Genomic differentiation of brushtail possum (*Trichosurus vulpecula*) populations

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Zoology

at Massey University, Manawatū, New Zealand

David Carmelet-Rescan

2023

Abstract

Toxin resistance, an ecologically functional trait, has emerged as a result of the coevolutionary arms race between plants and herbivores. This adaptive response allows organisms to counteract the detrimental effects of toxins. Such adaptation occurs within three primary ecological contexts: predator resistance, prey resistance, and auto-resistance. In the context of plant-herbivore interactions, the production of toxic secondary metabolites by plants plays a pivotal role, triggering a dynamic arms race with herbivores. This ongoing battle leads to the development of diverse defensive strategies in plants and corresponding counter-adaptations in herbivores. Toxin resistance studies have not only shed light on coevolutionary dynamics but also provided insights into regional adaptations and population fragmentation. The case of sodium fluoroacetate, a potent toxin found in certain plant species, offers a specific example of toxin resistance in brushtail possums (*Trichosurus vulpecula*). Deepening our understanding of the underlying mechanisms driving toxin resistance becomes particularly interesting in this case because of the notable variation in sodium fluoroacetate resistance observed among subspecies. The brushtail possum (*Trichosurus vulpecula*) is protected in its natural range of Australia but as an invasive pest in New Zealand populations are controlled by the application of human made sodium fluoroacetate, providing a strong selective force for the potential parallel evolution of toxin resistance.

I first focus on the population fragmentation and past connectivity of brushtail possum subspecies in their native range using ecological niche modelling and genetic analysis. My results suggest that Pliocene and Pleistocene climate oscillations played a significant role in shaping the distribution and structure of these widespread marsupials. The findings highlight the limited gene flow between subspecies that would have facilitated their adaptations to regional plant assemblages.

I studied gene expression using RNA sequencing to reveal significant differences in transcription levels between adult and juvenile brushtail possums. These findings provide insights into the mechanisms associated with possum development, tissue development, cell cycle, and extracellular matrix. Furthermore, the downregulation of specific genes, such as Cytochrome P450, in juveniles may indicate their role in possums' dietary adaptations. Differential gene expression among subspecies of brushtail possum shed light on genomic

differentiation and identified candidate genes involved in regional adaptations involving toxin tolerance. Further analyses comparing subspecies has identified genes under positive selection and enriched biological pathways that may be associated with sodium fluoroacetate resistance. These findings support the idea of genomic divergence in spatial populations and provide hypotheses on the metabolic pathways involved in toxin resistance.

Collectively, these studies offer valuable insights into the coevolutionary dynamics, toxin resistance mechanisms, gene expression and selection patterns in brushtail possums (*Trichosurus vulpecula*) and formulated strong hypotheses related to sodium fluoroacetate resistance. This study presents a reliable method to use transcriptomics to investigate the expression and genomic differences behind observed phenotypic variation within a single species.

Acknowledgment

I would like to begin by expressing my gratitude to my supervisors for their invaluable support throughout this project. A heartfelt thanks to Steve for welcoming me into the Phoenix Lab team during my internship in 2018 and for granting me the opportunity to pursue my research with him for my PhD. Your constant encouragement to explore new directions and your shared passion for evolutionary ecology has been truly inspiring. Mary, I am incredibly grateful for your warm and caring support over the past four years, as well as your guidance through thought-provoking questions. I consider myself fortunate to have crossed paths with both of you and to have had you as my supervisors.

I would also like to extend a special thanks to Nim, my wonderful colleague and dear friend, the second member of team possum. Your friendship has made my stay in New Zealand all the more memorable and fulfilling.

Kwan, you hold a special place in my heart as the primary reason for my desire to return to New Zealand. I am deeply thankful for your love and support, whether it came from New Zealand, Thailand, or France.

To Mathieu and Julien, thank you for your friendship and for warmly welcoming me into your circle and showing me around.

Being a part of the Phoenix Lab group has been an absolute joy, and I am grateful for the opportunity to work with each member. Emily, Amanda, Mari, Michelle, Leo, Gillian, Simon, Yuta, Shogo, Suman, and Charlotte, I appreciate our collaborations, the assistance provided, the friendships formed, and the stimulating conversations we have shared.

I would also like to express my gratitude to all the friends I have made along the way, making all those years in New Zealand go so fast.

