *Journal of cellular physiology*, *235*(6), 4999-5007.

Zhao, H., Gan, T.-X., Liu, X.-D., Jin, Y., Lee, W.-H., Shen, J.-H., & Zhang, Y. (2008). Identification and characterization of novel reptile cathelicidins from elapid snakes. *Peptides*, *29*(10), 1685-1691.

Zhao, Z., Zhao, J., Song, K., Hussain, A., Du, Q., Dong, Y., Liu, J., & Yang, X. (2020). Joint DBN and Fuzzy C-Means unsupervised deep clustering for lung cancer patient stratification. *Engineering Applications of Artificial Intelligence*, *91*, 103571.

Zhou, C., McCarthy, S. A., & Durbin, R. (2023). YaHS: yet another Hi-C scaffolding tool. *Bioinformatics*, *39*(1), btac808.

Ziklo, N., Huston, W. M., Hocking, J. S., & Timms, P. (2016). Chlamydia trachomatis genital tract infections: when host immune response and the microbiome collide. *Trends in microbiology*, *24*(9), 750-765.

Wangsa, D., et al., *Induced chromosomal aneuploidy results in global and consistent deregulation of the transcriptome of cancer cells.* Neoplasia, 2019. **21**(7): p. 721-729.

APPENDIX 1. SUPPLEMENTARY MATERIAL TO CHAPTER 2

1. Genome Assemblies

The kowari genome was 3.21 GB and consisted of 851 scaffolds and 2,402 contigs. The scaffold N50 was 620 MB and the contig N50 was 4 MB. Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.8.0 (Simão et al., 2015) identified 95.6% mammalian genes, of which 94.5% were single copy, 1.2% were duplicated, 1.0% were fragmented and 3.4% were missing.

The bandicoot genome was 3.94 GB and consisted of 1,462 scaffolds and 4,595 contigs. The scaffold N50 was 722 MB and the contig N50 was 2 MB. BUSCO v 5.8.0 identified 89.5% mammalian genes, of which 87.1% were single copy and 2.5% were duplicated. 1.5% were fragmented and 8.9% were missing.

We could not identify the SRY gene in either the kowari or the bandicoot. In addition, genome coverage appeared to be largely equal across the entire genome for both species, suggesting both samples came from a homogametic individual (**Figure A1-1**). Contact maps for both species indicated 7 large scaffolds, corresponding with the 7 pairs of chromosomes for both orders (**Figure A1-2**).

2. Genome Annotations

BUSCO v5.8.0 (Simão et al., 2015) scores were calculated on the Galaxy Australia webserver (https://usegalaxy.org.au/) using the mammalia_odb10 database.

Repetitive regions of the genomes were masked using a 256 GB RAM, 64 vCPU, 3TB Pawsey Supercomputing Centre Nimbus cloud machine for the nine published genomes and the Galaxy webserver for the bandicoot and kowari. First, RepeatModeler v2.0.1

(Flynn et al., 2020) was used to identify and create a database of repetitive regions, then RepeatMasker v4.0.6 (Smit et al., 2013-2015) was used to mask them with the -nolow parameter to avoid masking low complexity regions.

A global assembly of all available tissue transcriptomes for each species was generated to guide annotation. Bioproject information for the raw data used for each genome is included in **Table A1-2**. Raw reads were quality checked using FastQC v0.11.8 (Andrews, 2010) then trimmed using Trimmomatic v 0.39 (Bolger et al., 2014) with the parameters ILLUMINACLIP:TruSeq3-PE.fa:2:30:10, SLIDINGWINDOW:4:5, LEADING:5, TRAILING:5, MINLEN:25. Reads were aligned to the repeat masked genome using hisat2 v2.1.0 (Kim et al., 2019) and StringTie v2.1.6 (Pertea et al., 2015) was used to merge the aligned reads into tissue-specific transcriptomes. The transcriptomes were then merged into a global transcriptome using Stringtie *merge* and filtered to only include those with FPKM > 0.1 and length > 30bp and then TransDecoder v2.0.1 (Haas et al., 2013) was used to predict coding regions in the global transcriptome with a minimum length of 20 amino acids.

FGENESH++v7.2.2 (Solovyev et al., 2006) was used for genome annotation using the longest open reading frame predicted from the global transcriptome, mammalian settings and optimised parameters supplied with the Tasmanian devil (*Sarcophilus harrisii*) gene finding matrix as this was the most closely related to most of the species in the study.

3. CAFE model selection

We tried a number of different parameter combinations to determine the best model for the data. First, we determined whether to use a base model or a discrete gamma model with k=2 (meaning gene families can belong to one of two different evolutionary rate categories) or k=3. Each model was run five times to check for convergence. Any models that did not converge were rejected and the gamma model with k=2 was retained.

We then tried both a global λ model and a multi-λ model (with λ_1 for Australian marsupials and λ_2 for American marsupials). Again, both models were run five times to check for convergence. To determine if the more constrained global λ model was a better fit than the multi-λ model, we performed a likelihood ratio test. We ran 100 simulations under the null hypothesis (i.e. the global λ model). For each simulation, we calculated the likelihood of the global λ model (L_{global}), the likelihood of the multi-λ model (L_{multi}) and then used these values to calculate the likelihood ratio: $2(lnL_{global} - lnL_{multi})$. We plotted a histogram of the likelihood ratio of all the simulations and used this to determine the probability of obtaining the value of our actual likelihood ratio under the null hypothesis (**Figure A1-7**). As $p < 0.05$, we rejected the null hypothesis and selected the multi-λ model.

SUPPLEMENTARY TABLES AND FIGURES

Table A1-1. Reported cases of neoplasia across marsupial taxa. Species are grouped by family, with the exception of Peramelemorphia and Phalangeriformes, as species in these taxa are often grouped together in reporting. ** indicates that this figure may not include all captive devils.

Taxon	Common name	Ratcliffe, 1933	Effron, Griner & Benirschke, 1977	Canfield & Cunningham, 1993	Canfield Hartley & Reddacliff, 1990a; Canfield, Hartley, & Reddacliff, 1990b	Ladds, 2009	Wildlife Health Registry (2004-2024)	Total
Dasyuridae	Dasyurids	3	3	7	70	140	59**	282
Didelphidae	Opossums	3	1	not reported	not reported	not reported	not reported	4
Macropodidae	Kangaroos, wallabies, quokkas	2	4	11	14	43	17	91
Peramelemorphia	Bandicoots and bilbies	1	not reported	not reported	3	19	12	35
Phascolarctidae	Koala	not reported	not reported	not reported	26	70	113	209
Phalangeriformes	Possums and gliders	not reported	not reported	0	22	74	33	129
Potoroidae	Bettongs, potoroos	not reported	not reported	not reported	not reported	not reported	3	3
Vombatidae	Wombats	not reported	not reported	0	2	2	not reported	4

Table A1-2. Details for genomes used in this study.

Species	Common Name	Genome assembly (NCBI accession number)	Transcriptome source (NCBI project number)	Genome Size (GB)	Contigs	Scaffolds	Scaffold N50 (MB)	Contig N50 (MB)	Complete BUSCOs [Single copy, Duplicated]
Dromiciops gliroides	Monito del monte	mDroGli1.pri (GCF_019393635.1)	PRJNA416414	3.3	277	17	670.8	38.2	96.7% [S:93.7%, D:3.0%]
Monodelphis domestica	Gray short-tailed opossum	mMonDom1.pri (GCF_027887165.1)	PRJNA200320	3.6	2,268	13	538.3	3.9	94.9% [S:92.5%, D:2.4%]
Macrotis lagotis	Greater bilby	bilby.v1.9 (GCF_037893015.1)	PRJNA1049866	3.7	5,027	608	343.9	1.2	95.6% [S:90.8%, D:4.8%]
Phascolarctos cinereus	Koala	phaCin_HiC (GCA_003287225.2)	PRJNA230900, PRJNA327021	3.2	1,913	1,245	428.2	11.4	97.2% [S:96.1%, D:1.1%]
Notamacropus eugenii	Tammar wallaby	Tammar_Male_v7_Final_Haploid_with_Y-001	PRJDB1934	3.4	56	9	483.6	194.0	96.9% [S:94.7%, D:2.3%]
Sminthopsis crassicaudata	Fat-tailed dunnart	dunnart_asm_12-2021_sm (Figshare)	PRJNA1028148	3.2	2,569	1,848	72.6	11.2	96.4% [S:94.9%, D:1.5%]
Antechinus flavipes	Yellow-footed antechinus	AdamAnt_v2 (GCF_016432865.1)	PRJNA565840	3.2	1,103	485	636.7	51.8	94.9% [S:93.7%, D:1.2%]

Species	Common Name	Genome assembly (NCBI accession number)	Transcriptome source (NCBI project number)	Genome Size (GB)	Contigs	Scaffolds	Scaffold N50 (MB)	Contig N50 (MB)	Complete BUSCOs [Single copy, Duplicated]
Sarcophilus harrisii	Tasmanian devil	mSarHar1.11 (GCF_902635505.1)	PRJEB34650	3.1	444	105	611.3	62.3	94.7% [S:93.6%, D:1.1%]
Dasyurus viverrinus	Eastern quoll	UniMelb_DasViv_v1.0 (GCA_020854095.1)	PRJNA963007	3.1	507	76	628.5	13.8	96.4% [S:95.4%, D:1.0%]
Dasyuroides byrneii	Kowari	mDasByr.1.2_20251403	this study	3.2	2,402	851	620.0	4.0	95.6% [S:94.5%, D:1.2%]
Perameles gunnii	Eastern barred bandicoot	mPerGun1.2_20251003	this study	3.9	4,595	1,462	722.0	2.0	89.5% [S:87.1%, D:2.5%]

Table A1-3. Reported cases of neoplasia across dasyurid genera. ** indicates that this figure may not include all captive devils.

Genus	Common name	Ratcliffe, 1933	Effron, Griner & Benirschke, 1977	Canfield, Hartley, & Reddacliff, 1990b	Ladds, 2009	Wildlife Health Registry (2004-2024)**	Total
Antechinus	Antechinus	not reported	not reported	4	10	1	15
Dasycercus	Mulgara	not reported	not reported	1	2	1	4
Dasyuroides	Kowari	not reported	not reported	11	32	not reported	43
Dasyurus	Quoll	2	not reported	20	45	12	79
Parantechinus	Dibbler	not reported	not reported	3	not reported	not reported	3
Phascogale	Phascogale	not reported	not reported	3	12	1	16
Planigale	Planigale	not reported	not reported	4	4	not reported	8
Pseudoantechinus	False antechinus	not reported	not reported	5	not reported	not reported	5
Sarcophilus	Devil	1	3	16	26	43**	89
Sminthopsis	Dunnart	not reported	not reported	3	8	1	12

Table A1-4. Number of protein-coding genes annotated by FGENESH in the 11 species, number annotated as cancer orthologs (Cosmic database), and percentage assigned to orthogroups.

Species	Total genes annotated by FGENESH	Total genes remaining after filtering	Annotated as cancer orthologs	Percentage of Genes in Orthogroups
Yellow footed antechinus	36,780	25,727	646	86.3
Eastern Barred Bandicoot	76,963	38,979	627	84.5
Greater Bilby	37,266	28,951	622	81.2
Tasmanian devil	32,130	24,172	648	83.4
Fat tailed dunnart	33,770	24,285	662	85.3
Koala	28,365	23,515	672	86.2
Kowari	73,135	39,773	674	88.9
Monito del Monte	28,719	23,917	588	85.7
Gray Short Tailed Opossum	36,640	30,251	576	68.8
Eastern Quoll	32,702	22,463	640	91.2
Tammar Wallaby	34,784	26,687	649	87.4

Table A1-5. Divergence dates for nodes in marsupial phylogeny that were obtained from timetree.org.

Node	Time (million years ago)
Quoll, devil	9.4
Kowari, (quoll and devil)	13.9
Antechinus, dasyurini	18
Dunnart, (antechinus, kowari, quoll, devil)	24.6
Bilby, bandicoot	30
Bilby, dasyurids	58
Koala, tammar	53
Bilby, diprotodontia	61
Monito, eomarsupialia	63
Monito, opossum	78

Table A1-6. Query sequences used for RAS annotations.

Species	Gene	Accession (Uniprot)
Homo sapiens	KRAS4A	P01116
Homo sapiens	KRAS4B	P01116-2
Homo sapiens	NRAS	P01111
Homo sapiens	HRAS	P01112
Homo sapiens	ERAS	Q7Z444
Homo sapiens	RRAS	P10301
Homo sapiens	RRAS2	P62070
Homo sapiens	MRAS	O14807
Homo sapiens	RIT1	Q92963
Homo sapiens	RIT2	Q99578
Homo sapiens	RAP1A	P62834
Homo sapiens	RAP1B	P61224
Homo sapiens	RAP2A	P10114
Homo sapiens	RAP2C	Q9Y3L5
Homo sapiens	RAP2B	P61225
Homo sapiens	RALA	P11233
Homo sapiens	RALB	P11234
Homo sapiens	REM1	O75628
Homo sapiens	REM2	Q8IYK8
Homo sapiens	RRAD	P55042
Homo sapiens	GEM	P55040
Homo sapiens	RERG	Q96A58
Homo sapiens	RASL11A	Q6T310
Homo sapiens	RASL11B	Q9BPW5
Homo sapiens	DIRAS1	O95057
Homo sapiens	RASL10A	Q92737
Homo sapiens	NKIRAS1	Q9NYS0
Homo sapiens	RASL12	Q9NYN1
Homo sapiens	RERGL	Q9H628
Homo sapiens	RHEB	Q15382
Homo sapiens	RHEBL1	Q8TAI7
Homo sapiens	DIRAS3	O95661
Homo sapiens	DIRAS2	Q96HU8
Homo sapiens	RASL10B	Q96S79
Homo sapiens	NKIRAS2	Q9NYR9
Homo sapiens	RASD2	Q96D21
Homo sapiens	RASD1	Q9Y272

Table A1-7. Accession numbers and gene names for nodes in the phylogenetic tree.

Accession number	Gene	Species	Phylo Group	Gene Subfamily
NM_181548.2	>ERAS_MOUSE	House mouse	Eutherian	ERAS
XM_059883764.1	>ERAS_COW	Domestic cattle	Eutherian	ERAS
XM_038448950.1	>ERAS_DOG	Dog	Eutherian	ERAS
XM_005878010.2	>ERAS_BAT	Brandt's bat	Eutherian	ERAS
XM_008272640.2	>ERAS_RABBIT	Rabbit	Eutherian	ERAS
XM_004690133.1	>ERAS_MOLE	Star-nosed mole	Eutherian	ERAS
XM_070258069.1	>ERAS_HORSE	Horse	Eutherian	ERAS
XM_036917191.2	>ERAS_PANGOLIN	Chinese pangolin	Eutherian	ERAS
XM_008571436.1	>ERAS_FLYING_LEMUR	Sunda flying lemur	Eutherian	ERAS
XM_006171940.1	>ERAS_TREE_SHREW	Chinese tree shrew	Eutherian	ERAS
XM_003417997.3	>ERAS_ELEPHANT	African savanna elephant	Eutherian	ERAS
XM_004464982.2	>ERAS_ARMADILLO	Nine-banded armadillo	Eutherian	ERAS
XM_004464982.2	>ERAS_SLOTH	Southern two-toed sloth	Eutherian	ERAS
NM_181532.3	>ERAS_HUMAN	Homo sapiens	Eutherian	ERAS
NG_042222.1	>RRAS_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NG_017058.1	>RRAS2_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_001085049.3	>MRAS_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NG_033885.1	>RIT1_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_001272077.2	>RIT2_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_001010935.3	>RAP1A_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_001010942.3	>RAP1B_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_021033.7	>RAP2A_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_001271186.2	>RAP2C_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_002886.4	>RAP2B_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_005402.4	>RALA_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_001369400.1	>RALB_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NG_046939.1	>REM1_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_173527.3	>REM2_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_001128850.2	>RRAD_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_005261.4	>GEM_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_001190726.2	>RERG_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_001331126.2	>RASL11A_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_023940.3	>RASL11b_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_145173.4	>DIRAS1_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_006477.5	>RASL10A_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_001377351.1	>NKIRAS1_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_001307930.2	>RASL12_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NG_052618.1	>RERGL_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_005614.4	>RHEB_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_001303126.2	>RHEBL1_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NG_011753.1	>DIRAS3_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_017594.5	>DIRAS2_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_033315.4	>RASL10b_HUMAN	Homo sapiens	Eutherian	Ras subfamily

Accession number	Gene	Species	Phylo Group	Gene Subfamily
NM_001001349.2	>NKIRAS2_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_001366725.1	>RASD2_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NG_028074.2	>RASD1_HUMAN	Homo sapiens	Eutherian	Ras subfamily
NM_001369786.1	>KRAS4A_HUMAN	Homo sapiens	Eutherian	Canonical
NM_001369787.1	>KRAS4B_HUMAN	Homo sapiens	Eutherian	Canonical
NG_007572.1	>NRAS_HUMAN	Homo sapiens	Eutherian	Canonical
NG_007666.1	>HRAS_HUMAN	Homo sapiens	Eutherian	Canonical
XM_064499648.1	>RRAS_BIRD	Emu	Aves	Ras subfamily
NM_204489.2	>MRAS_BIRD	Chicken	Aves	Ras subfamily
NM_001031327.3	>RIT1_BIRD	Chicken	Aves	Ras subfamily
XM_001233996.7	>RIT2_BIRD	Chicken	Aves	Ras subfamily
XM_046904296.1	>RAP1A_BIRD	Chicken	Aves	Ras subfamily
NM_001007852.1	>RAP1B_BIRD	Chicken	Aves	Ras subfamily
XM_001233103.6	>RAP2A_BIRD	Chicken	Aves	Ras subfamily
NM_001012572.3	>RAP2C_BIRD	Chicken	Aves	Ras subfamily
XM_015291825.4	>RAP2B_BIRD	Chicken	Aves	Ras subfamily
XM_046937036.1	>RALA_BIRD	Chicken	Aves	Ras subfamily
XM_025152428.3	>RALB_BIRD	Chicken	Aves	Ras subfamily
XM_015296468.4	>REM1_BIRD	Chicken	Aves	Ras subfamily
XM_064475528.1	>REM2_BIRD	Great cormorant	Aves	Ras subfamily
NM_001277606.3	>RRAD_BIRD	Chicken	Aves	Ras subfamily
NM_213579.2	>GEM_BIRD	Chicken	Aves	Ras subfamily
XM_046904998.1	>RERG_BIRD	Chicken	Aves	Ras subfamily
XM_417126.8	>RASL11A_BIRD	Chicken	Aves	Ras subfamily
XM_420710.8	>RASL11b_BIRD	Chicken	Aves	Ras subfamily
XM_015299968.4	>DIRAS1_BIRD	Chicken	Aves	Ras subfamily
NM_001030702.2	>RASL10A_BIRD	Chicken	Aves	Ras subfamily
XM_040696256.2	>NKIRAS1_BIRD	Chicken	Aves	Ras subfamily
XM_004943886.5	>RASL12_BIRD	Chicken	Aves	Ras subfamily
XM_416411.8	>RERGL_BIRD	Chicken	Aves	Ras subfamily
XM_040695391.2	>RHEB_BIRD	Chicken	Aves	Ras subfamily
XM_065043780.1	>RHEBL1_BIRD	Rock pigeon	Aves	Ras subfamily
XM_021296473.2	>DIRAS3_BIRD	Rock pigeon	Aves	Ras subfamily
XM_423026.8	>DIRAS2_BIRD	Chicken	Aves	Ras subfamily
XM_001233673.7	>RASL10b_BIRD	Chicken	Aves	Ras subfamily
NM_001006333.2	>NKIRAS2_BIRD	Chicken	Aves	Ras subfamily
XM_416293.8	>RASD2_BIRD	Chicken	Aves	Ras subfamily
NM_001044636.2	>RASD1_BIRD	Chicken	Aves	Ras subfamily
NM_001256162.1	>KRAS_BIRD	Chicken	Aves	Canonical
NM_001012549.2	>NRAS_BIRD	Chicken	Aves	Canonical
NM_001396746.1	>HRAS_BIRD	Chicken	Aves	Canonical
XM_033171113.1	>KRASBL_LIZARD	Sand lizard	Lizard	Canonical
NM_001114248.1	>RRAS_FROG	Tropical clawed frog	Amphibian	Ras subfamily
XM_031900379.1	>RRAS2_FROG	Tropical clawed frog	Amphibian	Ras subfamily

Accession number	Gene	Species	Phylo Group	Gene Subfamily
XM_002943157.5	>MRAS_FROG	Tropical clawed frog	Amphibian	Ras subfamily
XM_031891490.1	>RIT1_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_001102887.1	>RAP1A_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_001008194.2	>RAP1B_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_001035116.1	>RAP2A_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_001016898.2	>RAP2C_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_001035115.1	>RAP2B_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_001015915.2	>RALA_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_001102845.1	>RALB_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_001102791.1	>REM1_FROG	Tropical clawed frog	Amphibian	Ras subfamily
XM_002941509.5	>REM2_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_001016726.2	>RRAD_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_001097353.1	>GEM_FROG	Tropical clawed frog	Amphibian	Ras subfamily
XM_040345529.1	>RERG_FROG	Common frog	Amphibian	Ras subfamily
XM_002941535.4	>RASL11A_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_001015774.1	>RASL11b_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_001079312.1	>DIRAS1_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_001005037.1	>RASL10A_FROG	Tropical clawed frog	Amphibian	Ras subfamily
XM_002932411.5	>NKIRAS1_FROG	Tropical clawed frog	Amphibian	Ras subfamily
XM_004912687.4	>RASL12_FROG	Tropical clawed frog	Amphibian	Ras subfamily
XM_040345486.1	>RERGL_FROG	Common frog	Amphibian	Ras subfamily
NM_001015922.2	>RHEB_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_203606.2	>RHEBL1_FROG	Tropical clawed frog	Amphibian	Ras subfamily
XM_031899814.1	>DIRAS3_FROG	Tropical clawed frog	Amphibian	Ras subfamily
XM_002937026.4	>DIRAS2_FROG	Tropical clawed frog	Amphibian	Ras subfamily
XM_004911629.4	>RASL10b_FROG	Tropical clawed frog	Amphibian	Ras subfamily
XM_002939033.5	>NKIRAS2_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_001016006.2	>RASD2_FROG	Tropical clawed frog	Amphibian	Ras subfamily
NM_001078940.1	>RASD1_FROG	Tropical clawed frog	Amphibian	Ras subfamily
XM_040437469.1	>KRAS_TOAD	Common toad	Amphibian	Canonical
NM_001016763.2	>NRAS_FROG	Tropical clawed frog	Amphibian	Canonical
NM_001017003.2	>HRAS_FROG	Tropical clawed frog	Amphibian	Canonical
NM_001008033.1	>KRASBL_FROG	Tropical clawed frog	Amphibian	Canonical
NM_001005931.2	>RRAS_FISH	Zebrafish	Fish	Ras subfamily
NM_001017815.1	>RRAS2_FISH	Zebrafish	Fish	Ras subfamily
XM_033024843.1	>MRAS_FISH	Thorny skate	Fish	Ras subfamily
NM_001128781.1	>RIT1_FISH	Thorny skate	Fish	Ras subfamily
XM_033031348.1	>RIT2_FISH	Thorny skate	Fish	Ras subfamily
NM_001002152.1	>RAP1A_FISH	Zebrafish	Fish	Ras subfamily
NM_199533.1	>RAP1B_FISH	Zebrafish	Fish	Ras subfamily
NM_001145705.1	>RAP2A_FISH	Zebrafish	Fish	Ras subfamily
NM_001007055.1	>RAP2C_FISH	Zebrafish	Fish	Ras subfamily
NM_001001729.2	>RAP2B_FISH	Zebrafish	Fish	Ras subfamily
NM_201018.1	>RALA_FISH	Zebrafish	Fish	Ras subfamily

Accession number	Gene	Species	Phylo Group	Gene Subfamily
NM_001003649.1	>RALB_FISH	Zebrafish	Fish	Ras subfamily
NM_201174.1	>REM1_FISH	Zebrafish	Fish	Ras subfamily
NM_001123046.1	>REM2_FISH	Zebrafish	Fish	Ras subfamily
NM_199798.1	>RRAD_FISH	Zebrafish	Fish	Ras subfamily
NM_001045849.1	>GEM_FISH	Zebrafish	Fish	Ras subfamily
NM_001327837.1	>RERG_FISH	Zebrafish	Fish	Ras subfamily
NM_001017840.2	>RASL11A_FISH	Zebrafish	Fish	Ras subfamily
NM_200140.1	>RASL11b_FISH	Zebrafish	Fish	Ras subfamily
NM_199831.1	>DIRAS1_FISH	Zebrafish	Fish	Ras subfamily
NM_001128366.1	>RASL10A_FISH	Zebrafish	Fish	Ras subfamily
NM_001100076.1	>NKIRAS1_FISH	Zebrafish	Fish	Ras subfamily
NM_200395.1	>RASL12_FISH	Zebrafish	Fish	Ras subfamily
NM_001002494.1	>RERGL_FISH	Zebrafish	Fish	Ras subfamily
NM_200729.1	>RHEB_FISH	Zebrafish	Fish	Ras subfamily
NM_001076748.2	>RHEBL1_FISH	Zebrafish	Fish	Ras subfamily
XM_033028462.1	>DIRAS3_FISH	Thorny skate	Fish	Ras subfamily
XM_005155552.4	>DIRAS2_FISH	Zebrafish	Fish	Ras subfamily
XM_068221542.1	>RASL10b_FISH	Zebrafish	Fish	Ras subfamily
NM_001003433.1	>NKIRAS2_FISH	Zebrafish	Fish	Ras subfamily
NM_001030202.2	>RASD2_FISH	Zebrafish	Fish	Ras subfamily
NM_200532.1	>RASD1_FISH	Zebrafish	Fish	Ras subfamily
NM_001003744.2	>KRAS_FISH	Zebrafish	Fish	Canonical
NM_131145.2	>NRAS_FISH	Zebrafish	Fish	Canonical
NM_001017623.1	>HRAS_FISH	Zebrafish	Fish	Canonical
NM_001292570.1	>KRASBL_SHARK	Elephant Shark	Fish	Canonical
This study	>Af_HRas	Yellow-footed antechinus	Dasyuromorphia	Canonical
This study	>Af_KRas4a	Yellow-footed antechinus	Dasyuromorphia	Canonical
This study	>Af_KRas4b	Yellow-footed antechinus	Dasyuromorphia	Canonical
This study	>Af_NRas	Yellow-footed antechinus	Dasyuromorphia	Canonical
This study	>AfMgRas1	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas10	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas11	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas12	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas13	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas14	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas15	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas16	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas17	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas18	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas19	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas2	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas20	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas21	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas22	Yellow-footed antechinus	Dasyuromorphia	MgRas

Accession number	Gene	Species	Phylo Group	Gene Subfamily
This study	>AfMgRas23	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas3	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas4	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas5	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas6	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas7	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas8	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>AfMgRas9	Yellow-footed antechinus	Dasyuromorphia	MgRas
This study	>Db_Hras	Kowari	Dasyuromorphia	Canonical
This study	>Db_Kras4A	Kowari	Dasyuromorphia	Canonical
This study	>Db_Kras4B	Kowari	Dasyuromorphia	Canonical
This study	>Db_Nras	Kowari	Dasyuromorphia	Canonical
This study	>DbMgRas1	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas10	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas11	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas12	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas13	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas14	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas15	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas16	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas17	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas18	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas19	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas2	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas20	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas21	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas3	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas4	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas5	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas6	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas7	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas8	Kowari	Dasyuromorphia	MgRas
This study	>DbMgRas9	Kowari	Dasyuromorphia	MgRas
This study	>Dg_HRas	Monito del monte	Microbiotheria	Canonical
This study	>Dg_KRas4A	Monito del monte	Microbiotheria	Canonical
This study	>Dg_KRas4B	Monito del monte	Microbiotheria	Canonical
This study	>Dg_NRas	Monito del monte	Microbiotheria	Canonical
This study	>DgMgRas1	Monito del monte	Microbiotheria	MgRas
This study	>DgMgRas2	Monito del monte	Microbiotheria	MgRas
This study	>Dv_HRas	Eastern quoll	Dasyuromorphia	Canonical
This study	>Dv_KRas4A	Eastern quoll	Dasyuromorphia	Canonical
This study	>Dv_KRas4B	Eastern quoll	Dasyuromorphia	Canonical
This study	>Dv_NRas	Eastern quoll	Dasyuromorphia	Canonical
This study	>DvMgRas1	Eastern quoll	Dasyuromorphia	MgRas