I would like to extend my heartfelt appreciation to Bethany Jackson from Murdoch University, Meg Rodgers from WA Wildlife, and Dale Chittenden from the Department of Conservation on Stewart Island providing essential assistance during the sampling process.

Lastly, I want to thank my family and friends in France who have supported me unwaveringly, despite the physical distance and the limited updates I have provided. Your constant belief in me has been a tremendous source of strength.

Table of Contents

Chapter 1: General Introduction

In this thesis, which consists of a General Introduction, four research chapters and General conclusions, I investigate genomic elements in the brushtail possum *Trichosurus vulpecula.* Using existing knowledge of the biology and ecology of this species in its native Australian range and in New Zealand I take a genomic approach to compare lineages with distinct characteristics. The rationale for this work is the fact that brushtail possums in west and east Australia display very different responses to toxins in the plants they eat, and one of the most potent plant toxins is widely applied in New Zealand to manage this major agricultural and biodiversity pest. Using comparative analyses, I seek an understanding of the molecular mechanisms associated with toxin metabolism. In this introduction (chapter 1), I summarize current knowledge of animal toxin resistance and its relevance in brushtail possums. I also describe the concepts and methods used in this study, provide an overview of the objectives of my study and summarise the content covered in each chapter to give a synopsis of the thesis.

Coevolution toward toxin resistance

Coevolution, the reciprocal influence between two or more separate evolutionary lineages, has played a fundamental role in shaping the diversity of life (Dawkins and Krebs 1979; Ehrlich and Raven 1964; Endara et al. 2017). Toxin resistance and the coevolutionary dynamics it encompasses have been the subject of extensive research in the field of evolutionary biology, with fundamental literature presented by Erlich and Raven (1964). Their seminal paper highlighted the coevolutionary ‘arms race’ between the caterpillars of certain butterflies and the plants they eat, emphasizing defensive compounds deployed by plants and the insect adaptations to counter these toxins. Their work highlighted how intense biotic interactions could drive diversification in phytophagous organisms, including butterflies, and plants (Edger et al. 2015). Although coevolution is often regarded as a major driver of species diversification, Suchan and Alvarez (2015) questioned its significance in explaining the current diversity of plants and associated insects. Through a review of classical and recent case studies, they suggested that coevolutionary processes might occur through relatively short "interludes," making the pattern difficult to detect.

The production of plant secondary metabolites that are toxic to herbivores and so protect plant foliage are implicated in the reciprocal evolution of toxin resistance, which involves the development of mechanisms by organisms to counteract the harmful effects of toxins produced

by their counterparts. Toxin resistance has been extensively studied in numerous ecological contexts, and understanding the mechanisms of toxin resistance is essential for elucidating the complex arms race between species and the coevolutionary processes underlying it (e.g. de Castro et al. 2018). Toxin resistance can also be considered an ecologically functional trait that mediates antagonistic interactions between predators and their prey, driven by natural selection. Van Thiel et al. (2022) conducted a comprehensive review of toxin resistance in animals, highlighting its convergent evolution and the diverse ecological contexts in which it has evolved. Three primary ecological contexts in which toxin resistance has been observed were identified: predator resistance, prey resistance, and autoresistance. Predator resistance refers to the situation where a predator has evolved resistance to the toxins produced by its prey (eg. grasshopper mice (*Onychomys spp.*) to Arizona bark scorpion (*Centruroides exilicauda*), Rowe and Rowe 2008; or Hedgehog (*Erinaceus europaeus*) to European viper (*Vipera berus*), de Wit and Weström 1987). Prey resistance involves prey species developing resistance to the toxins of their predators (eg. hispid cotton rat *(Sigmodon hispidus)*, gray woodrat *(Neotoma micropus)*, and the opossum (*Didelphis virginiana)* to the Western diamondback rattlesnake (*Crotalus atrox)*, Perez, Pichyangkul, and Garcia 1979). Autoresistance, on the other hand, occurs when an organism becomes resistant to its own toxins (Wang and Wang 2017; Yuan and Wang 2018). These distinct ecological contexts demonstrate the versatility and adaptive nature of toxin resistance mechanisms.

In natural environments, plants are engaged in reciprocal selective interactions with animal herbivores (Ehrlich and Raven 1964). This coevolutionary arms race (Dawkins and Krebs 1979; Endara et al. 2017) involves many types of weapons with plant primary and secondary metabolites playing a major role in the armoury. Plant communities richly endowed with these chemical defences impose strong natural selection on potential browsing animals. Resistance in herbivores that eat plants containing toxic compounds can be both behavioural and physiological traits that limit toxicosis (Ujvari et al. 2015). This coevolution operates in a spatial mosaic of community composition, population density, trait expression and other factors such that populations evolve different traits and so contribute to functional diversity and ecological complexity (Chaves-Campos, Johnson, and Hulsey 2011; Scheuerl et al. 2019). The genomic processes underlying these interactions also have implications for the evolution of toxin resistance by pest species around the world (Hawkins et al. 2019).

The adaptation of generalist herbivores to diverse plant secondary metabolites have been proposed as making them genetically predisposed (pre-adaptation) to rapidly evolve resistance to synthetic pesticides (Dermauw et al. 2018). The concept of host plant adaptation provides a theoretical pathway to pesticide resistance in herbivores. Dermauw et al. (2018) discussed the uses of *resistance* to insecticides and *tolerance* of plant toxins, suggesting it might be important in the context of resistance management. Both terms refer to genotypes that can biochemically and physiologically reduce toxicity but Dermauw et al. (2018) used *resistance* in the context of human-modified selection while *tolerance* was used for phenotypes produced by natural selection, highlighting the value of a clear understanding of the target under selection and the selective agent for effective resistance management strategies. However, the distinction is not clear cut and other authors use the term *resistance* in all contexts (Van Thiel et al. 2022). In the case of brushtail possums, the plant toxin that some populations are *tolerant* of (selective agent), is the same toxin applied by humans to kill this pest in New Zealand where *resistance* might evolve, so within this thesis, I will continue to use "resistance" to describe the phenotype of low susceptibility.

More important is the cellular and biochemical pathways that create the phenotype of surviving poisonous compounds. Toxin resistance has evolved numerous times across Metazoa (Arbuckle, Rodríguez de la Vega, and Casewell 2017). Gene-product-based resistance has been categorized into three main types: toxin scavenging, target-site insensitivity, and off-target repurposing. These mechanisms allow animals to either bind and inactivate toxins, render target molecules refractory to toxins, or exploit toxin functions to alter their physiological effects, respectively. The exploration of these resistance mechanisms has been studied within many plant/insect interactions providing a deeper understanding of the coevolutionary dynamics between organisms producing toxins and those evolving resistance (Karageorgi et al. 2019). Invertebrates, as key players in various ecosystems, have provided valuable insights into the mechanisms of toxin resistance. For example invertebrates resistant to toxins produced by the bacterium *Bacillus thuringiensis* ,mostly throught off-target repurposing, shed light on the mode of action of the Cry protein produced by *B. thuringiensis* (Griffitts and Aroian 2005). Understanding the cellular mechanisms is crucial for managing and mitigating the development of resistance to human-applied toxins (Peterson, Bezuidenhout, and Van den Berg 2017).

Resistance (or tolerance) can also happen through a more general response that could be referred to as damage control instead of a toxin-specific response. One example is in the

butterfly *Eumaeus atala* which developed tolerance to cycad toxins putatively through multiple mechanisms such as autophagy of damaged cells, removal of cell debris by macrophages, and more active cell proliferation (Robbins et al. 2021).

Mammalian resistance to fluoroacetate toxin

Sodium fluoroacetate, also known as the toxin "1080", occurs naturally in tropical and subtropical plant species in New Zealand (*Puha),* Brazil and Africa (Leong et al. 2017) but is especially high in 40 plants species in Australia (i.e. *Gastrolobium* spp, (Eason 2002; Ogilvie et al. 2009)). Sodium fluoroacetate is known to be a potent toxin in mammals (Gupta 2015) including dogs, rats and some populations of brushtail possums *Trichosurus vulpecula* (Alterio 2000; Eason et al. 2011). Its toxicity comes from its preferential binding of fluoroacetate to aconitase in the tricarboxylic cycle in vertebrates, leading to termination of cellular respiration and death (Leong et al. 2017, Figure 2). This toxin mediated plant-herbivore interaction prevents the plant from being eaten by most mammal species (Twigg et al. 1996).