Accession number	Gene	Species	Phylo Group	Gene Subfamily
This study	>DvMgRas10	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas11	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas12	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas13	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas14	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas15	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas16	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas17	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas18	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas19	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas2	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas20	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas21	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas22	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas3	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas4	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas5	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas6	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas7	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas8	Eastern quoll	Dasyuromorphia	MgRas
This study	>DvMgRas9	Eastern quoll	Dasyuromorphia	MgRas
This study	>Md_HRas	Grey short-tailed opossum	Didelphimorphia	Canonical
This study	>Md_KRas4B_DUP	Grey short-tailed opossum	Didelphimorphia	Canonical
This study	>Md_KRas4A	Grey short-tailed opossum	Didelphimorphia	Canonical
This study	>Md_KRas4B	Grey short-tailed opossum	Didelphimorphia	Canonical
This study	>Md_NRas	Grey short-tailed opossum	Didelphimorphia	Canonical
This study	>MdMgRas1	Grey short-tailed opossum	Didelphimorphia	MgRas
This study	>MdMgRas2	Grey short-tailed opossum	Didelphimorphia	MgRas
This study	>Me_HRas	Tammar Wallaby	Diprotodontia	Canonical
This study	>Me_KRas4A	Tammar Wallaby	Diprotodontia	Canonical
This study	>Me_KRas4B	Tammar Wallaby	Diprotodontia	Canonical
This study	>Me_NRas	Tammar Wallaby	Diprotodontia	Canonical
This study	>MeMgRas1	Tammar Wallaby	Diprotodontia	MgRas
This study	>MeMgRas2	Tammar Wallaby	Diprotodontia	MgRas
This study	>MeMgRas3	Tammar Wallaby	Diprotodontia	MgRas
This study	>Ml_HRas	Greater bilby	Peramelemorphia	Canonical
This study	>Ml_KRas4a	Greater bilby	Peramelemorphia	Canonical
This study	>Ml_Kras4b	Greater bilby	Peramelemorphia	Canonical

Accession number	Gene	Species	Phylo Group	Gene Subfamily
This study	>Ml_NRas	Greater bilby	Peramelemorphia	Canonical
This study	>MlMgRas10	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas11	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas12	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas13	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas14	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas15	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas16	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas17	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas18	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas19	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas20	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas21	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas22	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas23	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas24	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas3	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas4	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas5	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas6	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas7	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas8	Greater bilby	Peramelemorphia	MgRas
This study	>MlMgRas9	Greater bilby	Peramelemorphia	MgRas
This study	>Pc_HRas	Koala	Diprotodontia	Canonical
This study	>Pc_KRas4A	Koala	Diprotodontia	Canonical
This study	>Pc_KRas4B	Koala	Diprotodontia	Canonical
This study	>Pc_NRas	Koala	Diprotodontia	Canonical
This study	>PcMgRas1	Koala	Diprotodontia	MgRas
This study	>PcMgRas2	Koala	Diprotodontia	MgRas
This study	>PcMgRas3	Koala	Diprotodontia	MgRas
This study	>Pg_HRas	Eastern barred bandicoot	Peramelemorphia	Canonical
This study	>Pg_KRas4a	Eastern barred bandicoot	Peramelemorphia	Canonical
This study	>Pg_KRas4b	Eastern barred bandicoot	Peramelemorphia	Canonical
This study	>Pg_NRas	Eastern barred bandicoot	Peramelemorphia	Canonical
This study	>PgMgRas1	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas10	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas100	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas101	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas102	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas103	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas104	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas105	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas106	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas107	Eastern barred bandicoot	Peramelemorphia	MgRas

Accession number	Gene	Species	Phylo Group	Gene Subfamily
This study	>PgMgRas108	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas109	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas11	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas110	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas111	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas112	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas113	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas114	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas115	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas116	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas117	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas12	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas13	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas14	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas15	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas16	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas17	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas18	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas19	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas2	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas20	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas21	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas22	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas23	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas24	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas25	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas26	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas27	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas28	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas29	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas3	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas30	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas31	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas32	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas33	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas34	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas35	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas36	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas37	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas38	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas39	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas4	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas40	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas41	Eastern barred bandicoot	Peramelemorphia	MgRas

Accession number	Gene	Species	Phylo Group	Gene Subfamily
This study	>PgMgRas42	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas43	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas44	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas45	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas46	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas47	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas48	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas49	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas5	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas50	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas51	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas52	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas53	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas54	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas55	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas56	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas57	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas58	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas59	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas6	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas60	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas61	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas62	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas63	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas64	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas65	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas66	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas67	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas68	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas69	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas7	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas70	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas71	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas72	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas73	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas74	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas75	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas76	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas77	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas78	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas79	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas8	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas80	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas81	Eastern barred bandicoot	Peramelemorphia	MgRas

Accession number	Gene	Species	Phylo Group	Gene Subfamily
This study	>PgMgRas82	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas83	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas84	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas85	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas86	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas87	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas88	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas89	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas9	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas90	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas91	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas92	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas93	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas94	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas95	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas96	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas97	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas98	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>PgMgRas99	Eastern barred bandicoot	Peramelemorphia	MgRas
This study	>Sc_Hras	Fat-tailed dunnart	Dasyuromorphia	Canonical
This study	>Sc_KRas4a	Fat-tailed dunnart	Dasyuromorphia	Canonical
This study	>Sc_KRas4b	Fat-tailed dunnart	Dasyuromorphia	Canonical
This study	>Sc_Nras	Fat-tailed dunnart	Dasyuromorphia	Canonical
This study	>ScMgRas1	Fat-tailed dunnart	Dasyuromorphia	MgRas
This study	>ScMgRas10	Fat-tailed dunnart	Dasyuromorphia	MgRas
This study	>ScMgRas11	Fat-tailed dunnart	Dasyuromorphia	MgRas
This study	>ScMgRas12	Fat-tailed dunnart	Dasyuromorphia	MgRas
This study	>ScMgRas13	Fat-tailed dunnart	Dasyuromorphia	MgRas
This study	>ScMgRas14	Fat-tailed dunnart	Dasyuromorphia	MgRas
This study	>ScMgRas2	Fat-tailed dunnart	Dasyuromorphia	MgRas
This study	>ScMgRas3	Fat-tailed dunnart	Dasyuromorphia	MgRas
This study	>ScMgRas4	Fat-tailed dunnart	Dasyuromorphia	MgRas
This study	>ScMgRas5	Fat-tailed dunnart	Dasyuromorphia	MgRas
This study	>ScMgRas6	Fat-tailed dunnart	Dasyuromorphia	MgRas
This study	>ScMgRas7	Fat-tailed dunnart	Dasyuromorphia	MgRas
This study	>ScMgRas8	Fat-tailed dunnart	Dasyuromorphia	MgRas
This study	>ScMgRas9	Fat-tailed dunnart	Dasyuromorphia	MgRas
This study	>Sh_DIRAS1	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_DIRAS2	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_GEM	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_Hras	Tasmanian devil	Dasyuromorphia	Canonical
This study	>Sh_KRas4a	Tasmanian devil	Dasyuromorphia	Canonical
This study	>Sh_KRas4b	Tasmanian devil	Dasyuromorphia	Canonical
This study	>Sh_MRAS	Tasmanian devil	Dasyuromorphia	Ras subfamily

Accession number	Gene	Species	Phylo Group	Gene Subfamily
This study	>Sh_NKIRAS1	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_NKIRAS2	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_Nras	Tasmanian devil	Dasyuromorphia	Canonical
This study	>Sh_RAD	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RALA	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RALB	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RAP1A	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RAP1B	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RAP2A	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RAP2B	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RAP2C	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RASD1	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RASL10A	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RASL10b	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RASL11A	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RASL11b	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RASL12	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_REM1	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_REM2	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RERG	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RHEB	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RIT1	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RIT2	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RRAS	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>Sh_RRAS2	Tasmanian devil	Dasyuromorphia	Ras subfamily
This study	>ShMgRas1	Tasmanian devil	Dasyuromorphia	MgRas
This study	>ShMgRas10	Tasmanian devil	Dasyuromorphia	MgRas
This study	>ShMgRas11	Tasmanian devil	Dasyuromorphia	MgRas
This study	>ShMgRas12	Tasmanian devil	Dasyuromorphia	MgRas
This study	>ShMgRas13	Tasmanian devil	Dasyuromorphia	MgRas
This study	>ShMgRas14	Tasmanian devil	Dasyuromorphia	MgRas
This study	>ShMgRas16	Tasmanian devil	Dasyuromorphia	MgRas
This study	>ShMgRas17	Tasmanian devil	Dasyuromorphia	MgRas
This study	>ShMgRas2	Tasmanian devil	Dasyuromorphia	MgRas
This study	>ShMgRas3	Tasmanian devil	Dasyuromorphia	MgRas
This study	>ShMgRas4	Tasmanian devil	Dasyuromorphia	MgRas
This study	>ShMgRas5	Tasmanian devil	Dasyuromorphia	MgRas
This study	>ShMgRas6	Tasmanian devil	Dasyuromorphia	MgRas
This study	>ShMgRas7	Tasmanian devil	Dasyuromorphia	MgRas
This study	>ShMgRas8	Tasmanian devil	Dasyuromorphia	MgRas
This study	>ShMgRas9	Tasmanian devil	Dasyuromorphia	MgRas

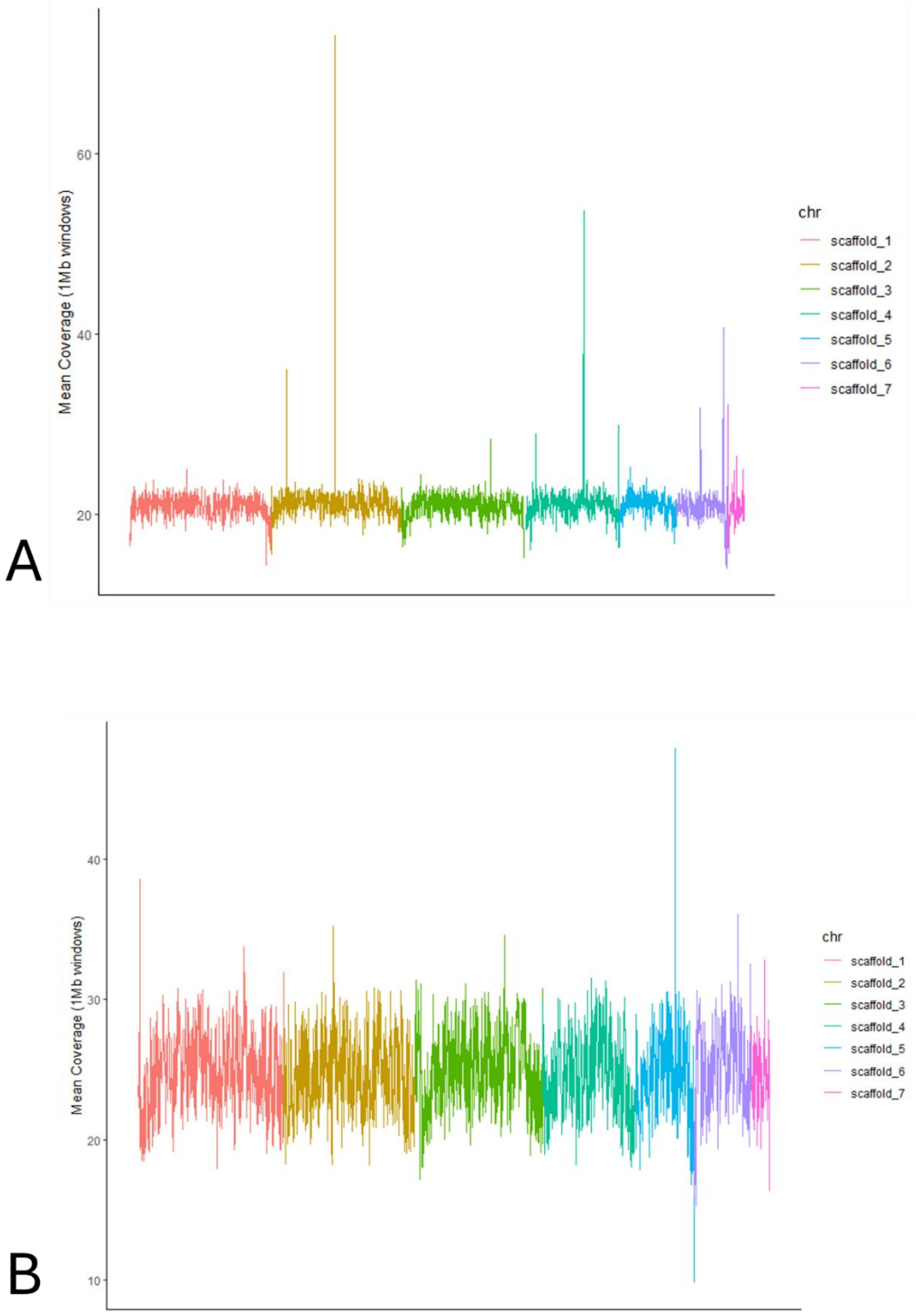


Figure A1-1. Plot showing mean coverage across the genome for A. Kowari and B. Bandicoot. Scaffolds are shown in different colours. For both species, coverage across the

genome was approximately equal across all scaffolds, confirming that the bandicoot sample came from a homogametic individual and suggesting that the kowari sample did as well.

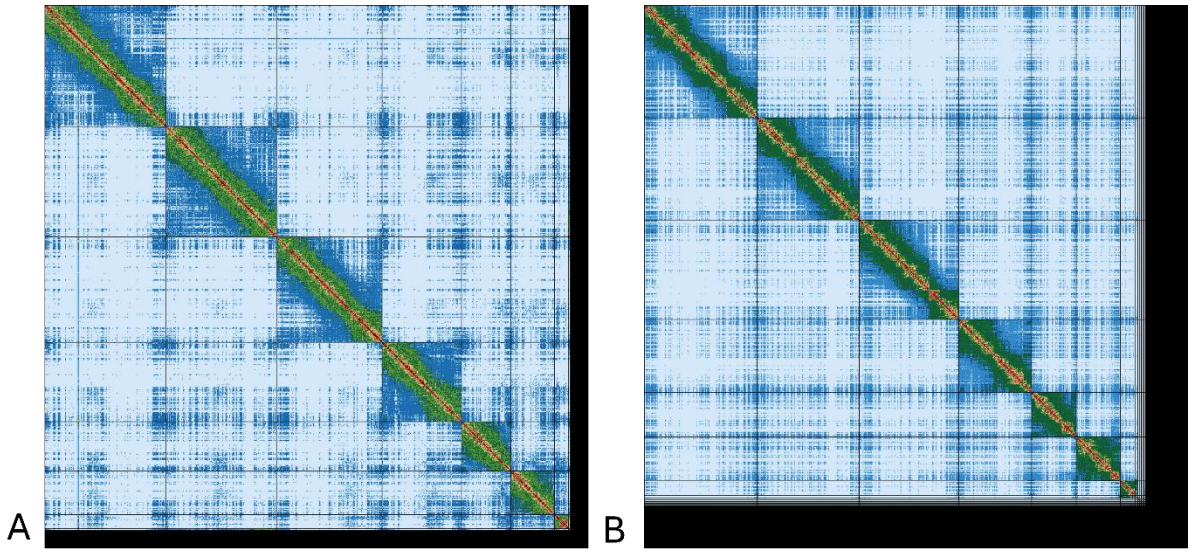


Figure A1-2. Hi-C contact map of the A. kowari and B. bandicoot assembly. Both indicate seven large scaffolds corresponding to the seven chromosomes expected in the species.

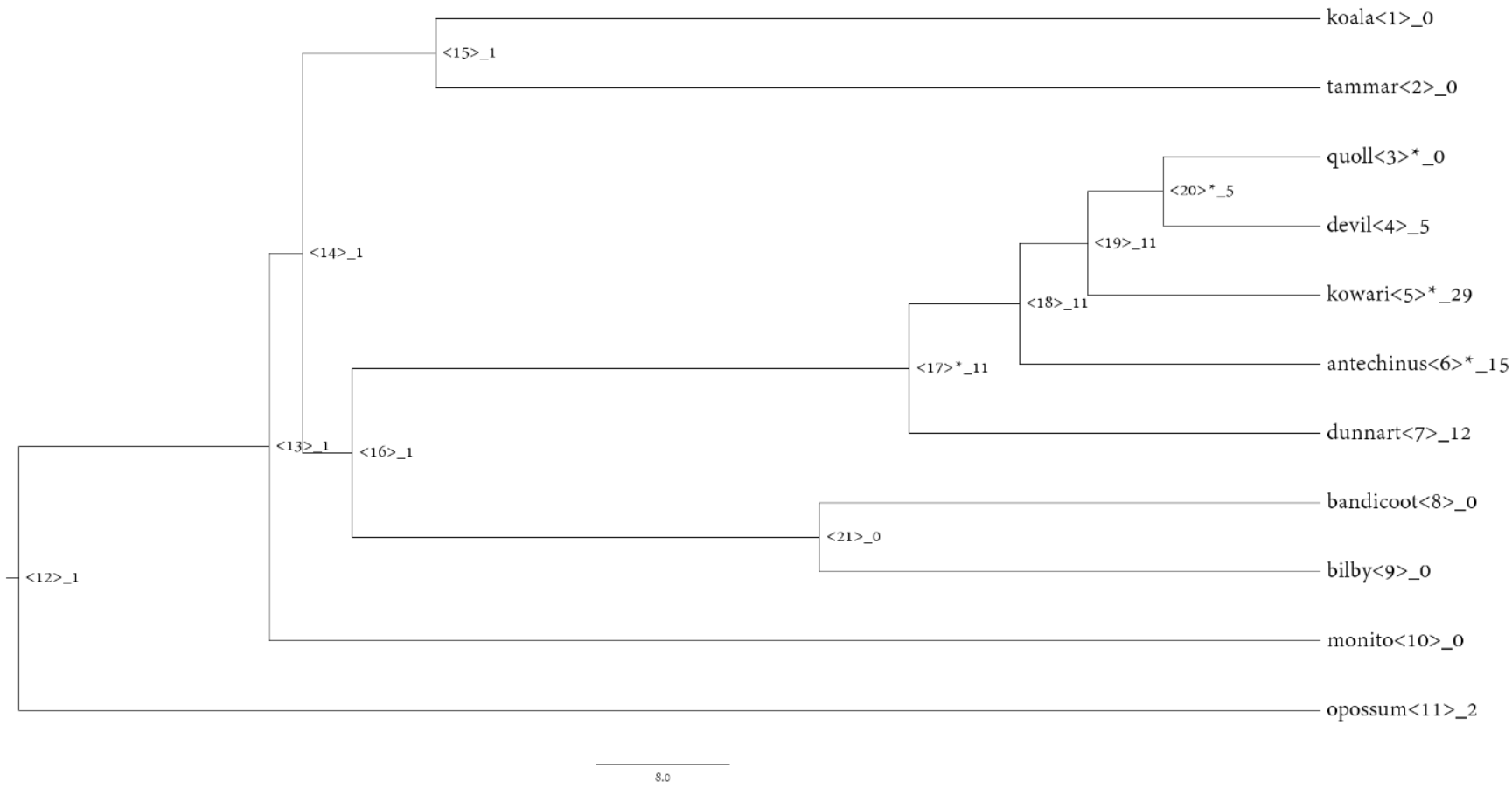


Figure A1-3. Gene tree for orthogroup HOG000767. The number after the underscore represents the number of genes each species has (or is predicted to have) in the orthogroup. Asterisk indicates that a statistically significant expansion or contraction occurred in this lineage. This orthogroup contains putative orthologs of ATRX, a tumour suppressor. The orthogroup underwent a significant expansion in the dasyurid ancestor but subsequent contractions in the ancestor of the quoll and devil.

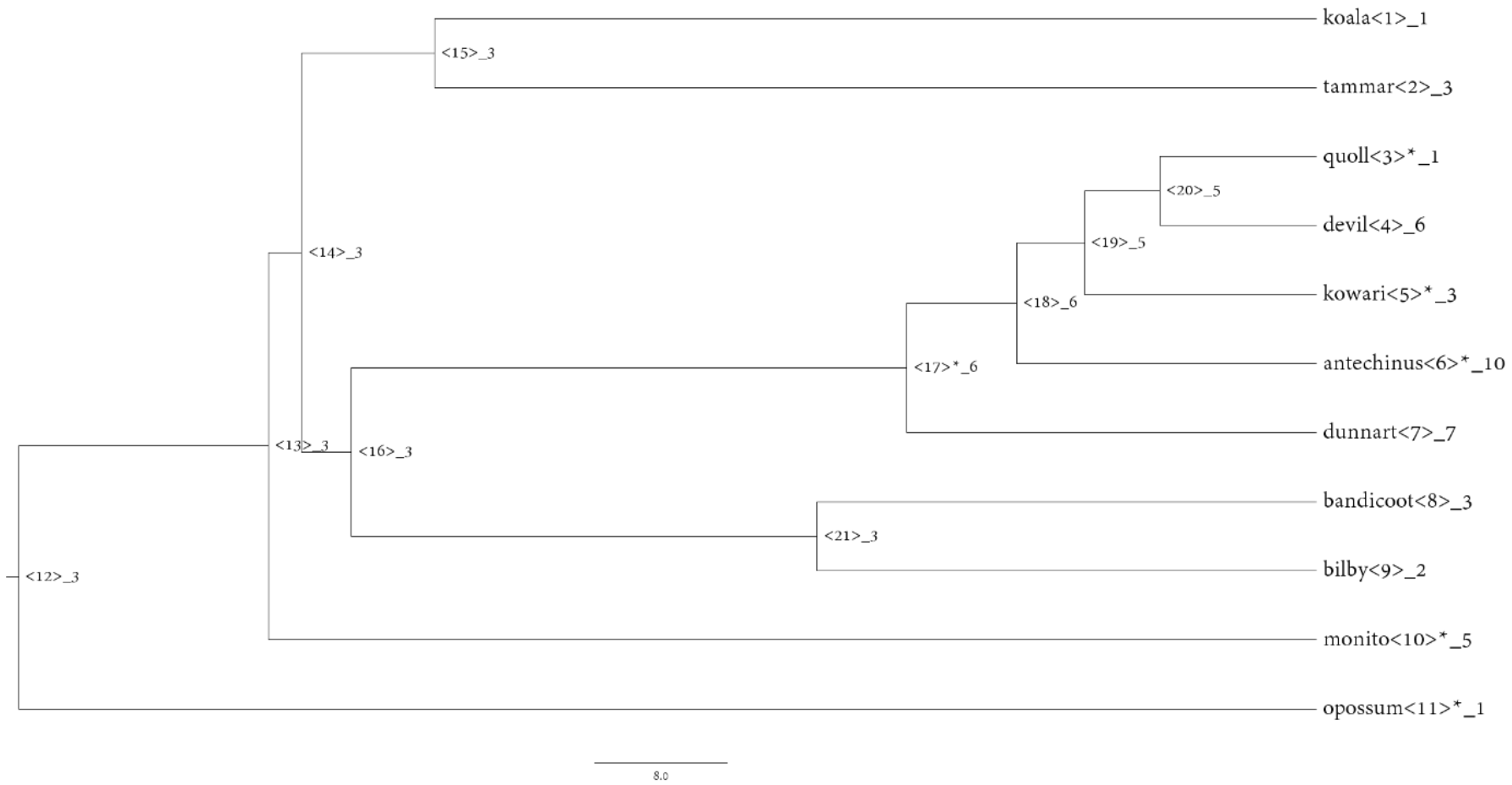


Figure A1-4. Gene tree for orthogroup HOG0001427. The number after the underscore represents the number of genes each species has (or is predicted to have) in the orthogroup. Asterisk indicates that a statistically significant expansion or contraction occurred in this lineage. This orthogroup contains copies of SLC34A2, a tumour supressor also involved in oncogenic fusion. The orthogroup underwent a significant expansion in the dasyurid ancestor but then subsequent contractions in the quoll and kowari.

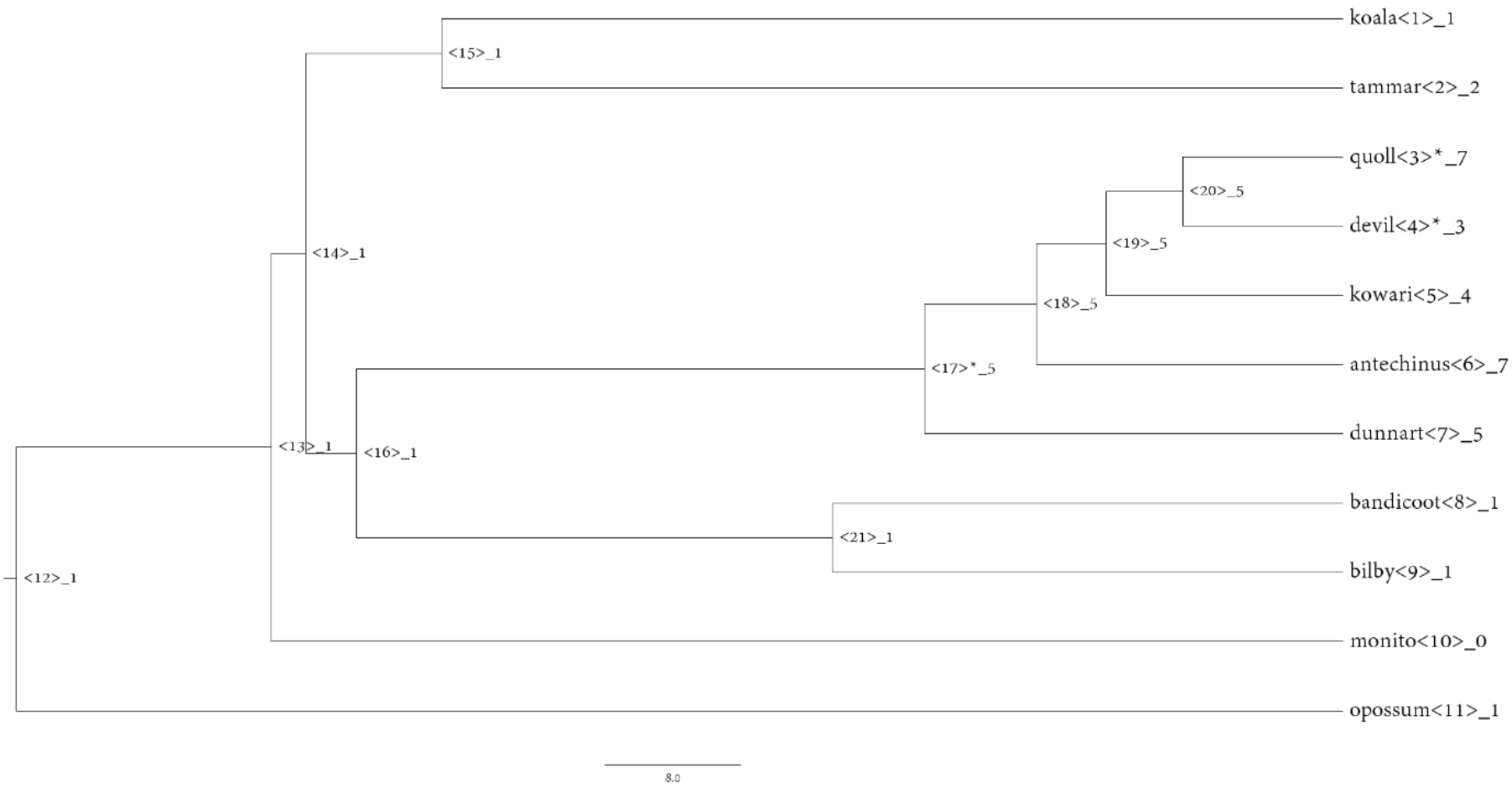


Figure A1-5. Gene tree for orthogroup HOG0002169. The number after the underscore represents the number of genes each species has (or is predicted to have) in the orthogroup. Asterisk indicates that a statistically significant expansion or contraction occurred in this lineage. This orthogroup contains copies of TCEA1, a gene involved in oncogenic fusions. The orthogroup underwent a significant expansion in the dasyurid ancestor but subsequently a contraction in the devil.


G1
G2
G3
G4
G5
AfMgRas1
AfMgRas2
AfMgRas3
AfMgRas4
AfMgRas5
PgMgRas1
PgMgRas2
PgMgRas3
PgMgRas4
PgMgRas5
MlMgRas1
MlMgRas3
MlMgRas4
MlMgRas5
ShMgRas1
ShMgRas2
ShMgRas3
ShMgRas4
ShMgRas5
ScMgRas1
ScMgRas2
ScMgRas3
ScMgRas4
ScMgRas5
DbMgRas1
DbMgRas2
DbMgRas3
DbMgRas4
DbMgRas5
DgMgRas1
DgMgRas2
MdMgRas1
MdMgRas2
DvMgRas1
DvMgRas2
DvMgRas3
DvMgRas4
DvMgRas5
MeMgRas1
MeMgRas2
MeMgRas3

Figure A1-6. Multisequence alignment of MgRas genes with the five conserved G box motifs outlined in black. If a species had more than five genes, only the first five were represented in the figure. Species in the alignment are yellow-footed antechinus (Af), eastern barred bandicoot (Pg), bilby (Ml), Tasmanian devil (Sh), fat-tailed dunnart (Sc), koala (Pc), kowari (Db), monito del monte (Dg), grey short-tailed opossum (Md), eastern quoll (Dv) and tammar wallaby (Me).

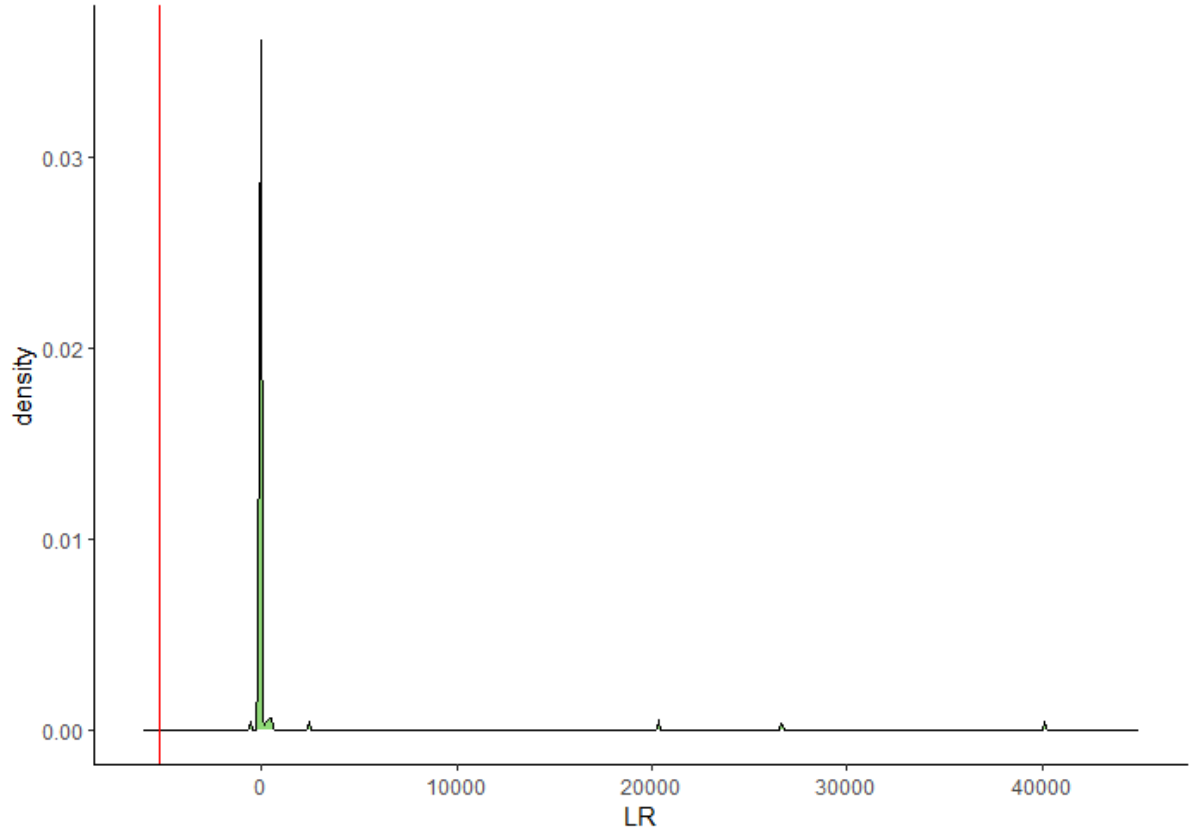


Figure A1-7. Likelihood ratio (LR) distribution under the null hypothesis. The green histogram represents the likelihood ratios obtained from the 100 simulations and the red line indicates the actual likelihood ratio (i.e. the likelihood of the more complex model divided by the likelihood of the simpler model, as determined by CAFE). 0 of the 100 simulations had a value that was equal to or more extreme than the actual likelihood ratio, indicating that the probability (p-value) of obtaining this value under the null hypothesis is <0.01.

APPENDIX 2. SUPPLEMENTARY MATERIAL TO CHAPTER 3

SUPPLEMENTARY TABLES AND FIGURES

Table A2-1. Observed intraspecific diversity in vertebrate AMPs.

Species	Common Name	Gene type	Gene	Total SNPs (exonic)	Total nsSNPs	CNV type	Copy number	Reference
Bos grunniens	Yak	Cathelicidin	CATHL1	N/A	N/A	duplication	2 to 9	(Zhang et al., 2016)
Bos taurus	Cattle	Cathelicidin	CATHL2	1	1	N/A	N/A	(Gillenwaters et al., 2009)
Bos taurus	Cattle	Cathelicidin	CATHL4	N/A	N/A	duplication	unknown	(Bickhart et al., 2012)
Bos taurus	Cattle	Cathelicidin	CATHL5	8	2	N/A	N/A	(Gillenwaters et al., 2009)
Bos taurus	Cattle	Cathelicidin	CATHL6	9	4	N/A	N/A	(Gillenwaters et al., 2009)
Bos taurus	Cattle	Cathelicidin	CATHL7	2	0	N/A	N/A	(Gillenwaters et al., 2009)
Bubalus bubalis	Buffalo	Cathelicidin	CATHL4	N/A	N/A	duplication	5 to 8	(Brahma et al., 2015)
Coturnix japonica	Quail	Cathelicidin	CjCATH1	2	0	N/A	N/A	(Ishige et al., 2021)
Coturnix japonica	Quail	Cathelicidin	CjCATH2	10	6	N/A	N/A	(Ishige et al., 2021)
Coturnix japonica	Quail	Cathelicidin	CjCATH3	5	0	N/A	N/A	(Ishige et al., 2021)
Coturnix japonica	Quail	Cathelicidin	CjCATHB1	2	1	N/A	N/A	(Ishige et al., 2021)
Homo sapiens	Human	Defensin	DEFB127	5	4	N/A	N/A	(Hollox & Armour, 2008)

Species	Common Name	Gene type	Gene	Total SNPs (exonic)	Total nsSNPs	CNV type	Copy number	Reference
Homo sapiens	Human	Defensin	DEFB120	3	3	N/A	N/A	(Hollox & Armour, 2008)
Homo sapiens	Human	Defensin	DEFB132	7	5	N/A	N/A	(Hollox & Armour, 2008)
Homo sapiens	Human	Defensin	DEFB118	2	2	N/A	N/A	(Hollox & Armour, 2008)
Homo sapiens	Human	Defensin	DEFB1	3	2	N/A	N/A	(Hollox & Armour, 2008)
Homo sapiens	Humans	Defensin	8p23.1 locus (gene cluster)	N/A	N/A	duplication	2 to 12	(Milanese et al., 2009)
Homo sapiens	Humans	Defensin	DEFA3	N/A	N/A	duplication	0 to 10	(Milanese et al., 2009)
Homo sapiens	Humans	Defensin	DEFA7	N/A	N/A	duplication	2 to 6	(Milanese et al., 2009)
Homo sapiens	Humans	Defensin	DEFB4	N/A	N/A	duplication	1 to 9	(Milanese et al., 2009)
Homo sapiens	Humans	Defensin	DEFB104	N/A	N/A	duplication	2 to 9	(Milanese et al., 2009)
Homo sapiens	Humans	Defensin	DEFB105	N/A	N/A	duplication	2 to 9	(Milanese et al., 2009)
Homo sapiens	Humans	Defensin	DEFB106	N/A	N/A	duplication	2 to 6	(Milanese et al., 2009)
Homo sapiens	Humans	Defensin	DEFB107	N/A	N/A	duplication	1 to 4	(Milanese et al., 2009)
Ovis aries	Sheep	Defensin	SBD1	2	0	N/A	N/A	(Monteleone et al., 2011)
Ovis aries	Sheep	Defensin	SBD2	5	2	N/A	N/A	(Monteleone et al., 2011)
Parus major	Great tit	Defensin	AvBD2	1	1	N/A	N/A	(Hellgren, 2015)
Parus major	Great tit	Defensin	AvBD4	3	2	N/A	N/A	(Hellgren, 2015)
Parus major	Great tit	Defensin	AvBD7	3	2	N/A	N/A	(Hellgren, 2015)
Parus major	Great tit	Defensin	AvBD9	0	0	N/A	N/A	(Hellgren, 2015)
Parus major	Great tit	Defensin	AvBD10	1	1	N/A	N/A	(Hellgren, 2015)
Parus major	Great tit	Defensin	AvBD12	2	2	N/A	N/A	(Hellgren, 2015)

Species	Common Name	Gene type	Gene	Total SNPs (exonic)	Total nsSNPs	CNV type	Copy number	Reference
Sus scrofa	Pig	Cathelicidin	NPG3	N/A	N/A	duplication	4 to 23	(Paudel et al., 2013)
Sus scrofa	Pig	Cathelicidin	PMAP23	N/A	N/A	duplication	2 to 12	(Paudel et al., 2013)
Sus scrofa	Pig	Cathelicidin	PR39	N/A	N/A	duplication	2 to 10	(Ahn et al., 2022)
Tachycineta bicolor	Tree swallow	Defensin	AvBD2	8	3	N/A	N/A	(Schmitt et al., 2017)
Tachycineta bicolor	Tree swallow	Defensin	AvBD4	5	3	N/A	N/A	(Schmitt et al., 2017)
Tachycineta bicolor	Tree swallow	Defensin	AvBD7	5	3	N/A	N/A	(Schmitt et al., 2017)
Tachycineta bicolor	Tree swallow	Defensin	AvBD8	8	5	N/A	N/A	(Schmitt et al., 2017)
Tachycineta bicolor	Tree swallow	Defensin	AvBD12	6	4	N/A	N/A	(Schmitt et al., 2017)
Tachycineta bicolor	Tree swallow	Defensin	AvBD13	6	3	N/A	N/A	(Schmitt et al., 2017)

Table A2-2. The coordinates of the AMPs used in this study.

Scaffold	Start	Stop	Gene Name	Strand	Full or Partial Gene
MSTS01000076.1	3071717	3071904	PhciCATH1_1	+	Full Gene
MSTS01000076.1	3073918	3074025	PhciCATH1_2	+	Full Gene
MSTS01000076.1	3074567	3074651	PhciCATH1_3	+	Full Gene
MSTS01000076.1	3075216	3075339	PhciCATH1_4	+	Full Gene
MSTS01000076.1	1695768	1695957	PhciCATH2_1	-	Full Gene
MSTS01000076.1	1694901	1695012	PhciCATH2_2	-	Full Gene
MSTS01000076.1	1694583	1694652	PhciCATH2_3	-	Full Gene
MSTS01000076.1	1693201	1693341	PhciCATH2_4	-	Full Gene
MSTS01000076.1	3086470	3086656	PhciCATH3_1	-	Full Gene
MSTS01000076.1	3085505	3085613	PhciCATH3_2	-	Full Gene
MSTS01000076.1	3084576	3084658	PhciCATH3_3	-	Full Gene
MSTS01000076.1	3083935	3084039	PhciCATH3_4	-	Full Gene
MSTS01000076.1	1783559	1783748	PhciCATH5_1	-	Full Gene
MSTS01000076.1	1781925	1782036	PhciCATH5_2	-	Full Gene
MSTS01000076.1	1781601	1781670	PhciCATH5_3	-	Full Gene
MSTS01000076.1	1780494	1780599	PhciCATH5_4	-	Full Gene
MSTS01000076.1	3029446	3029621	PhciCATH6_1	+	Full Gene
MSTS01000076.1	3032103	3032204	PhciCATH6_2	+	Full Gene
MSTS01000076.1	3033683	3033773	PhciCATH6_3	+	Full Gene
MSTS01000076.1	3034798	3034918	PhciCATH6_4	+	Full Gene
MSTS01000076.1	1715530	1715719	PhciCATH7_1	-	Full Gene
MSTS01000076.1	1713927	1714035	PhciCATH7_2	-	Full Gene
MSTS01000076.1	1713597	1713669	PhciCATH7_3	-	Full Gene
MSTS01000076.1	1712358	1712487	PhciCATH7_4	-	Full Gene
MSTS01000076.1	1737174	1737360	PhciCATH8_1	-	Full Gene
MSTS01000076.1	1735546	1735657	PhciCATH8_2	-	Full Gene
MSTS01000076.1	1735219	1735291	PhciCATH8_3	-	Full Gene
MSTS01000076.1	1734029	1734131	PhciCATH8_4	-	Full Gene
MSTS01000002.1	6003379	6003493	PhciDEFA1	+	Exon 2
MSTS01000002.1	6010703	6010874	PhciDEFA2_1	-	Full Gene
MSTS01000002.1	6008846	6008959	PhciDEFA2_2	-	Full Gene
MSTS01000002.1	5638382	5638442	PhciDEFB1_1	+	Full Gene
MSTS01000002.1	5645821	5645965	PhciDEFB1_2	+	Full Gene
MSTS01000002.1	5618000	5618063	PhciDEFB3_1	+	Full Gene
MSTS01000002.1	5619994	5620168	PhciDEFB3_2	+	Full Gene
MSTS01000040.1	10495293	10495455	PhciDEFB30	-	Exon 2
MSTS01000055.1	6614407	6614515	PhciDEFB4	+	Exon 2
MSTS01000002.1	5927334	5927472	PhciDEFB5_1	-	Full Gene
MSTS01000002.1	5924399	5924549	PhciDEFB5_2	-	Full Gene
MSTS01000002.1	6115224	6115359	PhciDEFB6	-	Exon 2
MSTS01000055.1	6763028	6763184	PhciDEFB7	+	Exon 2

Scaffold	Start	Stop	Gene Name	Strand	Full or Partial Gene
MSTS01000002.1	5762806	5762929	PhciDEFB8	-	Exon 2
MSTS01000002.1	6318426	6318591	PhciDEFB9	+	Exon 2
MSTS01000002.1	5695314	5695374	PhciDEFB10_1	-	Full Gene
MSTS01000002.1	5685952	5686093	PhciDEFB10_2	-	Full Gene
MSTS01000002.1	6291719	6291893	PhciDEFB11	+	Exon 2
MSTS01000040.1	10441725	10442028	PhciDEFB12	-	Exon 2
MSTS01000002.1	5854693	5854816	PhciDEFB13_2	-	Partial
MSTS01000002.1	5911412	5911472	PhciDEFB14_1	-	Full Gene
MSTS01000002.1	5907629	5907758	PhciDEFB14_2	-	Full Gene
MSTS01000002.1	5524143	5524203	PhciDEFB15_1	-	Full Gene
MSTS01000002.1	5512981	5513203	PhciDEFB15_2	-	Full Gene
MSTS01000002.1	6126286	6126343	PhciDEFB16_1	-	Full Gene
MSTS01000002.1	6123441	6123576	PhciDEFB16_2	-	Full Gene
MSTS01000002.1	6066307	6066442	PhciDEFB17	+	Exon 2
MSTS01000002.1	5795543	5795666	PhciDEFB18	-	Exon 2
MSTS01000002.1	6208019	6208139	PhciDEFB19	+	Exon 2
MSTS01000002.1	5793051	5793291	PhciDEFB20	-	Exon 2
MSTS01000040.1	10497494	10497644	PhciDEFB21	-	Exon 2
MSTS01000002.1	6231139	6231328	PhciDEFB22	+	Exon 2
MSTS01000040.1	10411009	10411207	PhciDEFB23	-	Exon 2
MSTS01000055.1	6616531	6616645	PhciDEFB24	+	Exon 2
MSTS01000002.1	39870321	39870483	PhciDEFB25	+	Exon 2
MSTS01000002.1	6339046	6339220	PhciDEFB26	+	Exon 2
MSTS01000002.1	6185071	6185206	PhciDEFB27	-	Exon 2
MSTS01000002.1	6253028	6253190	PhciDEFB28	+	Exon 2
MSTS01000040.1	10486271	10486511	PhciDEFB29	-	Exon 2

Table A2-3. Amino acid sequences for the genes used in this study.

Gene	Sequence
PhciCATH1	MEHLRKALLLASVATLIPTQAFPLSSLSYEQALSTAIHFYNEVHRGENAFRLLQTYSPSSNKDPQEQTLKR VNFTLKETVCPMTEDLVLYQCDFKTDGLVKECQGSLSNEQGIAAIILTCDPVAPEPSRFRRALFPRRRKG SNKPGKYSVLFAAKPSVGKTPHILTI
PhciCATH2	MEKGWTMWLSPLPLLLLGLMTPFATAQSLSYLDLVNRFIDNYNKKSISSNLFQLLVLNLQPEANNDPAT PRALNFTMMETVCPKTKQHNLVECRFKKKGVVKQCSGTISLDATQPSINISCGGPEDIKSGGFLHRIIRSF ANFIHQKYRILLDKYRKLQDIFSGSGDKV
PhciCATH3	MEPLRKLLLLASAAAVLPTRVLPQPSLSYEKALSAAIYFYNQGPGRENAFRVFQVHSFPSIQPLQEQTQKF LSFTLKETVCPVTEELLLDQCDFKTDGLVKECQVSVSNEQDMAAIILTCNQVPPEPLRFKRIRCLNGRKC NYHNLLLTIVPHWRIPKGK
PhciCATH5	MQSGWAMQVALLVLGLLSLMTPLVYARDRRYQDLVNEFIQEYNTKSGSENLFRLSILNLQSGENNDAA APRLLSFTMRETVCPNTENRNPDECDFKENGVVKECLGAIALDSPKPSADISCDGPEKIKRGGIWKLIRPL GRGAGRILRHFHIDFCGNC
PhciCATH6	MASTWRVLLLLGLATAVIALPRRKLTFRDASVLAARRFNGNLNEGAKYRVLVSSLQTPDSPLVLPLTFRI KETECPSSGLQNPETCAFKENGLEKNCTAKFTRLTRFGLGSVECQDVGNNNLVRFKRSASSGIIDTSSLPP KIRQIYNQAVYDTLVGILRNF
PhciCATH7	MQRGWTVQVALLVLGLLSLMTPLTCAQDQRYQDLVNRFIQEYNTKSGSENLFRLSILNLQSGENNDPA APPLLSFTIRETVCPNTENRDPDECDFKENGVVKECLGAIALDSPKPSANISCDGPEKTKRRKFFRSIRKRI KKLRKSIKKRLKKLPFEVPLVFSIPF

Gene	Sequence
PhciCATH8	MQSGWAMQVPLLVLGLLSLKNPFIYAQEQRYQDLVNELIQEYNKKSGSENLFRLSILNLPSGQNNDAAA PRLLSFTMRETVCPNTENRNPDECDFKENGVVKECLGAIALDSPKPSANISCDGPEKIKRRKKKGWKNV GKFINKVLKHFNIDICLNC
PhciDEFA1	DPQQRKNIVCYCRSSCLSRELSSGSCCISGVYYRLCCR
PhciDEFA2	MKSLCLLFALLLLAAQATAEKPADDIGLNSQAEDAKQTQVKVNSEGPQKFFLPSHTFGRTRGLTACSCF CQQRCGYAESRWGSCGSKLLCCRCR
PhciDEFB1	MRILYLMFVVVFVIFSSAGKVNDMVKDSKSCWKAEAICTVHRCMPPNIFKGHCSRSALFCCLLADTFR
PhciDEFB3	MRILYLVFMMLFVIFSCFGTGGYGGYFDNFVCLIRGGSCEKDLCTFPLVQKATCHKRRWHCCFYRPND FITGRPDIDRF
PhciDEFB30	GKSTCWNQKGFCRGQCRNKERFYIFCLNGKRCCVKPSYIPKDILEGTLDTKSKT
PhciDEFB4	DYSKKCRLAKGTCKQLCSEYETSVSYCIRPSMVCCI
PhciDEFB5	MRFFCFLFAFLFLFFQVQSFPGSQNETTQEDVAEVVEKTNFEDGEGDNPLQRSSSVNNAQKCKEISGLCR NGMCPWNENKLGSCGFAKPCCKRLRF
PhciDEFB6	GLLSRISSLICKVRKGKCRTIACTSKEEKIGTCSLGRRKCCRKKK
PhciDEFB7	RCGLDDGVLDTSTCWKVLGHCRVMCKEDEMQVGLCPTPKKMCCIYQPVIGDD
PhciDEFB8	DKAKDSRACKMAGGSCALICTPFLYDFGTCNNGRIMCCIHR
PhciDEFB9	SISMYHGSRLCAQLQGVCRKDICDTIEERIGRCTTHKSCCRKWWLSSFMRTPEPM
PhciDEFB10	MGNLYLVFMVLFVIFSYAETGFGPGFDTFKCFLAAGYCTKNPCKVSAIKRGTCFRKRETCCKSTKKD
PhciDEFB11	SIRLWASRMCAHLHGTCRKQFCNPTEENLRPCSKQKKCCQFLRGWSPVPTPEHKTKST

Gene	Sequence
PhciDEFB12	AGGWGKKKCWNNVGRCRHHCKSNEKYHSRCPNTKKCCLPRNKLSKDDSEWLVLRSSTHSAGRETRP SARARTPPAGAAPPGAGGAKSHSLYVPLPQFPPPP
PhciDEFB13	GADKETYSCLAKHGLCTRLCSDKFMERGTCFRGMLKCCLPH
PhciDEFB14	MRIYYLMFMVLFVISFAGTGYGQDSDTFNCYVNGGSCFYFTCRKKYRSTGTCYNGKAKCCKYK
PhciDEFB15	MRILYLVFMVLFVIFSDASADHVSTLDTTKCYHKHGICVYGSCQNIGRPKVTCFNKKAKCCKFKKFGQR KYKLKTMVRGDSGICKHRRIEDGRT
PhciDEFB16	MRLPSLMLLVFAILCQVLSVDGKKSSIPSCYFYRGICRNKKEIKCMPLPGRCPSSKKHCCRKGL
PhciDEFB17	GSARGFSYERPCYLRGGVCLKQGTPGCVPFKGPCREFTACCKRKN
PhciDEFB18	GRDKDTSECIKSGGSCAVICGEDYYPDGSCYNEQLMCCFPI
PhciDEFB19	PKICATCRLGRGKCRRKCKTDEIVSGSCKQSMLCCRKRIL
PhciDEFB20	LVMGGSCVQTCSLNYYEIGSCQNGKFKCCVHGRVREQKMKKIEMYERKVTLILIKETLSGNQGFFNTEI TPGSFLFLLKW
PhciDEFB21	CWKGAGTCRTFCTRKEVFLYFCKDNSMCCAYSFKMRKPEPEPKPENSQAT
PhciDEFB22	SLSTGGLDKLNDLSTLLLCLVRSGIKEICWKPTGTCRKKCHDNEIHVSRCTRGRKCCLPANVQ
PhciDEFB23	GEKKCWNNAGLCWDHCKSFEKKHSLCPNKRKRCLPKDKFPKYTTEPRNPPNPLPQQEEKHLPAPEP
PhciDEFB24	DAYNIRRCQRFLGRCKTDCESDEYEYGMCIKWRLQCCI
PhciDEFB25	GKIWGNDTMVCFDKQGNCYDICPRWKKQIGSCAEKVLKCCVLKEMKQKQKKVKG
PhciDEFB26	AHIKAARCWAGLGRCRKTCKSTEIHFLLCQQVTLCCIHKKLVDIDIPSPTLDRSLIRP
PhciDEFB27	AGAGFLDEKCQKYQGRCVSQCRKNEELAALCNKFQKCCKLMEPCQ

Gene	Sequence
PhciDEFB28	DLSPSILCAARNESQSFCWTVGGACQKQCLPGEFILEKCIANQFCCLGQRMSRP
PhciDEFB29	GETWFPEKCWNGKGRCRIICISDEVYFSRCENRKKCCLPPHVQTIPSVVIDDQLLYPSSFPFTEETITTTN KQLNGTGR

Table A2-4. Predicted active peptide sequences for the genes used in this study.

Gene	Sequence
PhciCATH1	LFPRRRKGSNKPGKYSVLFAAKPSVGKTPHILTI
PhciCATH2	NFIHQKYRILLDKYRKLQDIFSGSGDKV
PhciCATH3	PPEPLRFKRIRCLNGRKCNYHNLLLTIVPHWRIPKGK
PhciCATH5	KRGGIWKLIRPLGRGAGRILRHFHIDFCGNC
PhciCATH6	ASSGIIDTSSLPPKIRQIYNQAVYDTLVGILRNF
PhciCATH7	NISCDGPEKTKRRKFFRSIRKRIKKLRKSIKKRLKKLPFEVPLVFSIPF
PhciCATH8	NISCDGPEKIKRRKKKGWKNVGKFINKVLKHFNIDICLNC
PhciDEFA1	NIVCYCRSSCLSRELSSGSCCISGVYYRLCCR
PhciDEFA2	GLTACSCFCQQRCGYAESRWGSCGSKLLCCRCR
PhciDEFB1	DSKSCWKAEAICTVHRCMPPNIFKGHCSRSALFCCLLADTFR
PhciDEFB3	DNFVCLIRGGSCEKDLCTFPLVQKATCHKRRWHCCFYRPNDFITGRPDIDRF
PhciDEFB30	GKSTCWNQKGFCRGQCRNKERFYIFCLNGKRCCVKPSYIPKDILEGTLDTKSKT
PhciDEFB4	YSKKCRLAKGTCKQLCSEYETSVSYCIRPSMVCCI
PhciDEFB5	NAQKCKEISGLCRNGMCPWNENKLGSCGFAKPCCKRLRF
PhciDEFB6	SSLICKVRKGKCRTIACTSKEEKIGTCSLGRRKCCRKKK
PhciDEFB7	DTSTCWKVLGHCRVMCKEDEMQVGLCPTPKKMCCIYQPVIGDD
PhciDEFB8	DSRACKMAGGSCALICTPFLYDFGTCNNGRIMCCIHR
PhciDEFB9	GSRLCAQLQGVCRKDICDTIEERIGRCTTHKSCCRKWWLSSFMRTPEPM
PhciDEFB10	DTFKCFLAAGYCTKNPCKVSAIKRGTCFRKRETCCKSTKKD
PhciDEFB11	ASRMCAHLHGTCRKQFCNPTEENLRPCSKQKKCCQFLRGWSPVPTPEHKTKST
PhciDEFB12	GKKKCWNNVGRCRHHCKSNEKYHSRCPNTKKCCLPRNKLSKDDSEWLVLRSSTHSAGRETRPSARARTPPA GAAPPGAGGAKSHSLYVPLPQFPPPP
PhciDEFB13	ETYSCLAKHGLCTRLCSDKFMERGTCFRGMLKCCLPH
PhciDEFB14	DTFNCYVNGGSCFYFTCRKKYRSTGTCYNGKAKCCKYK

Gene	Sequence
PhciDEFB15	DTTKCYHKHGICVYGSCQNIGRPKVTCFNKKAKCCKFKKFGQRKYKLKTMVRGDSGICKHRRIEDGRT
PhciDEFB16	SIPSCYFYRGICRNKKEIKCMPLPGRCPSSKKHCCRKGL
PhciDEFB17	YERPCYLRGGVCLKQGTPGCVPFKGPCREFTACCKRKN
PhciDEFB18	DTSECIKSGGSCAVICGEDYYPDGSCYNEQLMCCFPI
PhciDEFB19	PKICATCRLGRGKCRRKCKTDEIVSGSCKQSMLCCRKRIL
PhciDEFB20	MGGSCVQTCSLNYYEIGSCQNGKFKCCVHGRVREQKMKKIEMYERKVTLILIKETLSGNQGFFNTEITPGSFL FLLKW
PhciDEFB21	CWKGAGTCRTFCTRKEVFLYFCKDNSMCCAYSFKMRKPEPEPKPENSQAT
PhciDEFB22	IKEICWKPTGTCRKKCHDNEIHVSRCTRGRKCCLPANVQ
PhciDEFB23	GEKKCWNNAGLCWDHCKSFEKKHSLCPNKRKRCLPKDKFPKYTTEPRNPPNPLPQQEEKHLPAPEP
PhciDEFB24	NIRRCQRFLGRCKTDCESDEYEYGMCIKWRLQCCI
PhciDEFB25	DTMVCFDKQGNCYDICPRWKKQIGSCAEKVLKCCVLKEMKQKQKKVKG
PhciDEFB26	KAARCWAGLGRCRKTCKSTEIHFLLCQQVTLCCIHKKLVDIDIPSPTLDRSLIRP
PhciDEFB27	LDEKCQKYQGRCVSQCRKNEELAALCNKFQKCCKLMEPCQ
PhciDEFB28	SQSFCWTVGGACQKQCLPGEFILEKCIANQFCCLGQRMSRP
PhciDEFB29	FPEKCWNGKGRCRIICISDEVYFSRCENRKKCCLPPHVQTIPSVVIDDQLLYPSSFPFTEETITTTNKQLNGTGR

Table A2-5. CNVs that overlapped with regions containing AMPs.