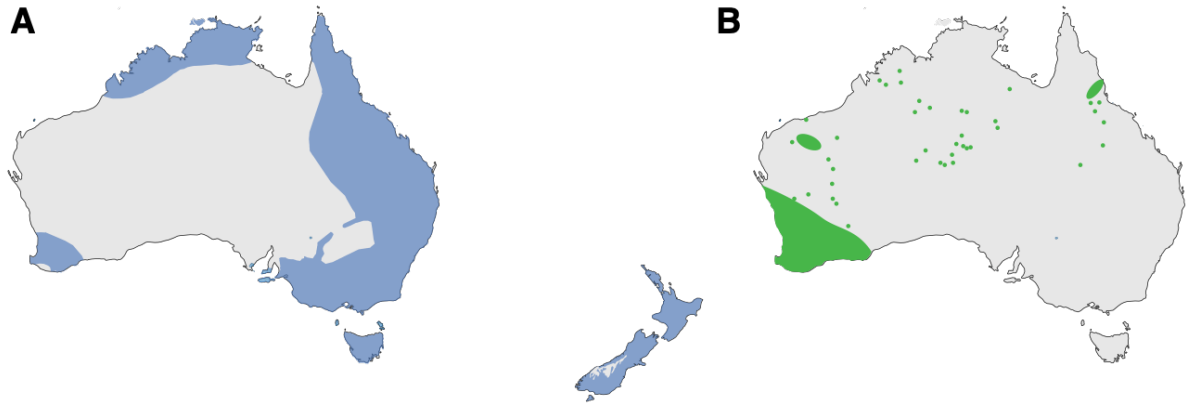


Figure 1: A) Brushtail possum *Trichosurus vulpecula* distribution. The distribution over its native country of Australia is spread along the coast, and introduced to New Zealand where it is now widespread expect on some high elevation alpine areas. B) Distribution of food plant *Gastrolobium* spp. in Australia.