AMP gene within CNV	Individual	Region	CNV type	scaffold	start	stop	length	Norm RD	eval1	eval2	eval3	eval4	q0
PhciCATH5	Armidale_F_M50273	M_NSW	duplication	MSTS01000076.1	1771001	1792000	21000	1.97153	0.000286	0.00363	0.001009	0.049321	0.04719
PhciCATH5	Armidale_M_M7070	M_NSW	duplication	MSTS01000076.1	1771001	1793000	22000	2.46317	0.000133	9.68E-11	0.00031	5.70E-09	0.038943
PhciCATH5	Dubbo_M_C11085	M_NSW	duplication	MSTS01000076.1	1772001	1793000	21000	1.88963	2.28E-06	3.08E-28	2.90E-06	2.67E-65	0.037801
PhciCATH5	GUH_F_DECC137F	M_NSW	duplication	MSTS01000076.1	1779001	1793000	14000	1.54009	0.00195	1.59E-20	0.02724	2.41E-16	0.063041
PhciCATH5	GUH_F_DECC162	M_NSW	duplication	MSTS01000076.1	1773001	1794000	21000	2.01617	1.29E-06	3.78E-09	4.31E-06	1.91E-07	0.077541
PhciCATH5	GUH_M_USYD040	M_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.97564	1.50E-06	5.11E-17	2.47E-05	1.92E-14	0.078157
PhciCATH5	GUH_M_USYD098M	M_NSW	duplication	MSTS01000076.1	1779001	1794000	15000	2.04772	0.000242	2.04E-39	0.000643	1.79E-42	0.080478
PhciCATH5	Liverpool_Plains_F_48964	M_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.86357	1.36E-06	9.48E-64	9.51E-06	1.41E-63	0.038442
PhciCATH5	Liverpool_Plains_F_49019	M_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.38965	0.014139	6.53E-05	0.043817	0.001511	0.044332
PhciCATH5	Liverpool_Plains_F_M48961	M_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	2.05338	4.85E-06	4.44E-23	3.14E-05	6.74E-20	0.038989
PhciCATH5	Liverpool_Plains_F_M48974	M_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.88767	4.80E-07	3.40E-26	4.18E-06	1.06E-22	0.038071
PhciCATH5	Liverpool_Plains_M_48967	M_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.43019	4.83E-05	4.33E-05	0.000318	0.001045	0.052332

AMP gene within CNV	Individual	Region	CNV type	scaffold	start	stop	length	Norm RD	eval1	eval2	eval3	eval4	q0
PhciCATH5	Liverpool_Plains_M_49001	M_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.76608	2.72E-05	2.12E-21	0.000188	2.19E-18	0.038547
PhciCATH5	Liverpool_Plains_M_M46161	M_NSW	duplication	MSTS01000076.1	1774001	1793000	19000	2.10478	0.000174	5.36E-14	0.00109	1.32E-11	0.043838
PhciCATH5	Liverpool_Plains_M_M48983	M_NSW	duplication	MSTS01000076.1	1772001	1793000	21000	1.83171	2.72E-07	7.59E-29	2.79E-07	7.49E-54	0.034027
PhciCATH5	Liverpool_Plains_M_M49020_001	M_NSW	duplication	MSTS01000076.1	1772001	1793000	21000	1.7985	3.42E-08	1.91E-24	3.84E-08	2.76E-21	0.03816
PhciCATH5	Liverpool_Plains_M_M49021_001	M_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.82727	2.09E-08	4.50E-27	4.20E-07	1.71E-23	0.038183
PhciCATH5	Liverpool_Plains_U_46160	M_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	2.03377	1.66E-07	1.22E-50	4.47E-06	1.05E-44	0.04009
PhciCATH5	Pilliga_M_M47035	M_NSW	duplication	MSTS01000076.1	1771001	1792000	21000	1.81332	0.000344	8.90E-07	0.001431	2.67E-05	0.042053
PhciCATH5	Pilliga_M_M47420	M_NSW	duplication	MSTS01000076.1	1771001	1793000	22000	1.55343	0.000312	5.88E-12	0.002355	4.47E-10	0.042619
PhciCATH5	Broadwater_M_M50338	N_NSW	duplication	MSTS01000076.1	1771001	1793000	22000	1.63374	0.000168	4.85E-09	0.000789	2.00E-07	0.046728
PhciCATH5	Byron_F_M50503	N_NSW	duplication	MSTS01000076.1	1771001	1793000	22000	2.02694	1.28E-07	2.30E-38	3.27E-07	4.40E-34	0.044715
PhciCATH5	Byron_M_M50373_001	N_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	2.45159	4.92E-09	5.03E-40	1.78E-07	3.78E-35	0.032255
PhciCATH5	Byron_M_M50393_001	N_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.8789	1.08E-09	3.92E-61	3.86E-08	3.80E-54	0.044197
PhciCATH5	Byron_M_M50435	N_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.91678	1.79E-05	1.40E-34	9.80E-05	3.01E-30	0.039746

AMP gene within CNV	Individual	Region	CNV type	scaffold	start	stop	length	Norm RD	eval1	eval2	eval3	eval4	q0
PhciCATH5	Kyogle_F_50433	N_NSW	duplication	MSTS01000076.1	1771001	1793000	22000	2.00672	3.10E-05	4.62E-11	0.000156	2.91E-09	0.035044
PhciCATH5	Lismore_F_2111152	N_NSW	duplication	MSTS01000076.1	1771001	1793000	22000	2.02227	0.000166	9.51E-18	0.001029	2.43E-15	0.033756
PhciCATH5	Lismore_F_M50340	N_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	2.23251	2.04E-06	2.31E-64	5.35E-06	1.98E-64	0.039493
PhciCATH5	Lismore_M_2111192	N_NSW	duplication	MSTS01000076.1	1771001	1792000	21000	1.90159	8.37E-05	9.95E-10	0.000299	5.70E-08	0.037316
PhciCATH5	Lismore_M_M50352	N_NSW	duplication	MSTS01000076.1	1772001	1793000	21000	1.6003	0.000272	5.63E-12	0.001166	5.28E-10	0.048169
PhciCATH5	Lismore_M_M50397	N_NSW	duplication	MSTS01000076.1	1771001	1794000	23000	1.99003	0.000187	1.64E+08	8.78E-06	3.16E-16	0.039073
PhciCATH5	Lismore_M_M50535	N_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	2.16561	8.33E-07	1.04E-37	7.06E-06	4.60E-33	0.037021
PhciCATH5	Fraser_Coast_F_51032	N_QLD	duplication	MSTS01000076.1	1771001	1792000	21000	2.33889	0.004082	3.44E-09	0.031802	1.75E-07	0.057241
PhciCATH5	Fraser_Coast_M_77266	N_QLD	duplication	MSTS01000076.1	1773001	1793000	20000	2.05346	1.08E-05	1.23E-57	0.000167	5.34E-51	0.049536
PhciCATH5	Gympie_F_92626	N_QLD	duplication	MSTS01000076.1	1773001	1793000	20000	1.48807	7.88E-06	6.98E-23	0.000161	1.01E-19	0.055548
PhciCATH5	Gympie_F_93694	N_QLD	duplication	MSTS01000076.1	1773001	1793000	20000	1.96989	2.91E-08	2.94E-62	1.01E-06	3.70E-55	0.046443
PhciCATH5	Gympie_M_95464	N_QLD	duplication	MSTS01000076.1	1771001	1793000	22000	1.53094	0.000901	1.76E-05	0.00294	0.000345	0.042013
PhciCATH5	South_Burnett_F_86919	N_QLD	duplication	MSTS01000076.1	1771001	1792000	21000	2.08998	0.000245	1.88E-07	0.000721	6.54E-06	0.048419
PhciCATH5	Sunshine_Coast_M_76877	N_QLD	duplication	MSTS01000076.1	1773001	1793000	20000	2.16569	1.63E-05	1.13E-32	7.68E-05	1.57E-28	0.042718

AMP gene within CNV	Individual	Region	CNV type	scaffold	start	stop	length	Norm RD	eval1	eval2	eval3	eval4	q0
PhciCATH5	Blue_Mountains_F_7094923	S_NSW	duplication	MSTS01000076.1	1773001	1792000	19000	1.55234	2.69E-05	2.57E-19	7.48E-05	2.30E-16	0.057075
PhciCATH5	Blue_Mountains_F_7094976	S_NSW	duplication	MSTS01000076.1	1773001	1792000	19000	1.45683	4.67E-06	5.46E-15	2.86E-05	1.71E-12	0.047323
PhciCATH5	Blue-Mountains_M_7041193	S_NSW	duplication	MSTS01000076.1	1771001	1793000	22000	1.97895	0.000202	6.24E-20	0.000737	2.52E-17	0.042024
PhciCATH5	Blue-Mountains_M_7041200	S_NSW	duplication	MSTS01000076.1	1771001	1792000	21000	2.01411	0.000403	2.01E-12	0.001504	2.08E-10	0.040367
PhciCATH5	Campbelltown_F_C11K2	S_NSW	duplication	MSTS01000076.1	1771001	1793000	22000	1.88507	0.000253	7.04E-07	0.000752	1.85E-05	0.04171
PhciCATH5	Campbelltown_F_C12K2	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.79225	1.19E-08	2.29E-84	5.62E-08	4.68E-75	0.044929
PhciCATH5	Campbelltown_F_C12K4	S_NSW	duplication	MSTS01000076.1	1773001	1794000	21000	1.80026	4.48E-06	5.78E+07	4.43E-08	2.84E-88	0.044643
PhciCATH5	Campbelltown_F_C2K4	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.8082	7.89E-08	3.37E-97	8.50E-07	1.32E-86	0.045493
PhciCATH5	Campbelltown_F_C3K2	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.75419	8.70E-09	4.07E-65	1.22E-07	9.87E-58	0.043017
PhciCATH5	Campbelltown_F_C5K7	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.7214	3.92E-07	5.10E-33	3.83E-06	7.63E-29	0.036754
PhciCATH5	Campbelltown_F_C6K5	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.84995	8.23E-08	5.15E-68	5.04E-07	2.43E-60	0.041759
PhciCATH5	Campbelltown_F_C7K2	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.81493	2.84E-08	3.15E-67	3.40E-07	1.24E-59	0.044618
PhciCATH5	Campbelltown_M_C10K1	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.86052	2.11E-08	6.16E-86	1.70E-07	1.81E-76	0.036106

AMP gene within CNV	Individual	Region	CNV type	scaffold	start	stop	length	Norm RD	eval1	eval2	eval3	eval4	q0
PhciCATH5	Campbelltown_M_C10K4	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.89256	2.02E-09	9.60477E-110	1.77E-08	6.76E-98	0.047196
PhciCATH5	Campbelltown_M_C11K1	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.79392	9.26E-08	5.05E-44	4.26E-07	9.53E-39	0.04012
PhciCATH5	Campbelltown_M_C2K2	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.79826	6.25E-07	8.25E-43	6.15E-06	1.18E-37	0.0445
PhciCATH5	Campbelltown_M_C3K1	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.84009	1.20E-08	9.14E-90	5.26E-08	5.54E-94	0.044419
PhciCATH5	Campbelltown_M_C3K3	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.75828	5.02E-07	3.09E-47	7.54E-06	1.22E-41	0.040189
PhciCATH5	Campbelltown_M_C5K8	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.84851	1.31E-07	9.04E-62	5.74E-07	1.01E-54	0.038995
PhciCATH5	Campbelltown_M_C6K1	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.8777	1.10E-07	1.26E-63	7.70E-07	2.16E-56	0.038754
PhciCATH5	Campbelltown_M_Marble	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.92358	1.40E-08	1.26E-13	3.30E-07	2.17E-11	0.036657
PhciCATH5	Dubbo_M_C11212	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.79627	2.45E-09	5.23E-78	6.03E-08	2.47E-69	0.042399
PhciCATH5	Dubbo_M_C11358	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.80293	3.59E-09	3.01E-65	6.78E-08	7.53E-58	0.041505
PhciCATH5	Dubbo_M_C11413	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.86558	0	##########	1.42E-10	##########	0.049494
PhciCATH5	Dubbo_M_C11512	S_NSW	duplication	MSTS01000076.1	1771001	1793000	22000	1.97218	3.59E-06	7.37E-29	1.03E-05	1.93E-25	0.037549
PhciCATH5	Monaro_F_Alex	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.83073	1.08E-09	3.39E-30	4.78E-08	2.65E-26	0.036997
PhciCATH5	Monaro_F_Annie	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.84497	6.18E-08	1.58E-10	1.68E-06	1.33E-08	0.0366

AMP gene within CNV	Individual	Region	CNV type	scaffold	start	stop	length	Norm RD	eval1	eval2	eval3	eval4	q0
PhciCATH5	Monaro_F_Brandy	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.95152	1.14E-06	3.18E-05	1.30E-05	0.000791	0.08208
PhciCATH5	Monaro_F_CS	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.82883	1.57E-08	0.00288	5.48E-07	0.045633	0.038758
PhciCATH5	Monaro_F_Mia	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.83489	3.24E-08	0.001359	1.43E-06	0.023212	0.031815
PhciCATH5	Monaro_F_Tahlia	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.86874	1.95E-08	1.35E-13	2.14E-07	2.30E-11	0.035918
PhciCATH5	Monaro_M_Drew	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.8802	1.59E-11	1.02E-63	1.59E-09	1.80E-56	0.036758
PhciCATH5	Monaro_M_Evan	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.86111	2.77E-09	3.66E-44	7.99E-08	7.13E-39	0.039952
PhciCATH5	Monaro_M_Hamish	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.92846	1.91E-10	3.91E-46	1.07E-08	1.20E-40	0.038044
PhciCATH5	Monaro_M_Jett	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.89713	5.20E-09	7.33E-39	1.77E-07	4.21E-34	0.034968
PhciCATH5	Monaro_M_Kenny	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.87485	1.59E-08	3.35E-06	5.22E-07	0.000104	0.036179
PhciCATH5	Monaro_M_Malu	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.82476	2.63E-09	7.71E-20	9.11E-08	5.55E-17	0.037312
PhciCATH5	Monaro_M_Matt	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.80078	1.42E-08	8.11E-16	3.95E-07	2.31E-13	0.039266
PhciCATH5	Monaro_M_Murray	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.94465	1.84E-06	2.39E-05	2.53E-05	0.000611	0.073117
PhciCATH5	Monaro_M_Tallow	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.8343	7.35E-08	1.47E-13	2.02E-06	2.49E-11	0.03741
PhciCATH5	Monaro_M_Xavier	S_NSW	duplication	MSTS01000076.1	1779001	1793000	14000	1.59477	0.001994	6.29E-08	0.037395	1.51E-05	0.074066
PhciCATH5	Narrandra_F_N17	S_NSW	duplication	MSTS01000076.1	1775001	1792000	17000	1.55031	0.000216	1.01E-08	0.004497	1.15E-06	0.043811

AMP gene within CNV	Individual	Region	CNV type	scaffold	start	stop	length	Norm RD	eval1	eval2	eval3	eval4	q0
PhciCATH5	Narrandra_F_N2	S_NSW	duplication	MSTS01000076.1	1773001	1792000	19000	1.51387	1.59E-05	6.94E-15	0.000286	2.12E-12	0.045655
PhciCATH5	Narrandra_M_N15	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	2.17721	1.56E-09	1.67E-38	1.94E-08	8.82E-34	0.032419
PhciCATH5	Narrandra_M_N3	S_NSW	duplication	MSTS01000076.1	1771001	1793000	22000	1.93027	1.61E-07	1.99E-19	1.44E-07	7.22E-17	0.035395
PhciCATH5	PTS_F_MidnightR	S_NSW	duplication	MSTS01000076.1	1779001	1792000	13000	1.56943	0.002309	6.36E-27	0.040202	1.94E-21	0.085959
PhciCATH5	PTS_M_Tai	S_NSW	duplication	MSTS01000076.1	1779001	1791000	12000	1.62846	1.10E-05	1.56E-45	0.001153	1.73E-36	0.072652
PhciCATH5	STHD_F_Annie-L	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.92882	4.97E-09	3.14E-16	2.65E-07	9.84E-14	0.0409
PhciCATH5	STHD_F_Arya	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	2.06306	6.17E-09	8.84E-37	3.32E-08	3.15E-32	0.036096
PhciCATH5	STHD_F_Beyonce	S_NSW	duplication	MSTS01000076.1	1772001	1793000	21000	1.45555	6.70E-05	0.001036	0.000116	2.04E-08	0.040775
PhciCATH5	STHD_F_Daenery	S_NSW	duplication	MSTS01000076.1	1772001	1793000	21000	1.99437	1.12E-07	3.73E-23	4.19E-08	4.06E-20	0.039114
PhciCATH5	STHD_F_Falia-F	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.97939	7.59E-09	5.40E-43	2.24E-07	8.04E-38	0.037941
PhciCATH5	STHD_F_Liv-NJ	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.91908	4.78E-07	7.03E-08	4.15E-06	3.22E-06	0.036179
PhciCATH5	STHD_F_Stevie-N	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.96942	3.31E-09	1.70E-26	4.84E-08	5.67E-23	0.037448
PhciCATH5	STHD_F_Yara	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	2.02171	3.82E-10	3.69E-32	1.46E-08	4.53E-28	0.035439
PhciCATH5	STHD_M_Clegane	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.53831	3.59E-07	1.33E-09	5.00E-06	9.09E-08	0.050892
PhciCATH5	STHD_M_Jimmy-B	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.91184	1.59E-10	6.00E-71	1.21E-08	5.58E-63	0.036907

AMP gene within CNV	Individual	Region	CNV type	scaffold	start	stop	length	Norm RD	eval1	eval2	eval3	eval4	q0
PhciCATH5	STHD_M_Qyburn	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.9796	3.16E-09	1.25E-31	1.06E-07	1.36E-27	0.043455
PhciCATH5	STHD_M_Rhaegar	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.97861	7.50E-08	1.27E-08	1.84E-06	6.89E-07	0.037339
PhciCATH5	STHD_M_St-james	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.96765	1.75E-09	6.80E-48	7.86E-08	3.13E-42	0.042854
PhciCATH5	STHD_M_Varys	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.99139	7.51E-08	1.40E-08	2.43E-06	7.53E-07	0.037827
PhciCATH5	STHD_M_Walter	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.94414	1.15E-08	3.25E-33	4.01E-07	5.08E-29	0.038098
PhciCATH5	SW_Sydney_M_Biya	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.8593	3.36E-06	1.15E-25	1.91E-05	3.17E-22	0.033694
PhciCATH5	SW_Sydney_M_Dalang	S_NSW	duplication	MSTS01000076.1	1771001	1793000	22000	1.89623	1.93E-06	2.48E-12	7.54E-06	2.04E-10	0.036282
PhciCATH5	SW_Sydney_M_Giloa	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.88161	2.60E-07	2.44E-55	4.61E-06	6.23E-49	0.037228
PhciCATH5	SW_Sydney_M_Kimber	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.93703	9.58E-09	5.72E-89	1.32E-07	3.37E-79	0.039891
PhciCATH5	SW_Sydney_M_Mirribi	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.95852	1.45E-08	1.09E-71	9.88E-08	1.20E-63	0.035624
PhciCATH5	SW_Sydney_M_Yung	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.85203	1.81E-06	1.36E-36	3.87E-05	4.62E-32	0.048222
PhciCATH5	SWSyd_F_M48896	S_NSW	duplication	MSTS01000076.1	1773001	1793000	20000	1.9916	3.19E-11	3.76E-97	1.17E-09	1.46E-86	0.03979
PhciCATH5	Wollemi_NP_U_M50543_001	S_NSW	duplication	MSTS01000076.1	1773001	1794000	21000	1.95974	9.39E-07	5.67E-08	3.22E-07	2.45E-44	0.039007
PhciCATH5	Toowoomba_F_79193	S_QLD	duplication	MSTS01000076.1	1775001	1793000	18000	1.61149	3.62E-05	6.94E-16	0.000108	3.77E-13	0.053015

AMP gene within CNV	Individual	Region	CNV type	scaffold	start	stop	length	Norm RD	eval1	eval2	eval3	eval4	q0
PhciCATH5	Western_Downs_M_75384	S_QLD	duplication	MSTS01000076.1	1771001	1793000	22000	1.49898	0.00019	3.50E-19	0.000767	1.53E-21	0.044837
PhciCATH5	Cape_Otway_F_Y32	Victoria	duplication	MSTS01000076.1	1771001	1793000	22000	1.91081	4.13E-05	5.03E-19	0.000102	1.68E-16	0.04299
PhciCATH5	Cape_Otway_F_Y37	Victoria	duplication	MSTS01000076.1	1773001	1792000	19000	1.864	3.93E-05	2.52E-29	0.000122	1.88E-33	0.043016
PhciCATH5	Cape_Otway_F_Y49	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	1.86667	6.03E-06	1.30E-39	3.84E-05	8.89E-35	0.036689
PhciCATH5	Gelantipy_F_K11	Victoria	duplication	MSTS01000076.1	1771001	1793000	22000	1.86641	6.65E-06	3.31E-09	3.68E-05	1.41E-07	0.037635
PhciCATH5	Gelantipy_F_K16	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	1.91773	8.73E-07	6.59E-52	5.86E-06	7.63E-46	0.039276
PhciCATH5	Gelantipy_M_K14	Victoria	duplication	MSTS01000076.1	1773001	1794000	21000	1.85872	0.000174	421186	2.13E-05	7.94E-42	0.034367
PhciCATH5	Gelantipy_M_K19	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	1.9143	5.61E-08	7.07E-44	3.99E-07	1.29E-38	0.039839
PhciCATH5	Gippsland_F_SAWS67	Victoria	duplication	MSTS01000076.1	1771001	1792000	21000	2.11156	7.15E-05	2.93E-20	4.59E-05	5.35E-18	0.037401
PhciCATH5	Gippsland_F_SAWS69	Victoria	duplication	MSTS01000076.1	1775001	1793000	18000	1.78542	0.005806	3.75E-08	0.009955	2.82E-06	0.047517
PhciCATH5	Gippsland_F_SAWS72	Victoria	duplication	MSTS01000076.1	1771001	1793000	22000	1.6177	0.003171	0.000321	0.023713	0.004823	0.044995
PhciCATH5	HDS_M_SAWS01	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	1.96976	6.77E-09	1.11E-45	1.78E-07	3.05E-40	0.042136
PhciCATH5	HDS_M_SAWS12	Victoria	duplication	MSTS01000076.1	1772001	1793000	21000	1.995	3.80E-08	4.13E-54	1.24E-08	5.50E-54	0.041759
PhciCATH5	Mallacoota_U_Annie	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	2.06318	5.38E-06	2.10E-50	2.39E-05	1.72E-44	0.046224

AMP gene within CNV	Individual	Region	CNV type	scaffold	start	stop	length	Norm RD	eval1	eval2	eval3	eval4	q0
PhciCATH5	Mallacoota_U_Blinky	Victoria	duplication	MSTS01000076.1	1773001	1792000	19000	1.57979	0.000143	1.61E-14	0.000758	4.51E-12	0.050995
PhciCATH5	Mallacoota_U_Frankie	Victoria	duplication	MSTS01000076.1	1773001	1794000	21000	2.01993	9.09E-05	27447.3	1.87E-05	3.89E-29	0.045319
PhciCATH5	Mallacoota_U_Solo	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	2.02861	1.21E-06	6.07E-40	4.70E-06	4.47E-35	0.042175
PhciCATH5	Mallacoota_U_Trip	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	2.01572	3.42E-05	2.25E-34	0.000145	4.60E-30	0.038698
PhciCATH5	Murray_River_F_F6	Victoria	duplication	MSTS01000076.1	1771001	1793000	22000	1.89764	1.16E-05	1.33E-12	2.06E-05	1.16E-10	0.034112
PhciCATH5	Murray_River_F_F8	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	1.52821	2.65E-05	4.07E-10	0.000316	3.12E-08	0.045745
PhciCATH5	Murray_River_F_FA	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	1.9471	2.65E-09	1.92E-93	1.74E-08	3.16E-83	0.035693
PhciCATH5	Murray_River_M_MR001	Victoria	duplication	MSTS01000076.1	1771001	1793000	22000	1.52059	4.51E-05	3.09E-09	0.000277	1.33E-07	0.040977
PhciCATH5	Murray_River_M_MR003	Victoria	duplication	MSTS01000076.1	1773001	1792000	19000	2.0421	4.04E-07	1.07E-66	2.27E-06	9.30E-59	0.037931
PhciCATH5	Murray_River_M_MR005	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	1.55913	1.99E-05	2.62E-13	0.000237	4.19E-11	0.041947
PhciCATH5	Murray_River_M_MR006	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	1.47648	7.43E-06	5.20E-10	3.68E-05	3.89E-08	0.042043
PhciCATH5	Murray_River_M_MR007	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	1.99117	2.73E-07	2.37E-84	1.22E-06	3.21E-77	0.042317
PhciCATH5	Murray_River_M_MR008	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	1.94781	2.50E-08	2.83E-87	2.62E-07	1.13E-77	0.03389
PhciCATH5	Murray_River_M_MR009	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	1.96491	6.22E-07	6.48E-68	5.32E-06	3.00E-60	0.035858
PhciCATH5	Murray_River_M_MR010	Victoria	duplication	MSTS01000076.1	1771001	1793000	22000	1.91215	3.68E-07	5.92E-25	1.51E-06	6.84E-22	0.04011

AMP gene within CNV	Individual	Region	CNV type	scaffold	start	stop	length	Norm RD	eval1	eval2	eval3	eval4	q0
PhciCATH5	SHBS_F_SAWS04	Victoria	duplication	MSTS01000076.1	1773001	1794000	21000	2.06036	5.04E-08	8.94E-51	1.21E-07	4.16E-45	0.055064
PhciCATH5	SHBS_M_SAWS37	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	1.61126	3.98E-06	1.55E-11	4.92E-05	1.65E-09	0.047844
PhciCATH5	SHGLD_M_SAWS27	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	2.07651	5.74E-10	1.07E-64	1.39E-08	2.36E-57	0.03593
PhciCATH5	SHGLD_M_SAWS33	Victoria	duplication	MSTS01000076.1	1773001	1793000	20000	1.5728	1.98E-07	9.59E-07	6.08E-06	3.38E-05	0.048326
PhciDEFB23	Liverpool_Plains_M_48973	M_NSW	deletion	MSTS01000040.1	1040001	1042000	21000	0.471598	1.44E-07	5.65E-18	2.12E-06	1.97E-15	0.468777
PhciDEFB23	Liverpool_Plains_M_M48983	M_NSW	deletion	MSTS01000040.1	1040001	1042000	17000	0.426221	7.55E-07	2.06E-34	1.99E-05	2.45E-29	0.494565
PhciDEFB23	Broadwater_M_50439	N_NSW	deletion	MSTS01000040.1	1040001	1042000	21000	0.520264	1.28E-06	6.48E-15	1.22E-05	1.16E-12	0.493817
PhciDEFB23	Broadwater_M_50508	N_NSW	deletion	MSTS01000040.1	1040001	1042000	17000	0.436605	2.32E-07	6.07E-28	5.45E-06	1.25E-23	0.498744
PhciDEFB23	Byron_F_M50431	N_NSW	deletion	MSTS01000040.1	1040001	1042000	20000	0.541963	0.000574	0.00215	0.001554	0.007427	0.426506
PhciDEFB23	Byron_M_M50430	N_NSW	deletion	MSTS01000040.1	1040001	1042000	23000	0.527599	1.32E-06	0.001191	4.34E-07	1.46E-11	0.467485
PhciDEFB23	Kyogle_F_50433	N_NSW	deletion	MSTS01000040.1	1040001	1042000	21000	0.507868	2.91E-06	2.16E-06	1.13E-05	5.96E-05	0.497797
PhciDEFB23	Kyogle_F_M50471	N_NSW	deletion	MSTS01000040.1	1040001	1042000	22000	0.546791	6.76E-06	0.002494	1.54E-05	5.56E-05	0.44
PhciDEFB23	Kyogle_M_50384	N_NSW	deletion	MSTS01000040.1	1040001	1042000	22000	0.497009	1.07E-06	0.001345	1.44E-06	1.11E-08	0.441052
PhciDEFB23	Lismore_F_M50336	N_NSW	deletion	MSTS01000040.1	1040001	1042000	22000	0.510555	3.05E-06	0.001917	9.78E-06	0.024516	0.495912