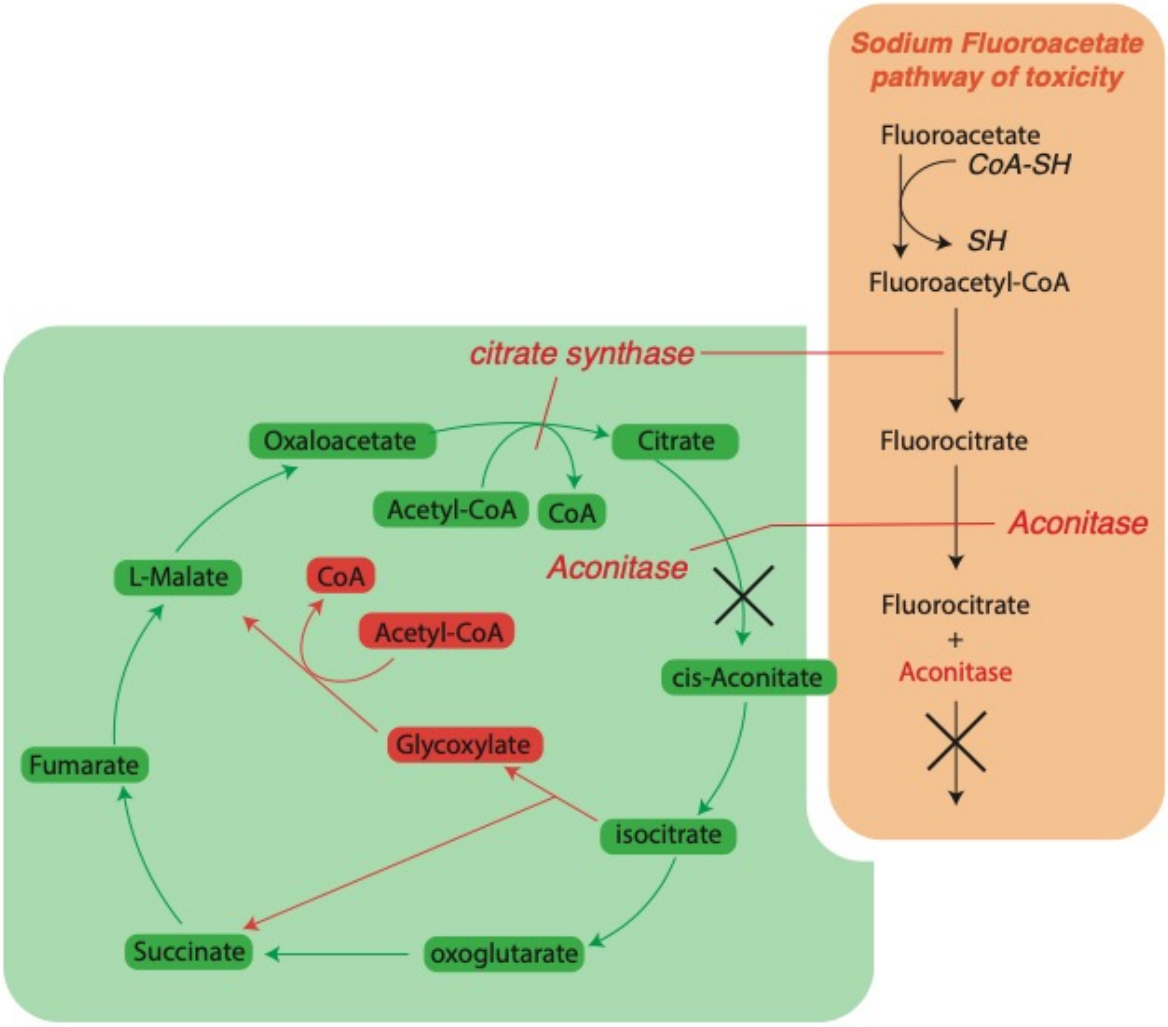


Figure 2: Pathway of toxicity of sodium fluoroacetate and the interaction with the tricarboxylic cycle.

In Western Australia plant-eating marsupials coexist with ~40 plant species including *Gastrolobium* spp. that produce high levels of 1080 to defend their foliage and these herbivores can tolerate 1080 (Twigg et al. 1996; Twigg and King 1991, Figure 1). In contrast, plant-eating marsupials in eastern Australia experience little of the toxin in nature and are sensitive to 1080. For example, brushtail possums in Western Australia have an LD_{50} (median lethal dose) for 1080 that is about 160 times higher than those in the east (Eason 2002; King, Oliver, and Mead 1978; Montague 2000; Twigg and King 1991). Thus, divergent trait expression has evolved in populations of the same species in a geographic mosaic, providing the basis for comparative analysis of the genomic pathways to resistance. However, a comparison of western and eastern possums found neither phenotypic difference in the defluorination rate of the liver (King, Oliver, and Mead 1978), nor functional genotypic differences in the aconitase gene (Deakin et

al. 2013). Decoding the genetic basis of fluoroacetate resistance requires a genome-wide analysis with associated population screening of loci that are implicated.

First developed in the United States in 1940s as a rodenticide prior to the Second World War, sodium fluoroacetate has since been used all over the world in pest management (e.g. introduced red foxes in Western Australia, and coyotes in USA (Robinson 1953; Thompson and Fleming 1994)). In New Zealand it has been used since the 1950s to kill mammalian pest species (Eason et al. 2011; Triggs and Green 1989). Its use in New Zealand is justified and particularly effective because there are only two native mammal species that might be affected (the bats *Mystacina tuberculata* and *Chalinolobus tuberculatus*) in terrestrial habitats, whereas other countries have a large number of non-target mammal species (Eason et al. 2011).

The brushtail possum problem in New Zealand

Brushtail possums (*Trichosurus vulpecula*) are endemic to Australia where they are found in the east and north and Tasmania, along with the south-western region (Figure 2). Despite being previously widespread in arid regions they are facing a decline in numbers over the last century with European settlement and predators that came with them (Kerle and How 2008). Since their introduction to New Zealand in the mid-1800s they have expanded to cover the country. Possum density is highly dependent on their foraging opportunities (Efford, Warburton, and Spencer 2000). In New Zealand, they are offered an extensive and attractive habitat. The climatic conditions in New Zealand allow more dense vegetation supporting the browsing of many more individuals compared to the Australian vegetation (Pracy 1974). Their mostly folivorous diet can lead to complete defoliation of some canopy species (Nugent et al. 2000), such as the endemic Myrtaceae northern rata (*Metrosideros robusta*) and southern rata (*Metrosideros umbellata*) which are in turn important habitat for many native birds and small animals (Nugent et al. 2000). As omnivores brushtail possums are also a direct threat to native birds as they feed on them, their eggs, as well as on native invertebrates (Nugent et al. 2000). The brushtail possum population in New Zealand has grown to around 48 million creating a huge threat to New Zealand biodiversity (M. Clout and Ericksen 2000). During their introduction from Australia they lost some parasites (M. Clout and Ericksen 2000; P. E. Cowan 1990), but in New Zealand brushtail possums are associated with the spread of bovine tuberculosis (Tb), as they are a reservoir and vector for *Mycobacterium bovis* that causes Tb in cattle and deer (Livingstone et al. 2015).

New Zealand, a real-time experiment, and a need for knowledge of potential toxin resistance

The development of pesticide and antibiotic resistance was first observed a long time ago and is now considered to be a potential threat to human and livestock health, and agriculture productivity and is a major challenge for the scientific community (Bass et al. 2015; Denholm and Rowland 1992; Georghiou and Mellon 1983; Orzech and Nichter 2008; Powles and Yu 2010; Roush and Tabashnik 2012; Triggs and Green 1989). For example, more than 700 pest species have been reported as resistant to pesticides (2001, Food and Agriculture Organization). Understanding the evolutionary processes behind the emergence and the spread of toxin resistance in populations is fundamental in order to manage pest species in a manner that will avoid the development of chemical resistance (Hawkins et al. 2019; Maclean 2010; Neve et al. 2014). The evolutionary process underlying the development of toxin resistance requires natural genetic variation within existing populations plus a particularly strong selective pressure from the use of the pesticide. Exposure to chemicals used for killing pest species is probably always the strongest directional selective pressure that the target species is subject to at a short temporal scale (Boyle 1960; Elena and Lenski 2003; Hawkins et al. 2019; Lenski 2017). These immense selective pressures create a selective sweep and strong bottleneck as "natural" selection rapidly changes allele frequencies in favour of genotypes that are resistant. Resistant individuals have high fitness resulting in the resistant phenotype becoming common in the population in just a few generations (Besnier et al. 2014; Palumbi 2001). Selective pressure for resistance often overcomes strong metabolic trade-offs as resistance is mostly associated with a costly metabolic process that need to be outweigh by the advantage associated with the resistance (Macnair 1991; Wood 1981). The emergence of resistance in populations prevents local extinction (Alexander et al. 2014). The development of toxin resistance can arise from several pathways, one being the so-called "biochemical pre-adaptation" implying that some species could have been already exposed to similar toxins in nature and already have genetic and metabolite detoxification abilities allowing rapid development of resistance to a number of toxins used as pesticide (Dermauw et al. 2013; Rosenheim et al. 1996).

Conveniently from an experimental point of view, brushtail possums released in New Zealand set in place an extraordinary real-time experiment in evolutionary toxicology. Multiple

introductions from Victoria and Tasmania (populations not resistant to sodium fluoroacetate) yielded a large population of possums peaking at around 48 million (M. N. Clout and Ericksen 2000; Pracy 1974) with high genomic diversity being evident in fur colour polymorphism and neutral genetic markers (Taylor et al. 2004; Pattabraman et al. 2021). The application in New Zealand of synthetic 1080 (McIlroy 1983; Ross, Bicknell, and Hickling 1999) to a million hectares each year results in repeated high local kill followed by population recovery (P. Cowan 2016; P. E. Cowan 1992; Gupta 2015; Ross, Bicknell, and Hickling 1999); intense episodic selection for resistance in space and time. This anthropogenic evolutionary manipulation parallels the coevolutionary arms races of chemical defence and resistance in nature (O'Reilly-Wapstra and Cowan 2010) leading to the prediction that toxin resistance will evolve in New Zealand populations, just as it does in an experimental selection of insects (T. M. Brown and Payne 1988). The development of "1080" resistance in possums in New Zealand could lead to a huge loss in biodiversity and have great impact on the economy, which is why exploring this prediction is crucial for the future of pest management in New Zealand aiming at the optimal way to protect New Zealand biodiversity and eradicate bovine tuberculosis in the country. In order to investigate the possibility of pre-adaptation to "1080" resistance of possums in New Zealand we need to understand the molecular mechanisms responsible for toxin resistance in Western Australia. Does the resistance come from the differential expression of alleles already present in the brushtail possums' genome or is it dependent on alleles unique to the western Australian population?

Measuring differential expression

Differential expression analysis is a key application of RNASeq (RNA sequencing) that is used to identify genes that are expressed at different levels between two or more conditions and/or traits, such as different cell types, disease states, developmental stages or toxin resistance (Anders and Huber 2010; Lamarre et al. 2018a; Marioni et al. 2008). The goal of differential expression analysis is to identify genes that are significantly up- or down-regulated between the conditions being compared. This is achieved by comparing the number of RNA sequence reads that map to each gene in each sample, this count data is then normalized to account for differences in library size and other technical factors, and statistical tests are performed to identify genes that are differentially expressed between the conditions being compared (Korthauer et al. 2019; Lamarre et al. 2018a). There are many different statistical methods that can be used for differential expression analysis, but they generally involve comparing the

expression levels of each gene between the different conditions and calculating a p-value or q-value to determine whether the observed differences are statistically significant. In addition to identifying individual differentially expressed genes, it is also common to perform pathway analysis to identify biological pathways that are enriched for differentially expressed genes (Harris et al. 2004; Kanehisa and Goto 2000). The results of differential expression analysis can provide insights into the biological processes and pathways that are affected by the conditions being studied.