AMP gene within CNV	Individual	Region	CNV type	scaffold	start	stop	length	Norm RD	eval1	eval2	eval3	eval4	q0
PhciDEFB23	Lismore_M_2111271	N_NSW	deletion	MSTS01000040.1	10404001	10426000	22000	0.486717	3.63E-07	1.30E-08	6.05E-07	1.60E-10	0.49718
PhciDEFB23	Northern-Rivers_M_M50329	N_NSW	deletion	MSTS01000040.1	10404001	10426000	22000	0.507729	3.71E-06	5.29E-07	8.02E-06	1.43E-05	0.499504
PhciDEFB23	Northern-Rivers_M_M50450	N_NSW	deletion	MSTS01000040.1	10404001	10426000	22000	0.494167	2.19E-07	2.81E-11	3.00E-07	1.10E-11	0.492654
PhciDEFB23	Tweed_Heads_M_57043	N_NSW	deletion	MSTS01000040.1	10403001	10425000	22000	0.508855	3.22E-06	1.15E-08	1.74E-05	4.40E-07	0.473325
PhciDEFB23	FarNorthQLD_K3	N_QLD	deletion	MSTS01000040.1	10402001	10426000	24000	0.53694	6.67E-06	1.21E-07	3.31E-05	2.80E-06	0.449335
PhciDEFB23	Fraser_Coast_M_93594	N_QLD	deletion	MSTS01000040.1	10403001	10426000	23000	0.53242	1.49E-07	3.75E-13	1.20E-06	3.00E-11	0.492193
PhciDEFB23	Noosa_M_79211	N_QLD	deletion	MSTS01000040.1	10402001	10426000	24000	0.561768	3.64E-05	2.88E-11	0.00011	1.34E-09	0.485528
PhciDEFB23	South_Burnett_F_86009	N_QLD	deletion	MSTS01000040.1	10403001	10426000	23000	0.541372	7.69E-06	9.66E-05	5.80E-05	0.001435	0.484261
PhciDEFB23	Sunshine_Coast_M_76877	N_QLD	deletion	MSTS01000040.1	10403001	10426000	23000	0.510171	4.85E-07	1.23E-08	9.08E-07	3.98E-07	0.46163
PhciDEFB23	Dubbo_M_C11212	S_NSW	deletion	MSTS01000040.1	10404001	10425000	21000	0.479519	2.51E-05	4.24E-17	0.00031	1.22E-14	0.474806
PhciDEFB23	Monaro_M_Kenny	S_NSW	deletion	MSTS01000040.1	10404001	10425000	21000	0.494895	8.46E-06	2.89E-13	7.56E-05	3.60E-11	0.49017
PhciDEFB23	STHD_F_Annie-L	S_NSW	deletion	MSTS01000040.1	10408001	10425000	17000	0.452009	7.90E-05	9.11E-35	0.00071	1.19E-29	0.4881
PhciDEFB23	Hidden-Vale_F_Claire	S_QLD	deletion	MSTS01000040.1	10408001	10426000	18000	0.45651	4.78E-06	1.93E-06	2.74E-05	9.36E-05	0.482415
PhciDEFB23	Toowoomba_F_79192	S_QLD	deletion	MSTS01000040.1	10403001	10426000	23000	0.525923	1.88E-07	1.57E-07	9.27E-07	4.08E-06	0.46087

AMP gene within CNV	Individual	Region	CNV type	scaffold	start	stop	length	Norm RD	eval1	eval2	eval3	eval4	q0
PhciDEFB23	Cape_Otway_F_Y60	Victoria	deletion	MSTS01000040.1	10405001	10426000	21000	0.482132	2.44E-06	7.67E-06	1.24E-05	0.000187	0.491781
PhciDEFB24	Clarence_Valley_M_M50344_001	M_NSW	duplication	MSTS01000055.1	6615001	6628000	13000	1.91145	4.93E-05	5.55E-72	6.40E-06	1.769633E-148	0.00591
PhciDEFB24	Liverpool_Plains_F_M48974	M_NSW	duplication	MSTS01000055.1	6616001	6629000	13000	1.67193	0.007958	5.70E-14	0.01525	2.84E-12	0.007608
PhciDEFB24	Pilliga_F_M47418	M_NSW	duplication	MSTS01000055.1	6616001	6630000	14000	1.64208	0.006557	6.24E-15	0.022846	4.26E-13	0.01069
PhciDEFB24	Pilliga_M_M47420	M_NSW	duplication	MSTS01000055.1	6615001	6629000	14000	1.78703	0.030339	0.004702	0.028568	4.56E-17	0.007161
PhciDEFB24	Port-Macquarie_F_81926	M_NSW	duplication	MSTS01000055.1	6615001	6631000	16000	2.27235	0.029194	1.27E-05	0.007209	1.14E-26	0.008858
PhciDEFB24	Port-Macquarie_F_81996	M_NSW	duplication	MSTS01000055.1	6616001	6630000	14000	2.58847	0.010687	2.34E-31	0.041753	1.25E-25	0.004539
PhciDEFB24	FarNorthQLD_K3	N_QLD	duplication	MSTS01000055.1	6615001	6627000	12000	1.82808	0.000853	2.27E-18	1.47E-05	7.05E-74	0.005857
PhciDEFB24	Gold_Coast_M_54980	S_QLD	duplication	MSTS01000055.1	6615001	6630000	15000	2.71807	0.021131	1.43E-12	0.006024	7.33E-85	0.004717
PhciDEFB24	Redland_F_95776	S_QLD	duplication	MSTS01000055.1	6616001	6630000	14000	2.42449	0.021009	2.35E-28	0.035226	1.90E-38	0.009329

Table A2-6. Haplotype frequencies for each allele in each geographic region.

Allele	North QLD	South QLD	North NSW	Mid NSW	South NSW	Victoria
PhciCATH1_Hap1	1.000	1.000	1.000	0.921	0.969	1.000
PhciCATH1_Hap2	0.000	0.000	0.000	0.079	0.031	0.000
PhciCATH2_Hap1	0.043	0.157	0.145	0.156	0.098	0.000
PhciCATH2_Hap2	0.053	0.000	0.072	0.119	0.339	0.007
PhciCATH2_Hap3	0.000	0.000	0.000	0.083	0.058	0.000
PhciCATH2_Hap4	0.021	0.000	0.000	0.000	0.000	0.000
PhciCATH2_Hap5	0.755	0.833	0.783	0.633	0.504	0.993
PhciCATH2_Hap6	0.128	0.009	0.000	0.009	0.000	0.000
PhciCATH3_Hap1	0.766	0.796	0.961	0.904	0.830	1.000
PhciCATH3_Hap2	0.000	0.000	0.000	0.096	0.036	0.000
PhciCATH3_Hap3	0.234	0.204	0.039	0.000	0.134	0.000
PhciCATH5_Hap1	0.851	0.833	0.770	0.816	0.531	0.993
PhciCATH5_Hap2	0.021	0.009	0.000	0.000	0.000	0.000
PhciCATH5_Hap3	0.000	0.000	0.000	0.105	0.192	0.000
PhciCATH5_Hap4	0.128	0.157	0.230	0.079	0.277	0.007
PhciCATH6_Hap1	0.681	0.593	0.382	0.868	0.911	1.000
PhciCATH6_Hap2	0.043	0.000	0.000	0.009	0.000	0.000
PhciCATH6_Hap3	0.277	0.407	0.618	0.044	0.058	0.000
PhciCATH6_Hap4	0.000	0.000	0.000	0.079	0.031	0.000
PhciCATH7_Hap1	0.989	1.000	1.000	0.921	0.929	1.000
PhciCATH7_Hap2	0.000	0.000	0.000	0.079	0.058	0.000
PhciCATH7_Hap3	0.011	0.000	0.000	0.000	0.013	0.000
PhciCATH8_Hap1	0.734	0.787	0.684	0.526	0.643	1.000
PhciCATH8_Hap2	0.000	0.000	0.000	0.061	0.147	0.000
PhciCATH8_Hap3	0.000	0.000	0.020	0.000	0.009	0.000
PhciCATH8_Hap4	0.032	0.009	0.000	0.193	0.063	0.000
PhciCATH8_Hap5	0.191	0.093	0.171	0.018	0.063	0.000
PhciCATH8_Hap6	0.000	0.000	0.000	0.079	0.058	0.000
PhciCATH8_Hap7	0.000	0.000	0.007	0.009	0.004	0.000
PhciCATH8_Hap8	0.000	0.093	0.072	0.105	0.013	0.000
PhciCATH8_Hap9	0.021	0.019	0.046	0.009	0.000	0.000
PhciDEFA1_Hap1	0.021	0.111	0.342	0.114	0.000	0.000
PhciDEFA1_Hap2	0.372	0.380	0.388	0.386	0.058	0.049
PhciDEFA1_Hap3	0.000	0.000	0.000	0.009	0.219	0.931
PhciDEFA1_Hap4	0.606	0.509	0.270	0.491	0.661	0.021
PhciDEFA2_Hap1	0.000	0.000	0.000	0.018	0.000	0.000
PhciDEFA2_Hap2	1.000	1.000	1.000	0.982	1.000	1.000
PhciDEFB3_Hap1	0.585	0.454	0.513	0.982	0.973	1.000
PhciDEFB3_Hap2	0.362	0.500	0.480	0.018	0.027	0.000
PhciDEFB3_Hap3	0.053	0.046	0.007	0.000	0.000	0.000
PhciDEFB4_Hap1	0.596	0.685	0.684	0.719	0.563	0.833
PhciDEFB4_Hap2	0.340	0.194	0.118	0.009	0.071	0.000
PhciDEFB4_Hap3	0.000	0.000	0.000	0.026	0.344	0.167

Allele	North QLD	South QLD	North NSW	Mid NSW	South NSW	Victoria
PhciDEFB4_Hap4	0.064	0.120	0.197	0.246	0.022	0.000
PhciDEFB5_Hap1	0.936	0.963	1.000	1.000	1.000	1.000
PhciDEFB5_Hap2	0.064	0.037	0.000	0.000	0.000	0.000
PhciDEFB7_Hap1	0.894	0.944	0.987	0.965	1.000	1.000
PhciDEFB7_Hap2	0.085	0.046	0.013	0.035	0.000	0.000
PhciDEFB7_Hap3	0.021	0.009	0.000	0.000	0.000	0.000
PhciDEFB8_Hap1	0.947	1.000	1.000	1.000	1.000	1.000
PhciDEFB8_Hap2	0.053	0.000	0.000	0.000	0.000	0.000
PhciDEFB9_Hap1	0.553	0.639	0.803	0.456	0.379	0.056
PhciDEFB9_Hap2	0.011	0.000	0.020	0.000	0.000	0.000
PhciDEFB9_Hap3	0.053	0.102	0.000	0.333	0.054	0.347
PhciDEFB9_Hap4	0.383	0.259	0.178	0.211	0.567	0.597
PhciDEFB10_Hap1	0.819	0.880	0.961	0.526	0.647	1.000
PhciDEFB10_Hap2	0.181	0.120	0.039	0.474	0.353	0.000
PhciDEFB11_Hap1	0.436	0.602	0.507	0.789	0.946	0.979
PhciDEFB11_Hap2	0.043	0.000	0.000	0.096	0.000	0.000
PhciDEFB11_Hap3	0.521	0.398	0.493	0.114	0.054	0.021
PhciDEFB12_Hap1	0.340	0.315	0.250	0.439	0.821	0.993
PhciDEFB12_Hap2	0.372	0.324	0.171	0.246	0.058	0.007
PhciDEFB12_Hap3	0.032	0.037	0.151	0.088	0.063	0.000
PhciDEFB12_Hap4	0.021	0.120	0.158	0.211	0.054	0.000
PhciDEFB12_Hap5	0.234	0.204	0.270	0.018	0.004	0.000
PhciDEFB14_Hap1	0.862	0.778	0.704	0.719	0.978	0.958
PhciDEFB14_Hap2	0.138	0.222	0.296	0.281	0.022	0.042
PhciDEFB16_Hap1	0.989	0.917	1.000	1.000	1.000	1.000
PhciDEFB16_Hap2	0.011	0.083	0.000	0.000	0.000	0.000
PhciDEFB18_Hap1	0.989	0.981	0.993	1.000	1.000	1.000
PhciDEFB18_Hap2	0.011	0.019	0.007	0.000	0.000	0.000
PhciDEFB19_Hap1	0.883	0.824	0.974	1.000	0.964	0.590
PhciDEFB19_Hap2	0.011	0.000	0.013	0.000	0.000	0.000
PhciDEFB19_Hap3	0.000	0.000	0.000	0.000	0.036	0.382
PhciDEFB19_Hap4	0.106	0.176	0.013	0.000	0.000	0.028
PhciDEFB20_Hap1	0.404	0.259	0.599	0.737	0.946	0.965
PhciDEFB20_Hap2	0.074	0.009	0.000	0.061	0.004	0.000
PhciDEFB20_Hap3	0.521	0.731	0.401	0.202	0.049	0.035
PhciDEFB24_Hap1	0.894	0.944	0.987	0.965	1.000	1.000
PhciDEFB24_Hap2	0.106	0.056	0.013	0.035	0.000	0.000
PhciDEFB26_Hap1	0.819	0.870	0.770	0.491	0.237	0.167
PhciDEFB26_Hap2	0.043	0.056	0.204	0.447	0.737	0.833
PhciDEFB26_Hap3	0.011	0.019	0.000	0.000	0.000	0.000
PhciDEFB26_Hap4	0.128	0.056	0.026	0.061	0.027	0.000
PhciDEFB27_Hap1	0.936	0.833	0.987	1.000	1.000	0.972
PhciDEFB27_Hap2	0.064	0.167	0.013	0.000	0.000	0.028
PhciDEFB28_Hap1	0.394	0.361	0.395	0.368	0.375	0.333
PhciDEFB28_Hap2	0.606	0.639	0.605	0.632	0.625	0.667

Table A2-7. Frequency of heterozygote individuals for each gene in each region.

AMP	North QLD	South QLD	North NSW	Mid NSW	South NSW	Victoria
PhciCATH1	0.0000	0.0000	0.0000	0.1579	0.0446	0.0000
PhciCATH2	0.3617	0.2593	0.3026	0.4737	0.4286	0.0139
PhciCATH3	0.2979	0.3333	0.0789	0.1228	0.1875	0.0000
PhciCATH5	0.2128	0.2222	0.3553	0.2807	0.4375	0.0139
PhciCATH6	0.3830	0.3704	0.5000	0.2632	0.1429	0.0000
PhciCATH7	0.0213	0.0000	0.0000	0.1579	0.0893	0.0000
PhciCATH8	0.4255	0.3519	0.4737	0.6842	0.4732	0.0000
PhciDEFA1	0.4043	0.6296	0.5789	0.5439	0.4286	0.1111
PhciDEFA2	0.0000	0.0000	0.0000	0.0351	0.0000	0.0000
PhciDEFB1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PhciDEFB10	0.2766	0.2037	0.0789	0.2807	0.3661	0.0000
PhciDEFB11	0.5745	0.4630	0.3026	0.3860	0.1071	0.0417
PhciDEFB12	0.6170	0.6481	0.6974	0.6842	0.2679	0.0139
PhciDEFB13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PhciDEFB14	0.1915	0.3704	0.3816	0.3860	0.0268	0.0833
PhciDEFB15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PhciDEFB16	0.0213	0.1296	0.0000	0.0000	0.0000	0.0000
PhciDEFB17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PhciDEFB18	0.0213	0.0370	0.0132	0.0000	0.0000	0.0000
PhciDEFB19	0.1702	0.2407	0.0526	0.0000	0.0714	0.5556
PhciDEFB20	0.4043	0.3889	0.4868	0.3684	0.0893	0.0417
PhciDEFB21	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PhciDEFB22	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PhciDEFB23	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PhciDEFB24	0.1702	0.0741	0.0263	0.0351	0.0000	0.0000
PhciDEFB25	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PhciDEFB26	0.2766	0.1481	0.3026	0.5614	0.3125	0.2500
PhciDEFB27	0.0851	0.2222	0.0263	0.0000	0.0000	0.0556
PhciDEFB28	0.4894	0.4259	0.4474	0.4561	0.5536	0.3611
PhciDEFB29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PhciDEFB3	0.4043	0.5370	0.4342	0.0351	0.0536	0.0000
PhciDEFB30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PhciDEFB4	0.4681	0.3333	0.3026	0.3509	0.4107	0.2222
PhciDEFB5	0.1277	0.0741	0.0000	0.0000	0.0000	0.0000
PhciDEFB6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PhciDEFB7	0.1915	0.0741	0.0263	0.0351	0.0000	0.0000
PhciDEFB8	0.1064	0.0000	0.0000	0.0000	0.0000	0.0000
PhciDEFB9	0.5532	0.4259	0.2105	0.5263	0.6071	0.6111

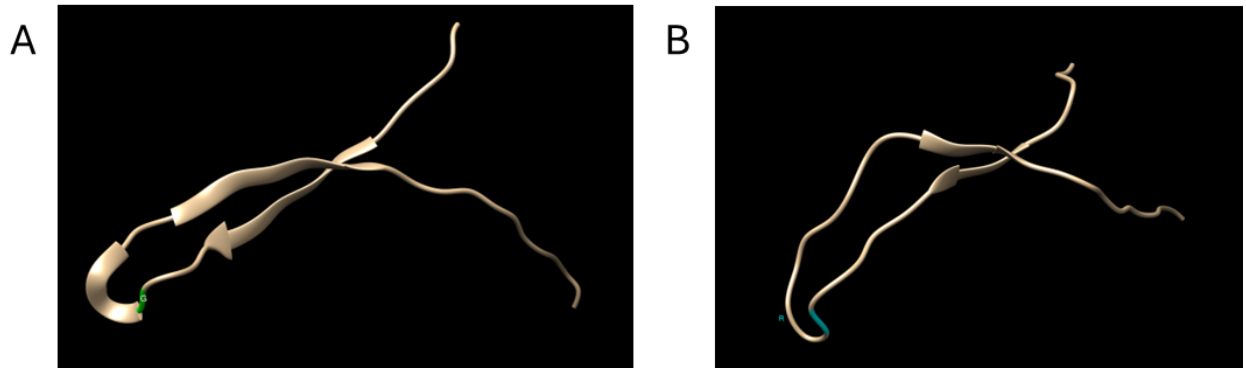


Figure A2-1. 3D visualisations of peptides with non-synonymous SNPs in the active peptide regions. Figures are of the active peptide with amino acid changes highlighted. A. PhciCATH3_Hap1 B. PhciCATH3_Hap2.

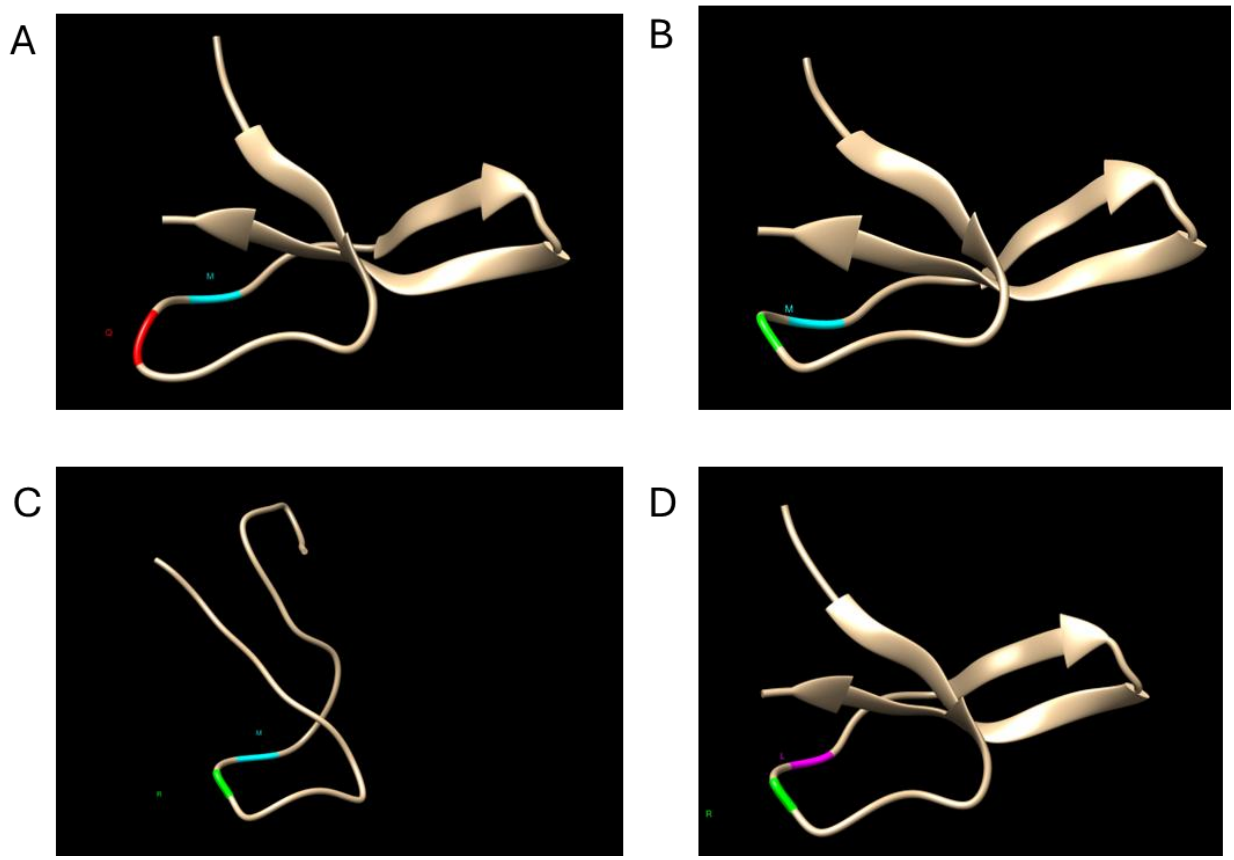


Figure A2-2. 3D visualisations of peptides with non-synonymous SNPs in the active peptide regions. Figures are of the active peptide with amino acid changes highlighted. A. PhciDEFA_Hap1 B. PhciDEFA1_Hap2 C. PhciDEFA1_Hap3 D. PHciDEFA1_Hap4.

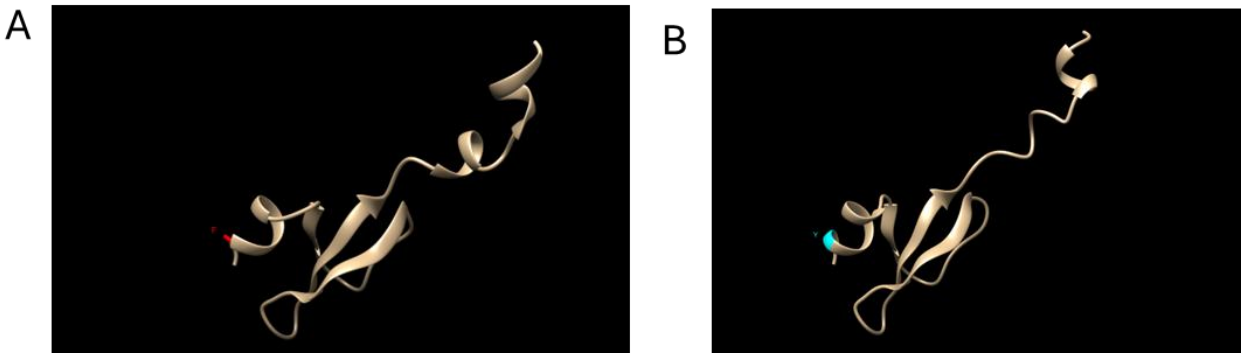


Figure A2-3. 3D visualisations of peptides with non-synonymous SNPs in the active peptide regions. Figures are of the active peptide with amino acid changes highlighted. A. PhciDEFB3_Hap1. B. PhciDEFB3_Hap3.

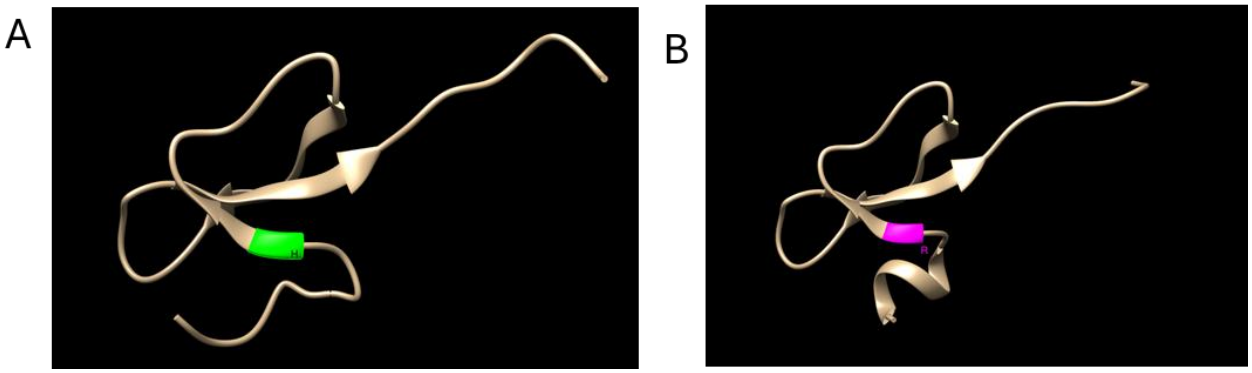


Figure A2-4. 3D visualisations of peptides with non-synonymous SNPs in the active peptide regions. Figures are of the active peptide with amino acid changes highlighted. A. PhciDEFB7_Hap1 B. PhciDEFB7_Hap3.

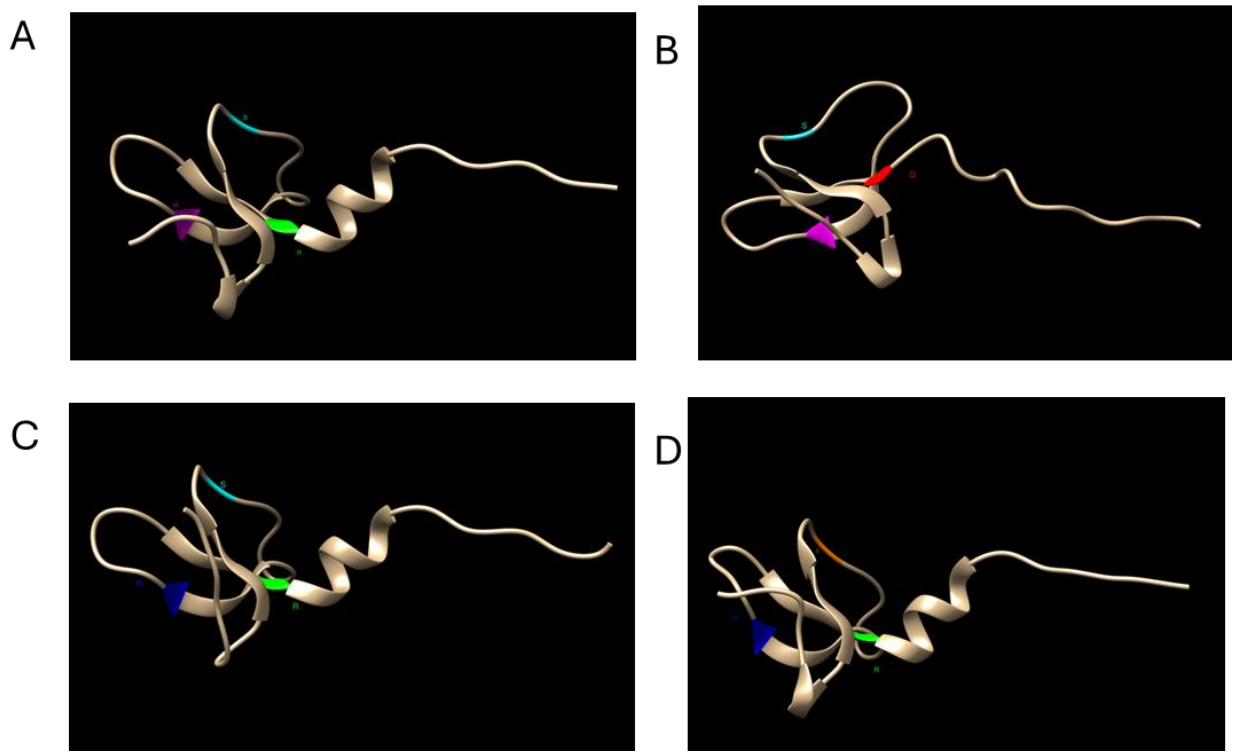


Figure A2-5. 3D visualisations of peptides with non-synonymous SNPs in the active peptide regions. Figures are of the active peptide with amino acid changes highlighted. A. PhciDEFB9_Hap1 B. PhciDEFB9_Hap2 C. PhciDEFB9_Hap3 D. PhciDEFB9_Hap4.

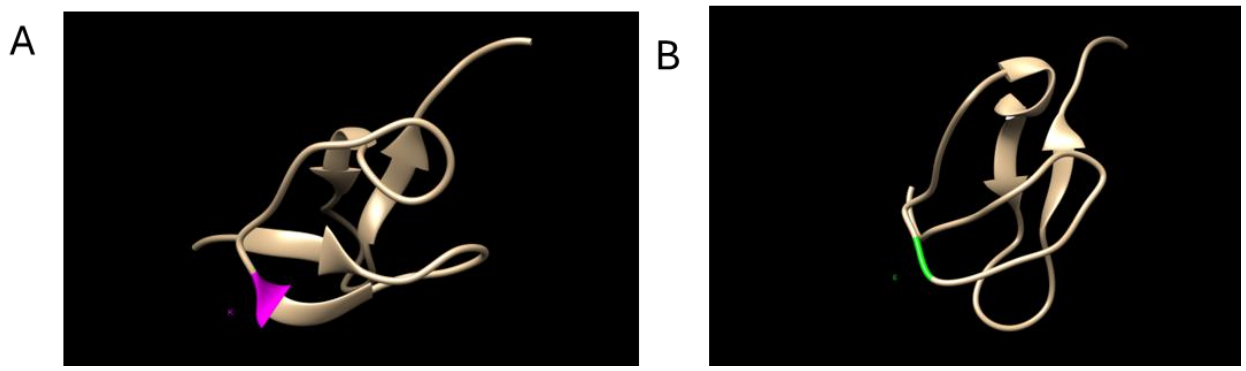


Figure A2-6. 3D visualisations of peptides with non-synonymous SNPs in the active peptide regions. Figures are of the active peptide with amino acid changes highlighted. A. PhciDEFB10_Hap1 B. PhciDEFB10_Hap2.

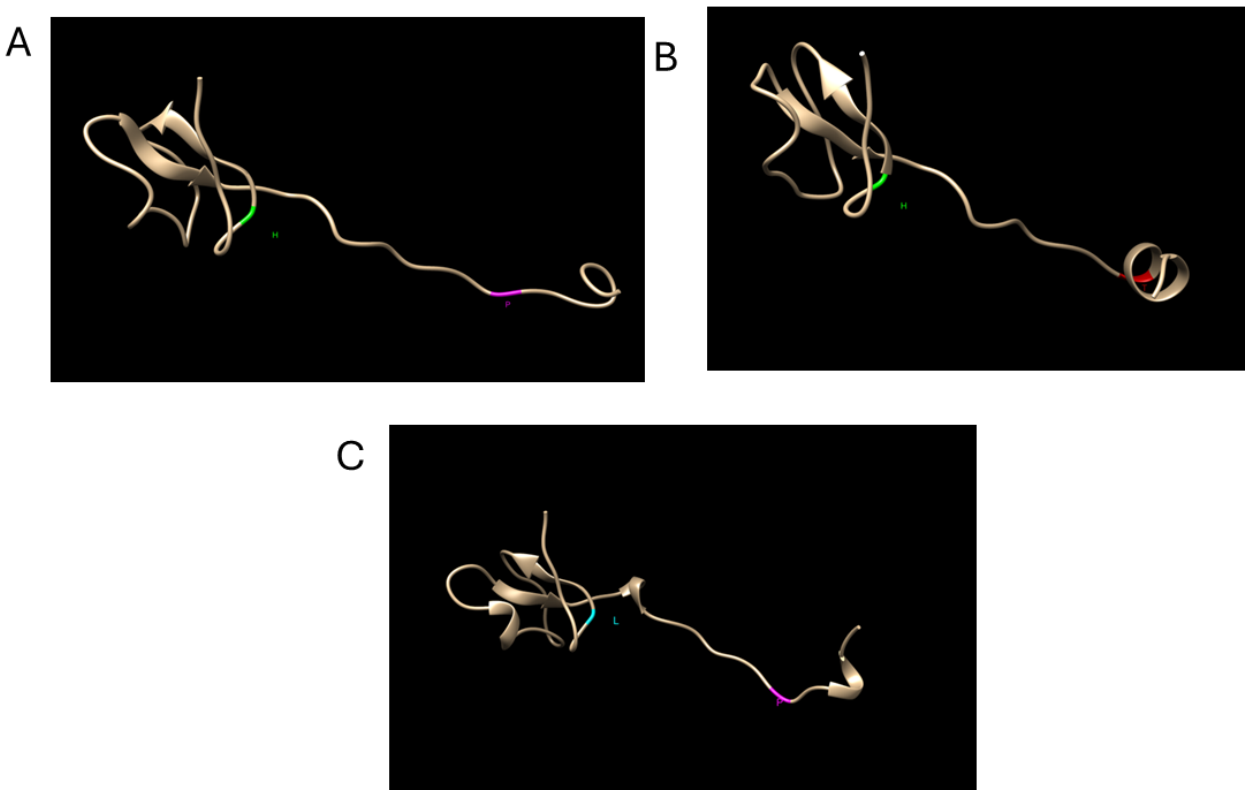


Figure A2-7. 3D visualisations of peptides with non-synonymous SNPs in the active peptide regions. Figures are of the active peptide with amino acid changes highlighted. A. PhciDEFB11_Hap1 B. PhciDEFB11_Hap2 C. PhciDEFB11_Hap3.

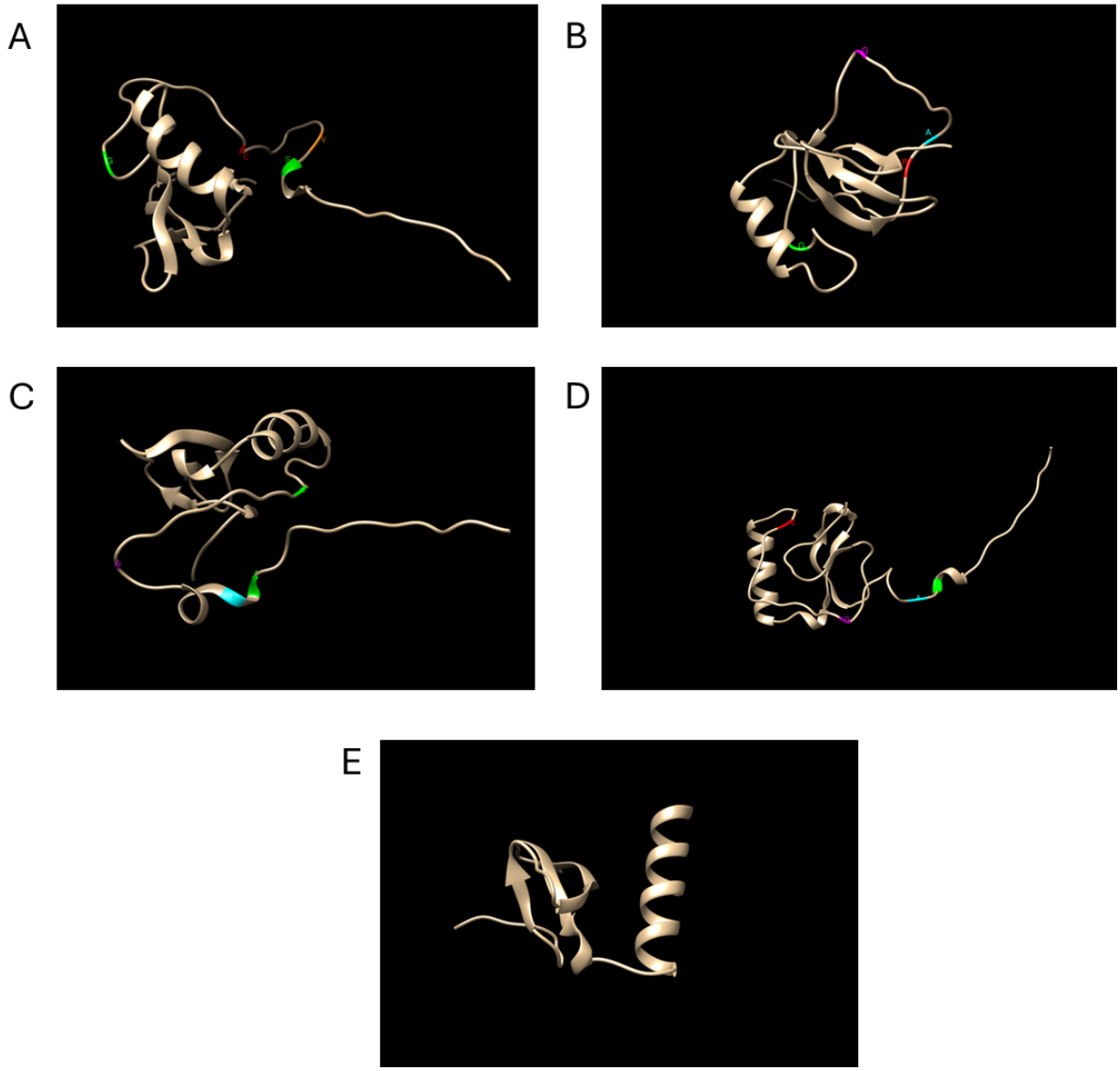


Figure A2-8. 3D visualisations of peptides with non-synonymous SNPs in the active peptide regions. Figures are of the active peptide with amino acid changes highlighted. A. PhciDEFB12_Hap1 B. PhciDEFB12_Hap2 C. PhciDEFB12_Hap3 D. PhciDEFB12_Hap4 E. PhciDEFB12_Hap5.

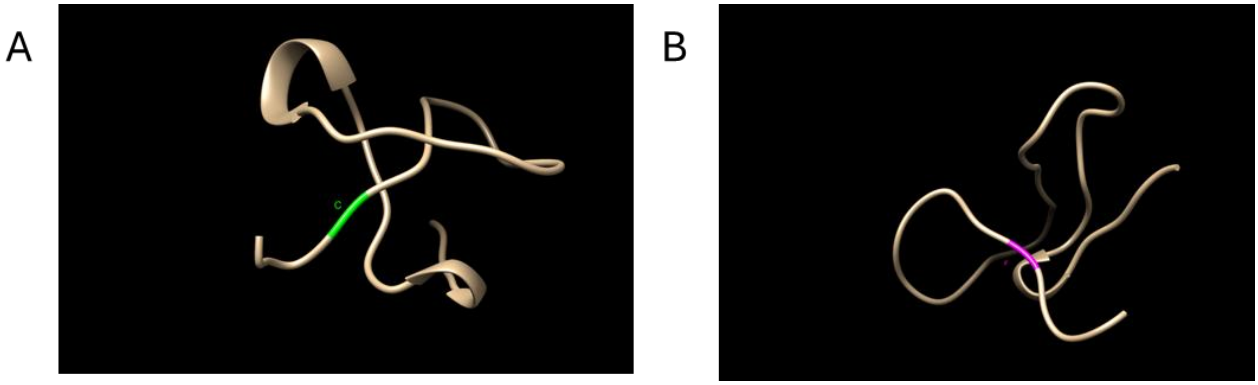


Figure A2-9. 3D visualisations of peptides with non-synonymous SNPs in the active peptide regions. Figures are of the active peptide with amino acid changes highlighted. A. PhciDEFB16_Hap1 B. PhciDEFB16_Hap2.

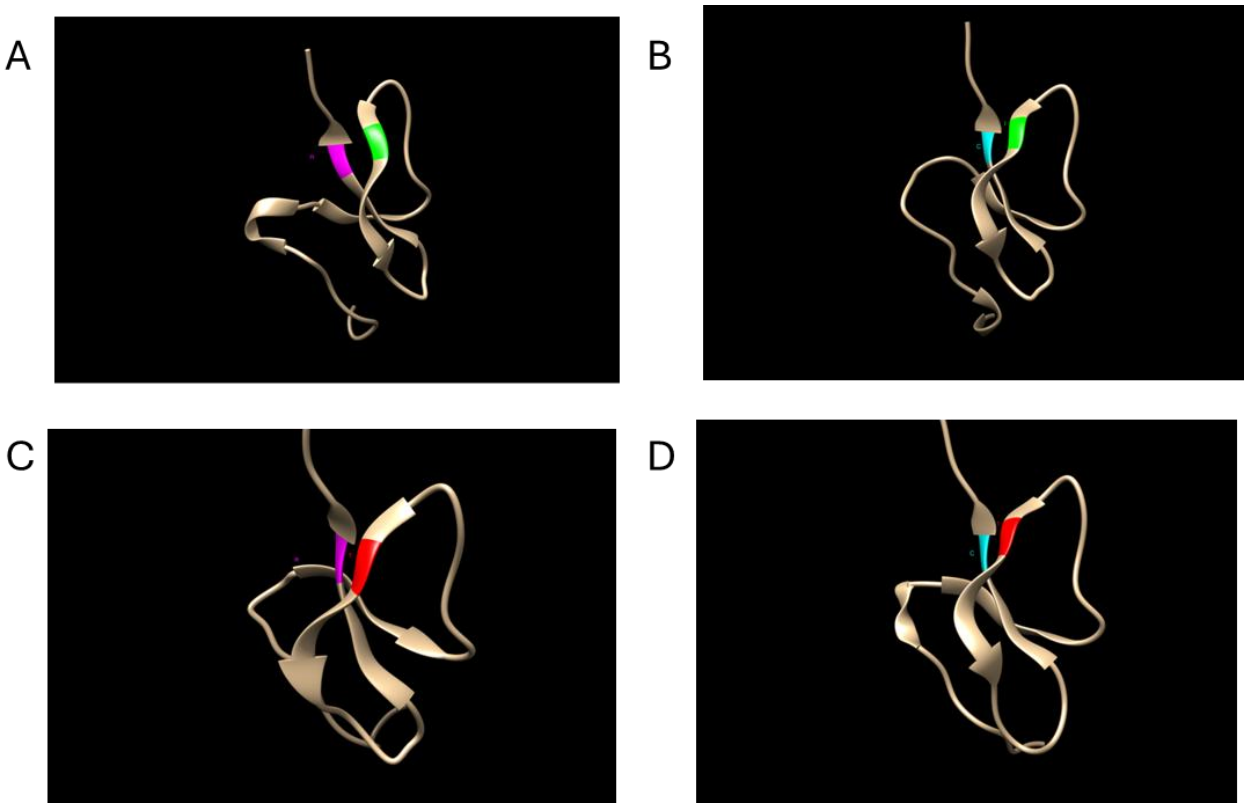


Figure A2-10. 3D visualisations of peptides with non-synonymous SNPs in the active peptide regions. Figures are of the active peptide with amino acid changes highlighted. A PhciDEFB19_Hap1 B. PhciDEFB19_Hap2 C. PhciDEFB19_Hap3 D. PhciDEFB19_Hap4.

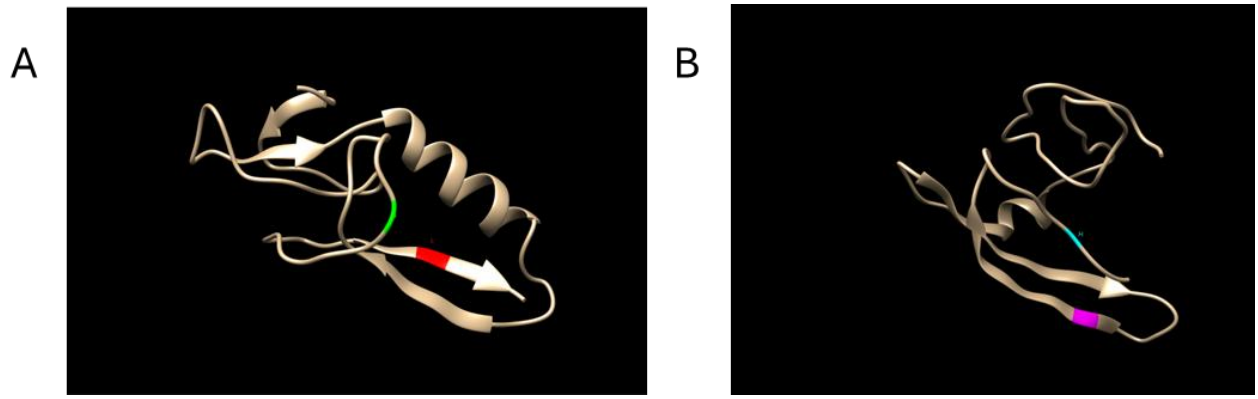


Figure A2-11. 3D visualisations of peptides with non-synonymous SNPs in the active peptide regions. Figures are of the active peptide with amino acid changes highlighted. A PhciDEFB20_Hap1 B. PhciDEFB20_Hap2.

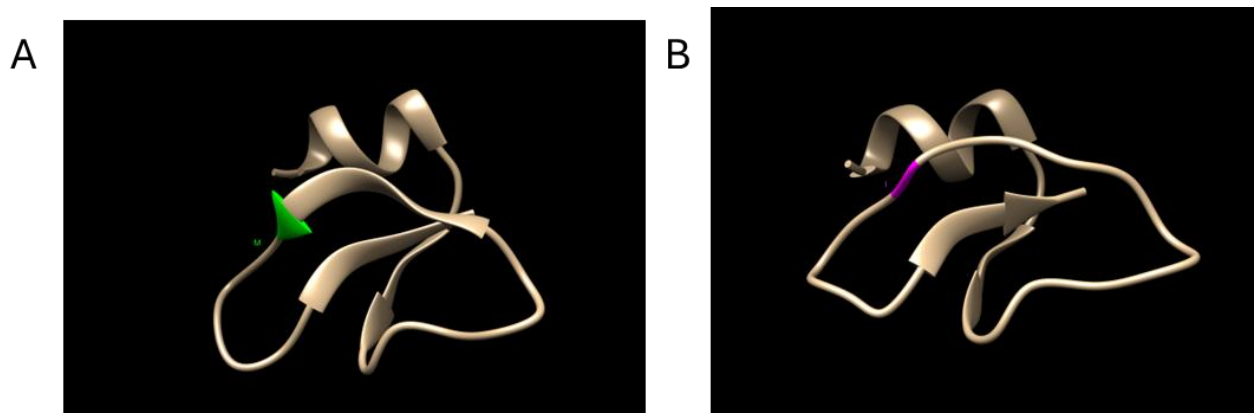


Figure A2-12. 3D visualisations of peptides with non-synonymous SNPs in the active peptide regions. Figures are of the active peptide with amino acid changes highlighted. A PhciDEFB24_Hap1. B PhciDEFB24_Hap2.

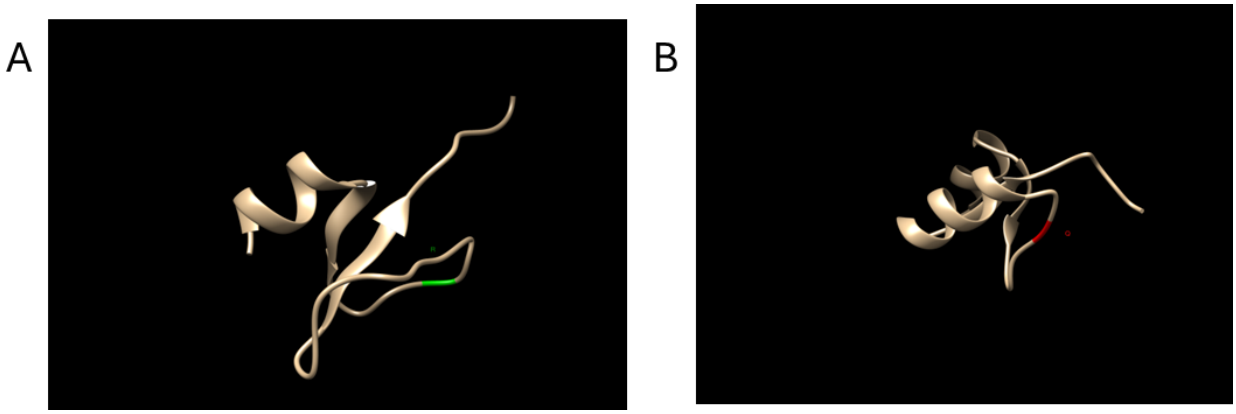


Figure A2-13. 3D visualisations of peptides with non-synonymous SNPs in the active peptide regions. Figures are of the active peptide with amino acid changes highlighted. A. PhciDEFB27_Hap1 B. PhciDEFB27_Hap2.

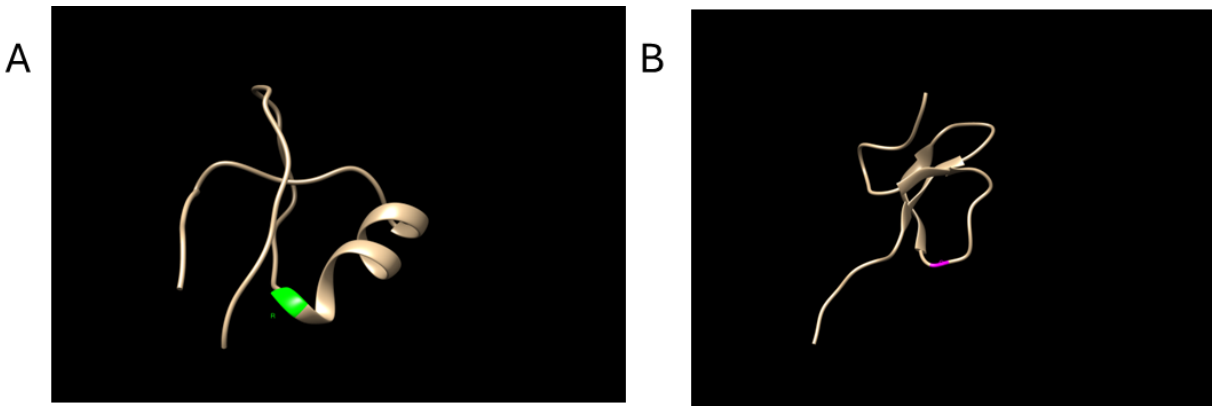


Figure A2-14. 3D visualisations of peptides with non-synonymous SNPs in the active peptide regions. Figures are of the active peptide with amino acid changes highlighted. A PhciDEFB28_Hap1. B. PhciDEFB28_Hap2.

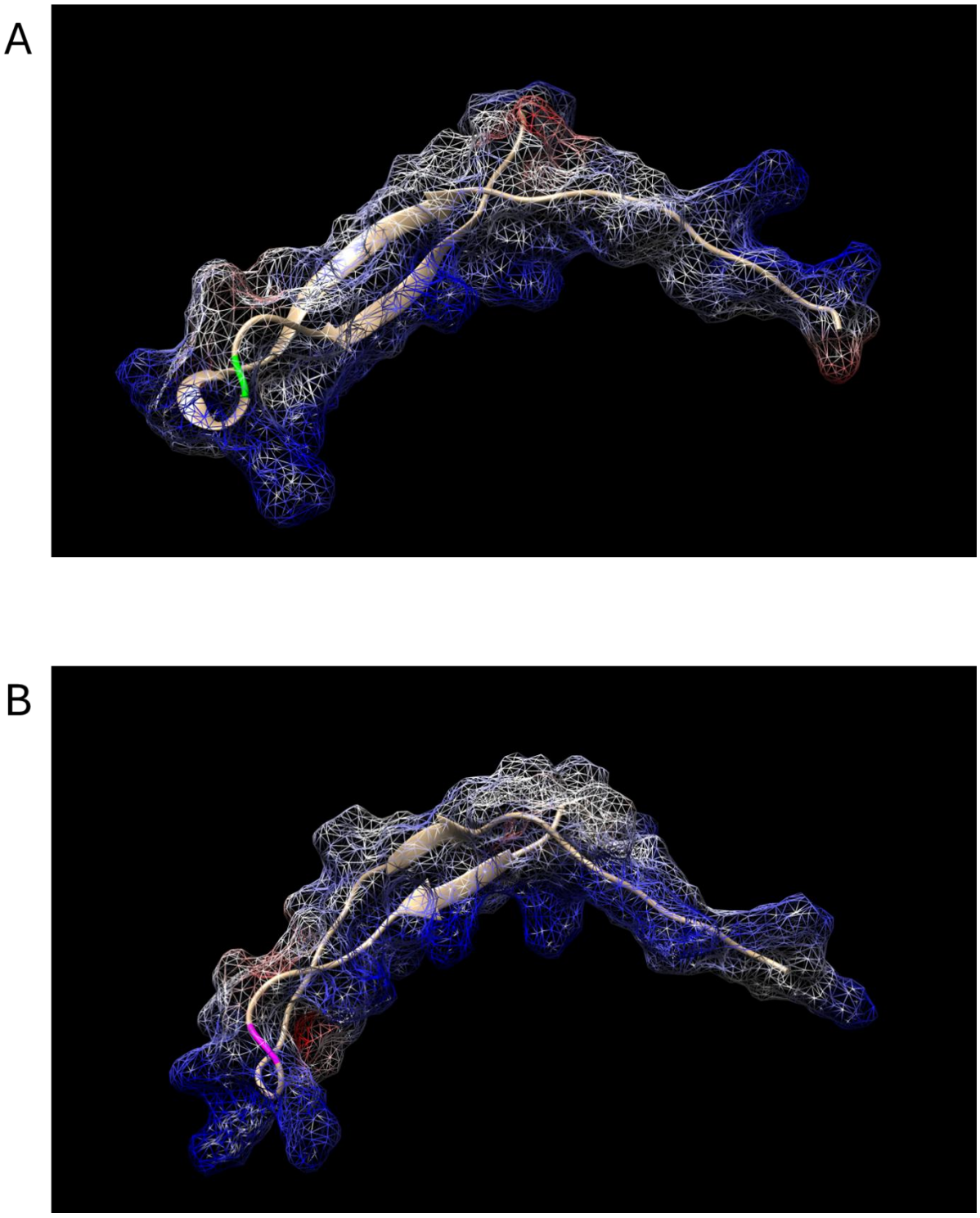


Figure A2-15. 3D visualization of (A) PhciCATH3_Hap1 and (B) PhciCATH3_Hap2 showing surface charge distribution changes with positively charged regions in blue and negatively charged regions in red.

APPENDIX 3. SUPPLEMENTARY MATERIAL TO CHAPTER 4

SUPPLEMENTARY TABLES AND FIGURES

Table A3-1. Metadata for 35 DFT1 samples used in this study. Phylogenomic clade was based on the analysis conducted by Kwon et al. (2020). Strain was determined by cytogenetic analysis conducted by Kwon et al. (2020).