Measuring positive selection using Synonymous and Nonsynonymous substitutions

Molecular evolution refers to the study of changes in DNA sequences of organisms over time. This field is closely linked to the theory of evolution, which posits that species evolve through natural selection and genetic drift. Genetic drift is a process that occurs when random events alter the frequency of alleles from one generation to the next, natural selection, on the other hand, alters the frequency of alleles, based on the differential reproductive success of individuals with certain alleles, favouring those that confer advantageous traits in a given environment. One way to study molecular evolution is to analyse the rates of synonymous (dS) and nonsynonymous (dN) mutations within genes (Kimura 1977). Synonymous mutations are changes in DNA that do not alter the amino acid sequence of a protein, while nonsynonymous mutations result in changes to the amino acid sequence. Once a new mutation has become fixed in a population it is referred to as a "substitution" The $\omega = dN/dS$ ratio is a measure of the relative rates of nonsynonymous and synonymous substitubtions. A ratio of dN/dS greater than one suggests that the gene has experienced positive selection, meaning that certain amino acid changes have been favoured by natural selection. A ratio of dN/dS less than one suggests that the gene has experienced purifying selection, meaning that most amino acid changes are deleterious and have been removed from the population (Yang 2007; Yang and Bielawski 2000). We can try to identify genes that undergo positive selection in a specific lineage in order to associate those genes (and alleles) with the specific trait associated with the lineage.

Objectives and thesis outline

The goal of this research is to investigate the overall molecular and expression differences associated with the phenotype of the western Australian lineage, focused on toxin resistance. The main hypothesis is that toxin resistance that appeared in the western Australian brushtail possums is associated with lineage-specific differential expression as well as genes under positive selection. To investigate this hypothesis, I studied the evolutionary history of brushtail possum populations across their natural range before generating transcriptomic data to isolate differentially expressed genes and looked for genes under positive selection comparing the west with the eastern Australian lineage. By studying the molecular and genetic basis of the toxin resistance phenotype in brushtail possums, this research has implications for understanding how species adapt to environmental pressures and could inform the development of new pest control strategies in New Zealand and other parts of the world.

Chapter 2: Time-calibrated phylogeny and ecological niche models indicate Pliocene aridification drove intraspecific diversification of brushtail possums in Australia.

The aridification events in Australia during the Pliocene likely affected the distribution and structure of widespread species. I studied the impact of Pliocene and Pleistocene climate changes on the brushtail possum. I used ecological niche modelling to estimate the possum's potential distribution over six million years and found that populations were unlikely to have been in contact during the Pleistocene. The possum's most recent common ancestor dates back to the early Pliocene, when continental aridification caused significant changes to Australian ecology. I found distinct mitochondrial lineages for different subspecies of the possum, which likely experienced little gene flow with one another since the Pliocene, explaining their adaptations to regional plant assemblages. This chapter has been published in Ecology and Evolution DOI: 10.1002/ece3.9633.

Chapter 3: Differential expression analysis and exploration of the molecular pathways to brushtail possums' development.

This chapter presents a study using RNA sequencing to analyse gene expression changes during development in the brushtail possum. The analysis identified differentially expressed genes between adult and juvenile possums, male and female possums, and different fur colour phenotypes. The study identified 475 genes with significant changes in transcriptional levels during development, some of which were associated with tissue development, cell cycle, and extracellular matrix. The findings provide insights into the mechanisms associated with possum development and have implications for understanding phenotypic variation and adaptation in non-model organisms. An analysis of the minimum sample size suitable for differential gene expression provided a guide for sampling strategies.

Chapter 4: Differential gene expression among brushtail possums reveals metabolism discrepancies between subspecies.