Sample Name	Transponder	Year sampled	Location sampled	Clade	Strain
07_2320	982009100758259	2007	Sorell	Clade A1	Strain 3
07_2296_T2	982009102737669	2007	Sorell	Clade A1	Strain 3
07_2297	982009100322308	2007	Sorell	Clade A1	Strain 3
07_2336_T2	985120015992583	2007	Sorell	Clade A1	Strain 3
08_1451_T2	982009104928332	2008	Kempton	Clade B	Strain 2
08_1963_T2	982009105188074	2008	Coles Bay	Clade A1	No data
08_1573_T1	982009104855230	2008	West Pencil Pine	Clade A2	Strain 2
08_0195_T1	982009104859904	2008	Narawntapu	Clade A2	Strain 1 tetraploid
08_2387_T2	982009105171583	2008	Sorell	Clade A1	Strain 3 evolved
08_2479_T1	982009100799027	2008	Sorell	Clade A1	No data
07_2258	982009100754517	2007	Sorell	Clade A1	Strain 3
11_3909	982000123133145	2011	Bronte	Clade B	No data
11_3918	982000123216973	2011	Bronte	Clade B	Strain 2 evolved
12_2065_T2	982009104841875	2012	West Pencil Pine	Clade A2	Strain 1
12_0705_T2	982009104253781	2012	West Pencil Pine	Clade C	Strain 1
11_2749_T1	982009104785985	2011	West Pencil Pine	Clade C	No data
11_4115_T3	982009106218282	2011	Mount Pleasant	Clade A1	Strain 4
12_0820	982009104719592	2012	West Pencil Pine	Clade C	Strain 1
06_3045	00065DC9C5	2006	Trowunna	Clade B	Strain 2
07_1152	982009100876802	2007	Coles Bay	Clade A2	Strain 2
07_1254_T1	982009102235882	2007	Fentonbury	Clade B	Strain 2
08_1868_T1	985120016082881	2008	Narawntapu	Clade A2	Strain 6a
08_2178	982009104358247	2008	Ringarooma	Clade A2	Strain 2 evolved
08_3033_T1	982009102236127	2008	Narawntapu	Clade A2	Strain 1 diploid and tetraploid
08_3696	982009104798550	2008	Sorell	Clade A1	No data

Sample Name	Transponder	Year sampled	Location sampled	Clade	Strain
09_1196	00065D7B0B	2009	Trowunna	Clade B	No data
12_3045	982000123208272	2012	DPIPWE (x Forestier)	Clade A1	Strain 3
15_1368_T1	982009106485186	2015	Narawntapu	Clade B	No data
15_1524	982000191009681	2015	Fentonbury	Clade B	No data
15_1610	982000363454290	2015	Fentonbury	Clade B	No data
15_2070_T1	982009106575584	2015	Takone	Clade C	No data
982009000000000	982009106207596	Unknown	Takone	Clade A2	No data
12_3856_T2	982000167789148	2012	West Pencil Pine	Clade A2	Strain 1
12_4284_T2	982009106180793	2012	West Pencil Pine	Clade B	Strain 2
12_4284_T3	982009106180793	2012	West Pencil Pine	Clade A2	Strain 2

Table A3-2. Purity estimates for the DFT1 samples used in this analysis based on VAF(het). Purity for the highlighted samples could not be estimated as there were less than 3 somatic variants present.

Sample	VAF(het)	Purity	
06_3045_[B]	0.4387149	87.74298	
07_1152_[A2]	0.4910077	98.20154	
07_1254_T1_[B]	0.4986897	99.73794	
07_2258_[A1]	0.4450971	89.01942	
07_2296_T2_[A1]	0.5183743	103.67486	
07_2297_S23_L003_Op	NA	NA	insufficient variants to calculate; also excluded on the basis of clustering
07_2320_[A1]	0.4713952	94.27904	
07_2336_T2_[A1]	0.4475268	89.50536	
08_0195_T1_[A2]	0.4290557	85.81114	
08_1451_T2_[B]	0.3249441	64.98882	
08_1573_T1_[A2]	0.4783602	95.67204	
08_1868_T1_[A2]	0.4176126	83.52252	
08_1963_T2_[A1]	0.4308512	86.17024	
08_2178_[A2]	0.4099615	81.9923	
08_2387_T2_[A1]	0.4250245	85.0049	
08_2479_T1_[A1]	0.4269843	85.39686	
08_3033_T1_[A2]	0.4269413	85.38826	
08_3696	NA	NA	insufficient variants to calculate; also excluded on the basis of clustering
09_1196_[B]	0.4140461	82.80922	
11_2749_T1_[C]	NA	NA	insufficient variants to calculate
11_3909	NA	NA	insufficient variants to calculate; also excluded on the basis of clustering
11_3918	NA	NA	insufficient variants to calculate; also excluded on the basis of clustering
11_4115_T3_[A1]	0.4267504	85.35008	
12_0705_T2_[C]	0.2662101	53.24202	
12_0820_[C]	0.4302974	86.05948	
12_2065_T2_[A2]	0.4631092	92.62184	
12_3045_[A1]	0.4199537	83.99074	
12_3856_T2_[A2]	0.422119	84.4238	
12_4284_T2_[B]	0.4342737	86.85474	
12_4284_T3_[A2]	0.3702872	74.05744	
15_1368_T1_[B]	0.4932978	98.65956	
15_1524_[B]	0.4501014	90.02028	
15_1610_[B]	0.4241794	84.83588	

Sample	VAF(het)	Purity
15_2070_T1_[C]	0.413281	82.6562
982009000000000_[A2]	0.4574514	91.49028

Table A3-3. RDA and variance partitioning analysis results. The full model (Gene count ~ study + tissue type) had an adjusted R_2=0.61, indicating that it accounts for ~61% of the variation in gene counts. Tissue type was significant as a predictor variable (p=0.001), but study was not (p=0.186). Variance partitioning analysis showed tissue alone (while controlling for the effect of study) accounted for 31% of the variation in gene count (p=0.001) and study (while controlling for the effect of tissue) only accounted for 0.7%. The ANOVA on the variance partitioning analysis also showed that study was not statistically significant (p=0.2).

	Adjust R^2	df	Variance	F	p-value
RDA					
Gene count ~ study + tissue type	0.61	10	39971	9.2501	0.001*
Tissue	-	8	38750	11.2094	0.001*
Study	-	2	1221	1.4129	0.186
Residual	-	43	18581	-	-
Variance partitioning analysis					
[a] = Tissue | Study	0.31	8	20908	6.0481	0.001*
[b] = Study | Tissue	0.007	2	1221	1.4129	0.2
[c] = Tissue + Study	0.29	0	-	-	-
[d] = Residuals	0.39	-	-	-	-

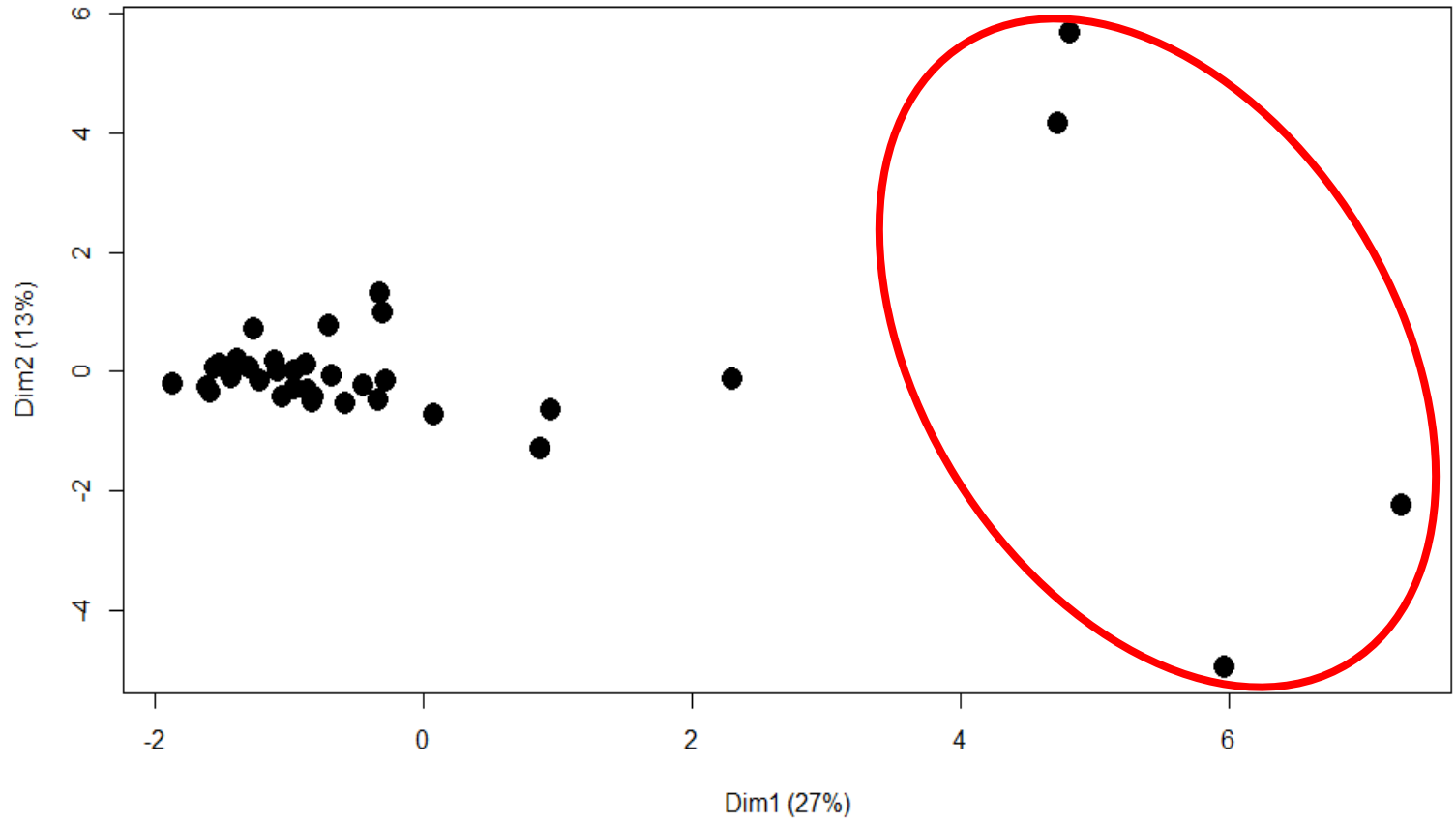


Figure A3-1. MDS plot of all 35 DFT1 samples. This plot revealed two pairs of outliers that were separated from the rest of the samples on both Dimension 1 and Dimension 2, as indicated by the red circle.

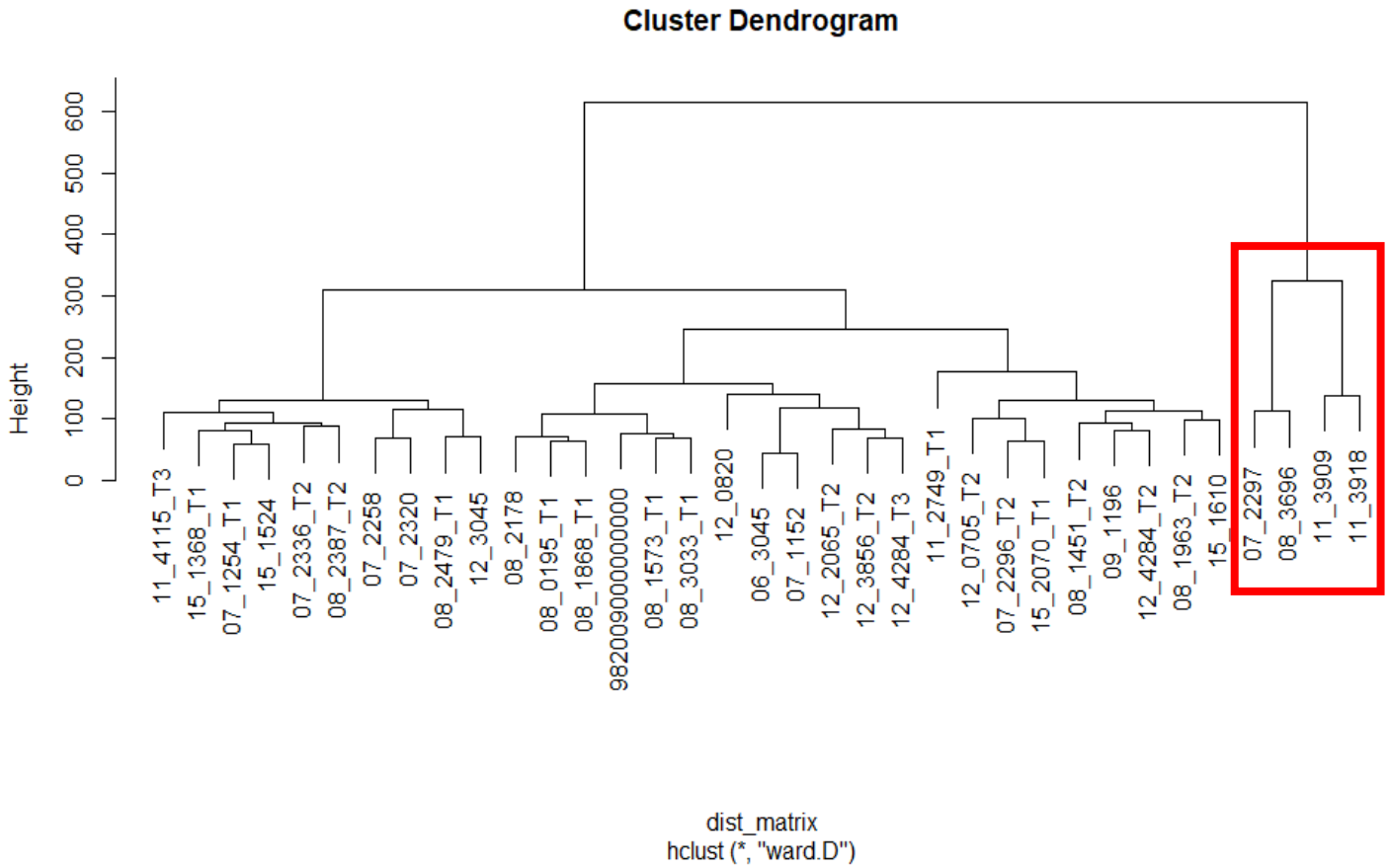


Figure A3-2. Hierarchical clustering of all 35 DFT1 samples using Ward's D and Euclidean distance. The same four outliers in figure 1a were assigned to their own cluster (red box), distinct from the other samples.

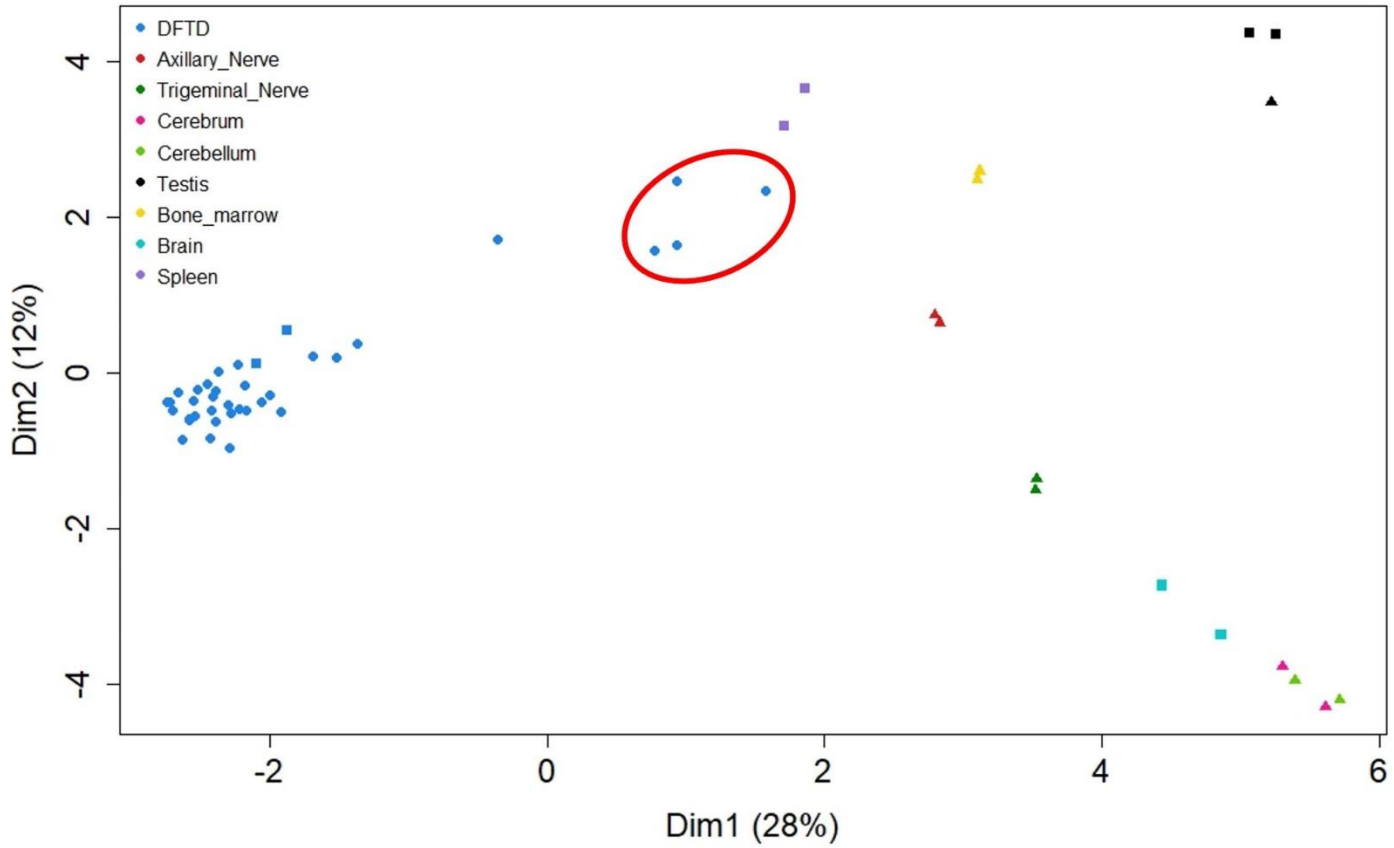


Figure A3-3. MDS plot of the entire dataset (n = 37 DFT1 biopsies, two axillary nerve (Stammnitz et al., 2023), two bone marrow (Stammnitz et al., 2023), two brain (Patchett et al., 2020), two cerebellum (Stammnitz et al., 2023), two cerebrum (Stammnitz et al., 2023), two spleen (Patchett et al., 2020), three testes (Patchett et al., 2020; Stammnitz et al., 2023) and two trigeminal nerve). Samples

from the current study are represented by circles, samples from Patchett (2020) are represented by squares and samples from Stammnitz (2023) are represented by triangles. The four samples identified as outliers in figure 1a and 1b clustered more closely to the healthy tissue biopsies than the DFT1 biopsies (circled).

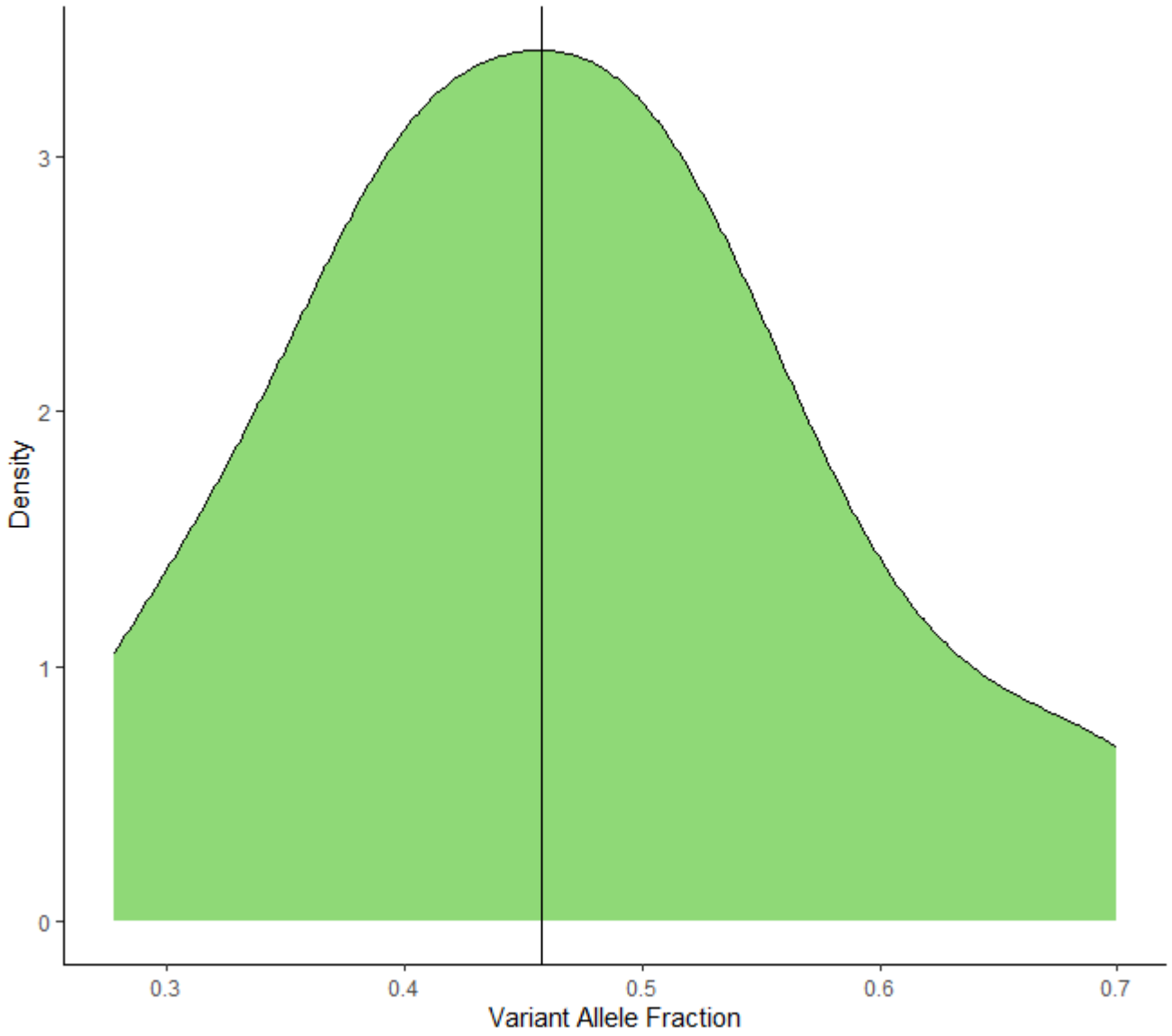


Figure A3-4. VAF distribution plot for sample 982009000000000. The vertical line is at the maximum density of the distribution and represents VAF_{HET} (the mode VAF of heterozygous variants). In this example, VAF_{HET}= 0.46 and so tumour purity (ρ) for this sample has been estimated to be 0.92 (i.e. 2*0.46).

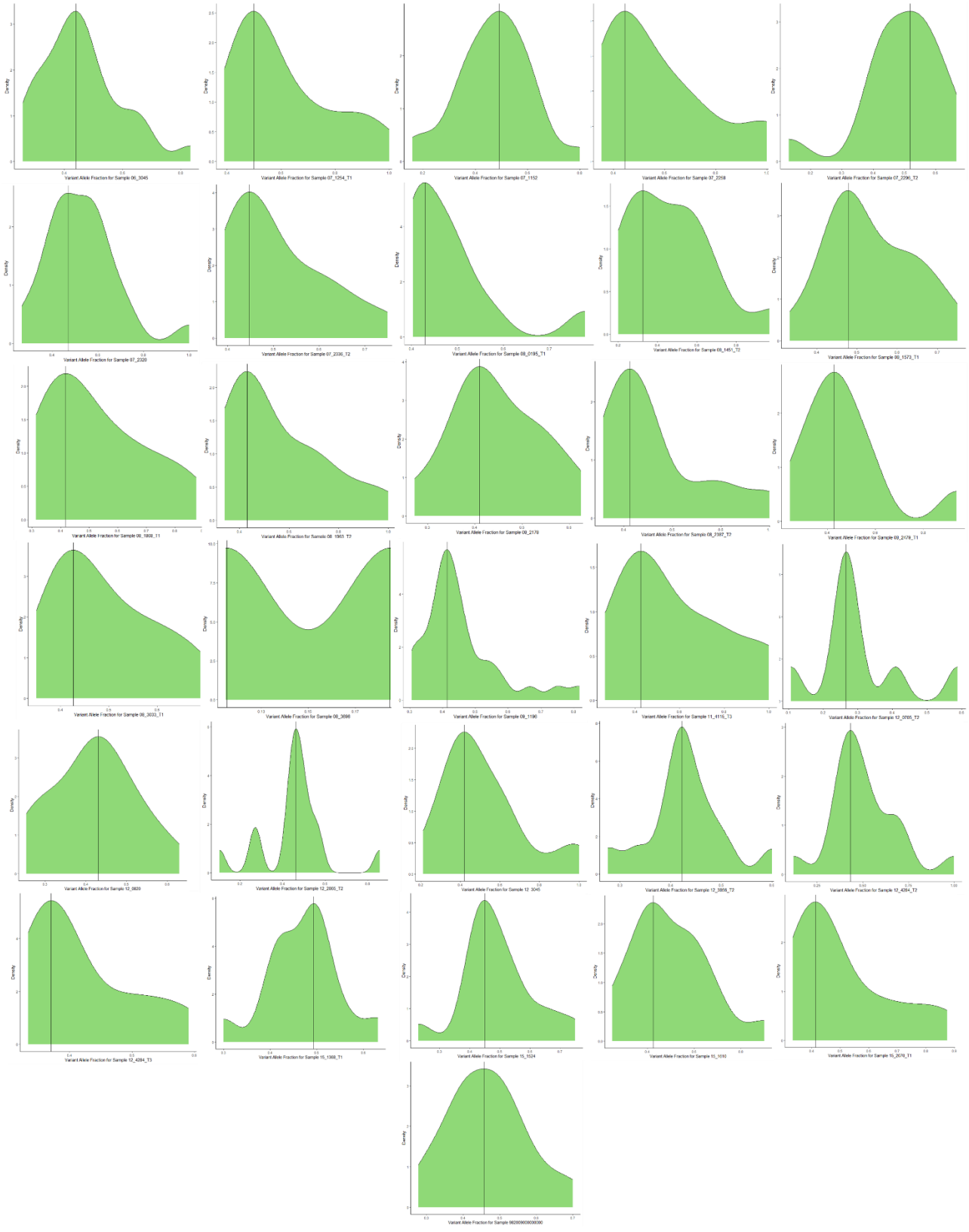


Figure A3-5. VAF distributions for all samples with 2 or more trunk variants.

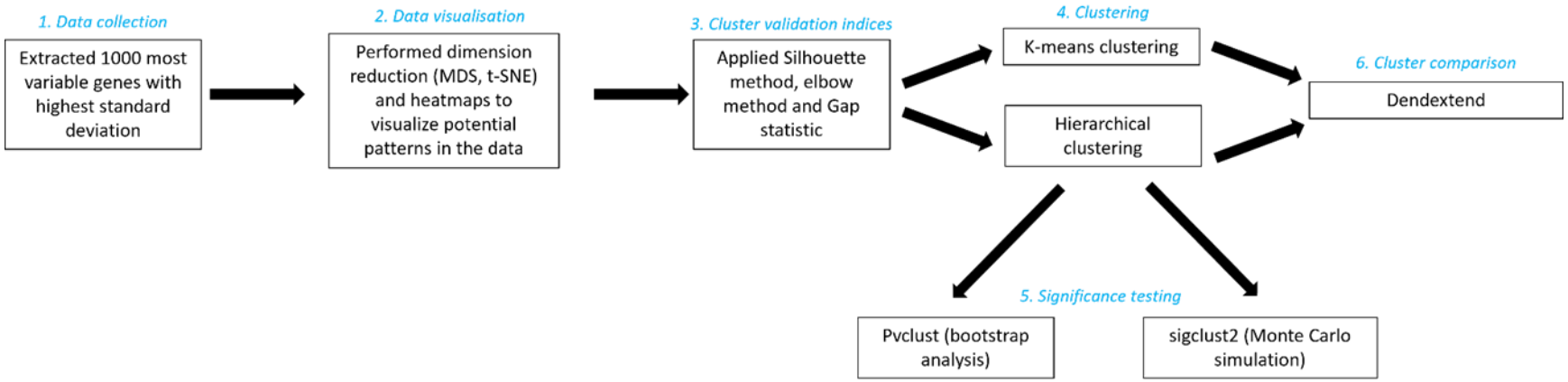


Figure A3-6. Experimental design.

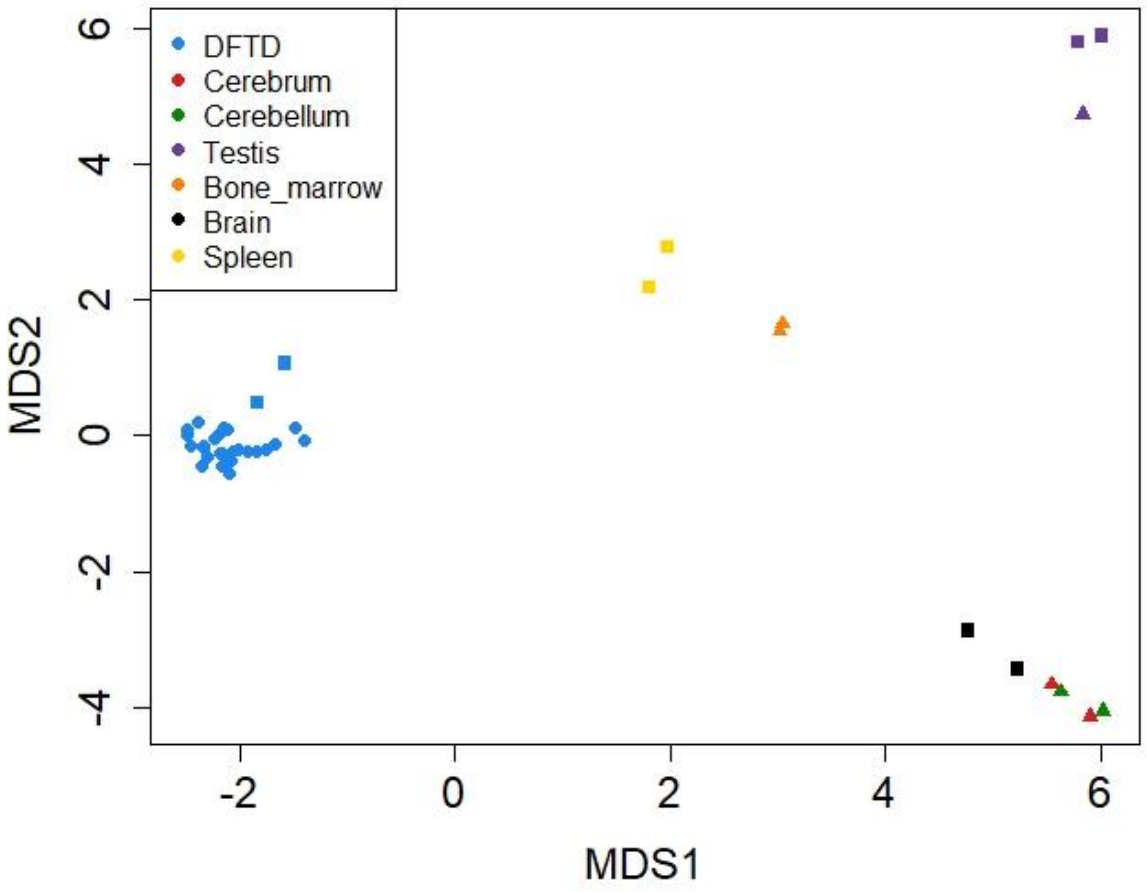


Figure A3-7. MDS plot of 29 DFT1 samples and 17 healthy tissue samples, excluding the samples with purity < 80%. Samples from the current study are represented by circles, samples from Patchett et al. (2020) are represented by squares and samples from Stammnitz et al. (2023) are represented by triangles.

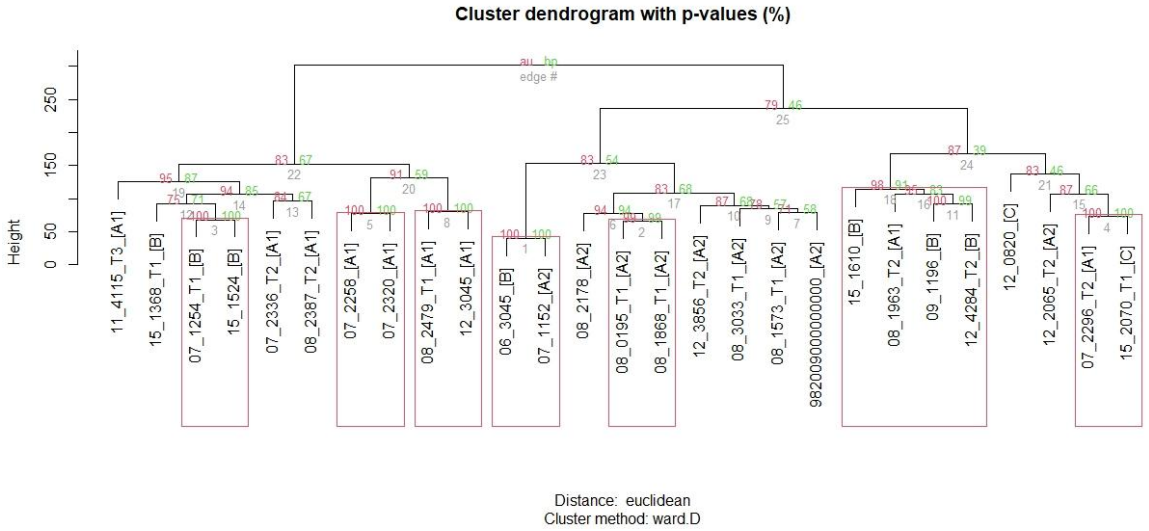


Figure A3-8. Results from significance testing of hierarchical clustering results using pvclust. Values in red are AU (unbiased probability values obtained by multiscale bootstrap resampling), values in green are BP (obtained by ordinary bootstrap resampling). Clusters with BP > 95 are contained in a red rectangle. AU values are low across most of the dendrogram, with a large portion unable to be assigned to any clade.

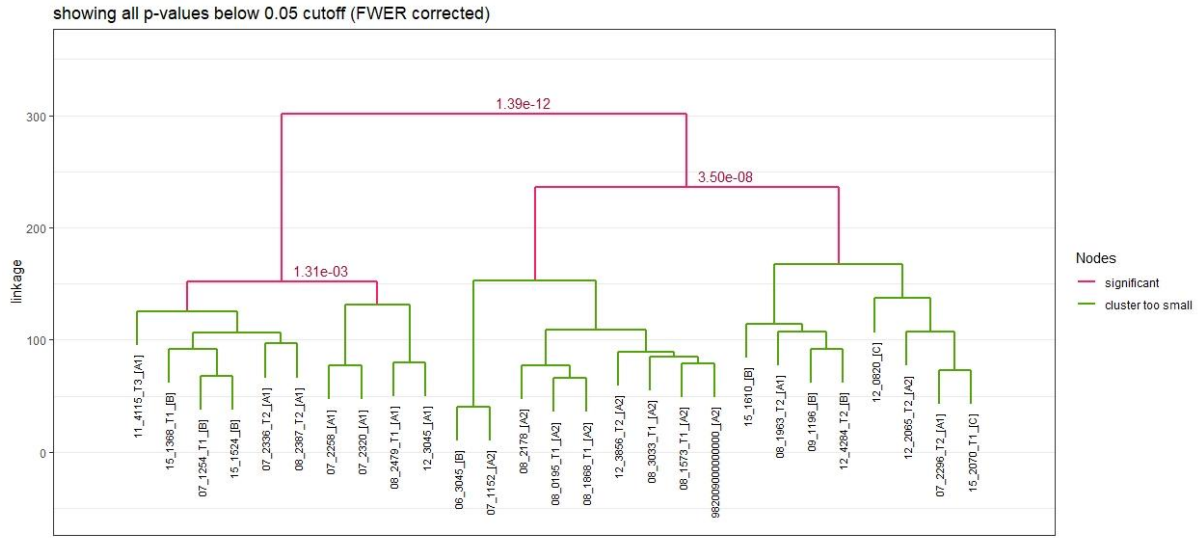


Figure A3-9. Statistical significance of hierarchical testing results using Monte-Carlo based method with sigclust2. Four nodes are statistically significant and the clusters within these are too small for significance testing.

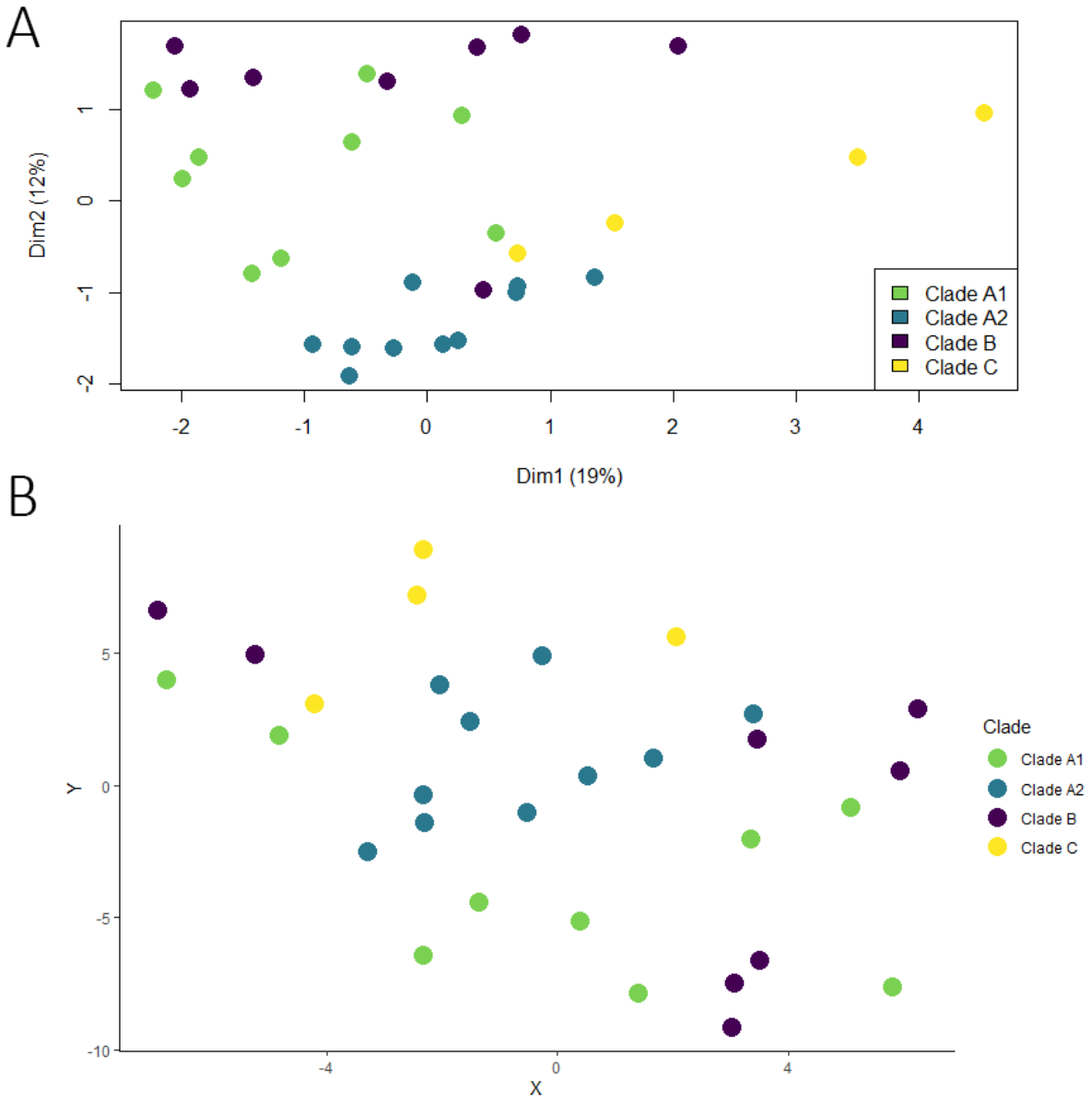


Figure A3-10. (A) MDS plot and (B) t-SNE plot of distances between TMM normalised gene counts for each DFT1 sample when not filtered for purity. Neither method of dimension reduction suggested the presence of natural groups in the data, with all samples largely clustering together. Samples are coloured by genotypic clade.

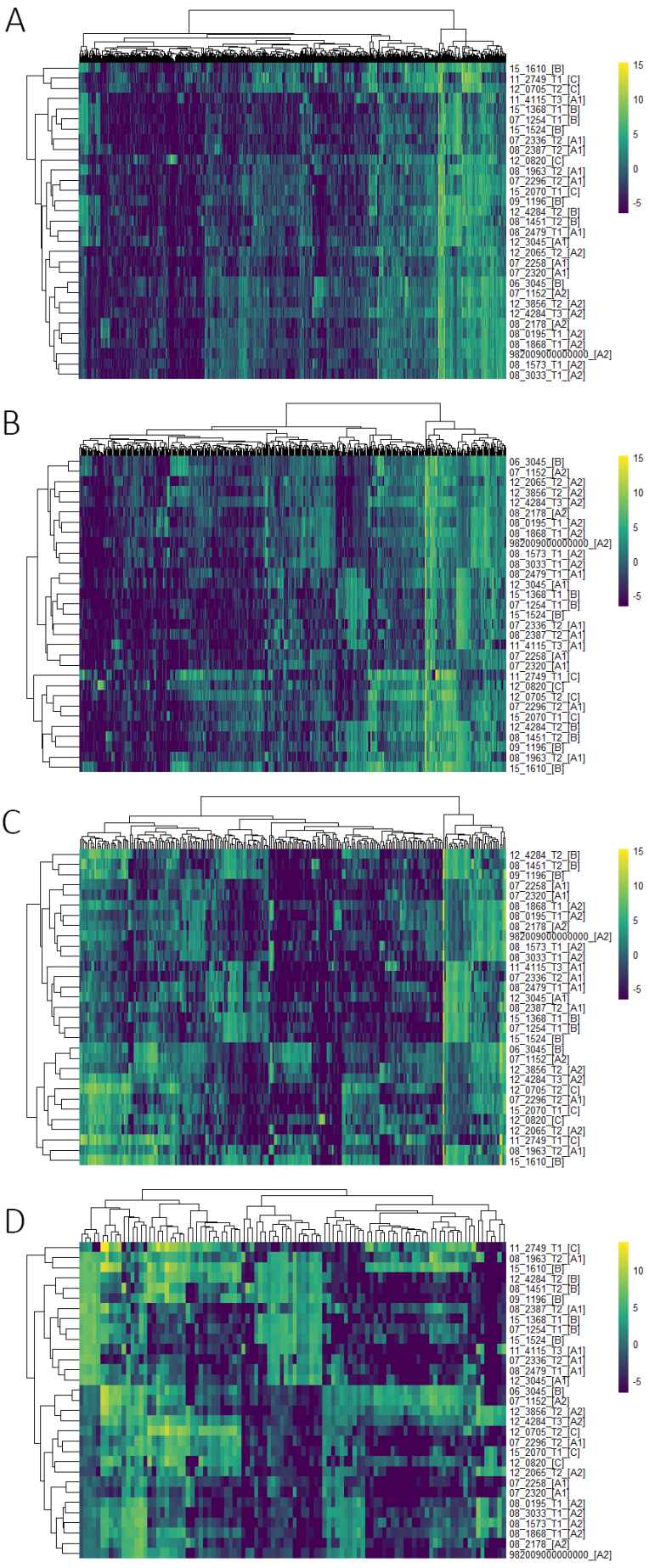

A
15_1610_[B]
11_2749_T1_[C]
12_0705_T2_[C]
11_4115_T3_[A1]
15_1368_T1_[B]
07_1254_T1_[B]
15_1524_[B]
07_2336_T2_[A1]
08_2387_T2_[A1]
12_0820_[C]
08_1963_T2_[A1]
07_2296_T2_[A1]
15_2070_T1_[C]
09_1196_[B]
12_4284_T2_[B]
08_1451_T2_[B]
08_2479_T1_[A1]
12_3045_[A1]
12_2065_T2_[A2]
07_2258_[A1]
07_2320_[A1]
06_3045_[B]
07_1152_[A2]
12_3856_T2_[A2]
12_4284_T3_[A2]
08_2178_[A2]
08_0195_T1_[A2]
08_1868_T1_[A2]
982009000000000_[A2]
08_1573_T1_[A2]
08_3033_T1_[A2]
15
10
5
0
-5
B
06_3045_[B]
07_1152_[A2]
12_2065_T2_[A2]
12_3856_T2_[A2]
12_4284_T3_[A2]
08_2178_[A2]
08_0195_T1_[A2]
08_1868_T1_[A2]
982009000000000_[A2]
08_1573_T1_[A2]
08_3033_T1_[A2]
08_2479_T1_[A1]
12_3045_[A1]
15_1368_T1_[B]
07_1254_T1_[B]
15_1524_[B]
07_2336_T2_[A1]
08_2387_T2_[A1]
11_4115_T3_[A1]
07_2258_[A1]
07_2320_[A1]
11_2749_T1_[C]
12_0820_[C]
12_0705_T2_[C]
07_2296_T2_[A1]
15_2070_T1_[C]
12_4284_T2_[B]
08_1451_T2_[B]
09_1196_[B]
08_1963_T2_[A1]
15_1610_[B]
15
10
5
0
-5
C
12_4284_T2_[B]
08_1451_T2_[B]
09_1196_[B]
07_2258_[A1]
07_2320_[A1]
08_1868_T1_[A2]
08_0195_T1_[A2]
08_2178_[A2]
982009000000000_[A2]
08_1573_T1_[A2]
08_3033_T1_[A2]
11_4115_T3_[A1]
07_2336_T2_[A1]
08_2479_T1_[A1]
12_3045_[A1]
08_2387_T2_[A1]
15_1368_T1_[B]
07_1254_T1_[B]
15_1524_[B]
06_3045_[B]
07_1152_[A2]
12_3856_T2_[A2]
12_4284_T3_[A2]
12_0705_T2_[C]
07_2296_T2_[A1]
15_2070_T1_[C]
12_0820_[C]
12_2065_T2_[A2]
11_2749_T1_[C]
08_1963_T2_[A1]
15_1610_[B]
15
10
5
0
-5
D
11_2749_T1_[C]
08_1963_T2_[A1]
15_1610_[B]
12_4284_T2_[B]
08_1451_T2_[B]
09_1196_[B]
08_2387_T2_[A1]
15_1368_T1_[B]
07_1254_T1_[B]
15_1524_[B]
11_4115_T3_[A1]
07_2336_T2_[A1]
08_2479_T1_[A1]
12_3045_[A1]
06_3045_[B]
07_1152_[A2]
12_3856_T2_[A2]
12_4284_T3_[A2]
12_0705_T2_[C]
07_2296_T2_[A1]
15_2070_T1_[C]
12_0820_[C]
12_2065_T2_[A2]
07_2258_[A1]
07_2320_[A1]
08_0195_T1_[A2]
08_3033_T1_[A2]
08_1573_T1_[A2]
08_1868_T1_[A2]
08_2178_[A2]
982009000000000_[A2]
10
5
0
-5

Figure A3-11. Heatmaps displaying (A) 1000 most variable genes (B) 500 most variable genes (C) 250 most variable genes (D) 100 most variable genes of all samples (not filtered for purity). None showed a clear mosaic pattern that would be expected if distinct clusters were present in the dataset. Yellow represents a higher value (indicating genes are upregulated in that sample) and dark blue represents a lower value (indicating that genes are downregulated in that sample). The name of the clade is in square brackets after the sample name.

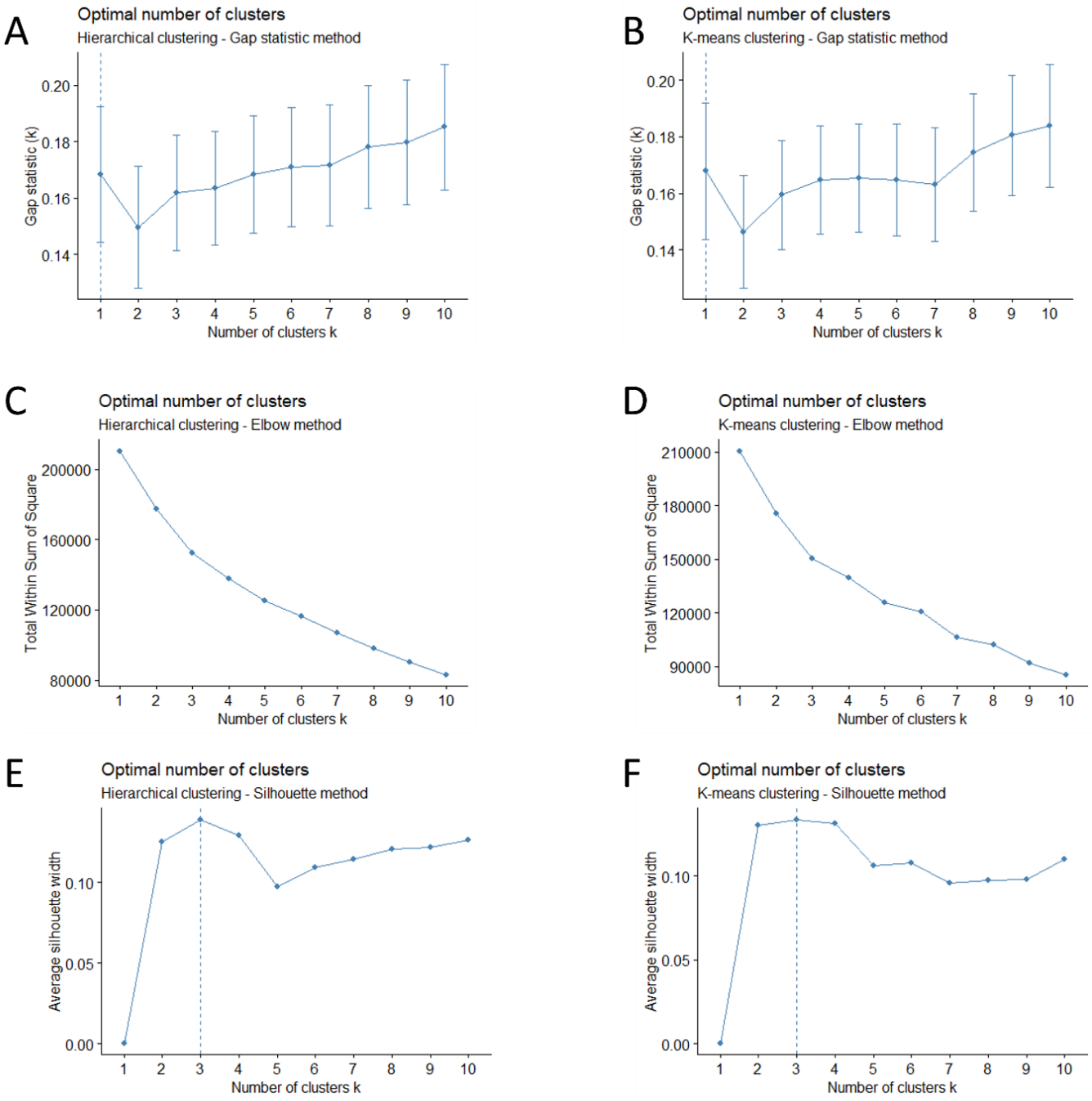


Figure A3-12. Clustering validation indices on the full dataset (when not filtered for purity). Gap statistic method indicated the optimum number of clusters was one for both (A) hierarchical clustering and (B) k-means clustering; the Elbow method was ambiguous for both (C) hierarchical clustering and (D) k-means clustering; and the Silhouette method indicated the optimum number of clusters was three for both (E) hierarchical clustering and (F) k-means clustering. The dotted line represents the highest value i.e. what the test deems to be the optimum number of clusters in the dataset.

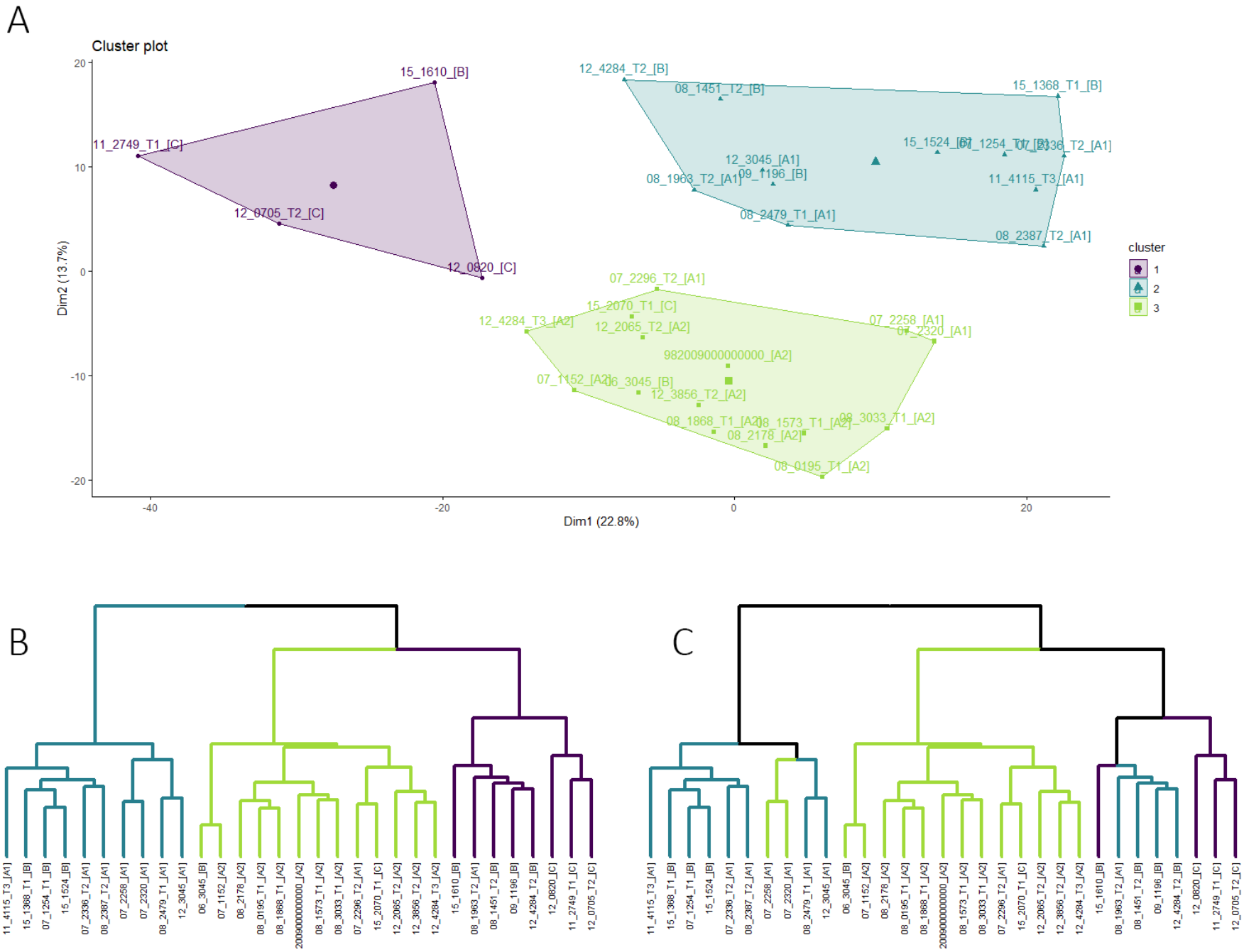

A
Cluster plot
Dim2 (13.7%)
Dim1 (22.8%)
cluster
1
2
3
15_1610_[B]
11_2749_T1_[C]
12_0705_T2_[C]
12_0820_[C]
12_4284_T2_[B]
08_1451_T2_[B]
15_1368_T1_[B]
12_3045_[A1]
09_1196_[B]
08_1963_T2_[A1]
11_4115_T3_[A1]
08_2479_T1_[A1]
08_2387_T2_[A1]
07_2296_T2_[A1]
15_2070_T1_[C]
12_4284_T3_[A2]
07_2258_[A1]
12_2065_T2_[A2]
07_2320_[A1]
982009000000000_[A2]
07_1152_[A2]
06_3045_[B]
12_3856_T2_[A2]
08_1868_T1_[A2]
08_1573_T1_[A2]
08_3033_T1_[A2]
08_2178_[A2]
08_0195_T1_[A2]
B
C

Figure A3-13. Clustering results on the full dataset when not filtered for purity: (A) K-means clustering results with k=3. (B) Dendrogram depicting the results of hierarchical clustering. The samples have been coloured to according to which group the hierarchical clustering assigned them. (C) Dendrogram depicting the results of hierarchical clustering. For comparison, the samples have been coloured to according to which group the k-means clustering assigned them. The clade is in square brackets after the sample name. This indicates that clusters defined by the two different methods overlapped but did not reach a consistent consensus.