The chapter examines the gene expression differences between brushtail possums from Western Australia and the eastern lineage (New Zealand). The study used RNA-Seq data to investigate gene expression differences, with a focus on identifying genetic pathways related to the western Australian possum's tolerance to the plant toxin sodium fluoroacetate. I used three approaches to identify differentially expressed genes associated with various metabolic pathways. The study found numerous significant differences in gene expression, demonstrating the evolutionary response to local selection resulting in genomic divergence of spatial populations. The identified gene expression differences provided candidate metabolic pathways most likely associated with sodium fluoroacetate resistance in these marsupials.

Chapter 5: Evidence of positive selection in the western Australian brushtail possum genome

This chapter reports on a selection analysis of the genome of brushtail possums, comparing the western lineage to the eastern lineage. The study used RNA-Seq data to identify genes under positive selection and the associated biological mechanisms. The main goal was to identify genetic pathways related to the western Australian possum's tolerance to the plant toxin sodium fluoroacetate. The study identified candidate genes under positive selection associated with glycan degradation and fatty acid metabolism. While the study allowed hypotheses about potential associations of these enriched pathways with fluoroacetate resistance, the list of candidate genes is still long.

Contributions

For each chapter, the research project was designed by myself (David Carmelet-Rescan), Steve A. Trewick and Mary Morgan Richard from Massey University. I carried the lab work, data analysis and writing under the supervision of Steve A. Trewick and Mary Morgan-Richards. Data were gathered from several different sources, sampling was realized thanks to Ralph Powlesland in the Manaroa region, Steve A. Trewick in the Manawatu area and thanks to Dale Chittenden on Stewart Island. In Australia, most of the data was generated locally with the association with Bethany Jackson at Murdoch University and Meg Rodgers at WA Wildlife for sampling. Additionally, some of the data used in Chapter 2 were generated by Nimeshika Pattabiraman from Massey University and the data in Chapter 3 by Donna Bond and Tim Hore from the University of Otago. Finally, the analyses needing high computation power were run using New Zealand eScience Infrastructure (NeSI) high-performance computing facilities (https://www.nesi.org.nz). This work was supported by the funding of OSPRI New Zealand Limited (Operational Solutions for Primary Industries) and Predator Free 2050.

References

Abbott, Ian. 2012. "Original Distribution of Trichosurus Vulpecula (Marsupialia : Phalangeridae) in Western Australia, with Particular Reference to Occurrence Outside the Southwest." *Journal of the Royal Society of Western Australia* 95: 83–93. https://www.rswa.org.au/publications/Journal/95(2)/Abbottpp.83-93.pdf (May 10, 2022).

Alanagreh, Lo'ai, Caitlin Pegg, Amritha Harikumar, and Mark Buchheim. 2017. "Assessing Intragenomic Variation of the Internal Transcribed Spacer Two: Adapting the Illumina Metagenomics Protocol." *PLoS ONE* 12(7): e0181491. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0181491 (February 10, 2022).

Alexander, Helen K., Guillaume Martin, Oliver Y. Martin, and Sebastian Bonhoeffer. 2014. "Evolutionary Rescue: Linking Theory for Conservation and Medicine." *Evolutionary Applications* 7(10): 1161–79.

Alterio, Nic. 2000. "Controlling Small Mammal Predators Using Sodium Monofluoroacetate (1080) in Bait Stations along Forestry Roads in a New Zealand Beech Forest." New Zealand Journal of Ecology 24(1): 3–9. https://www.jstor.org/stable/24054645

(September 25, 2023).

Anders, Simon, and Wolfgang Huber. 2010. "Differential Expression Analysis for Sequence Count Data." *Nature Precedings*: 1–1. https://www.nature.com/articles/npre.2010.4282.1 (July 5, 2022).

Andrews, Simon, and others. 2010. "FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010." *Https://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/* 1(1): http://www.bioinformatics.babraham.ac.uk/projects/. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/%0Ahttp://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/ (March 17, 2020).

Ansari, Mina Hojat et al. 2019. "Plio-Pleistocene Diversification and Biogeographic Barriers in Southern Australia Reflected in the Phylogeography of a Widespread and Common Lizard Species." *Molecular Phylogenetics and Evolution* 133: 107–19.

Archer, Michael et al. 2006. "Current Status of Species-Level Representation in Faunas from Selected Fossil Localities in the Riversleigh World Heritage Area, Northwestern Queensland." *Alcheringa* 30(4): 1–17. https://www.tandfonline.com/doi/abs/10.1080/03115510609506851 (October 1, 2021).

Bandelt, H J, P Forster, and A Rohl. 1999. "Median-Joining Networks for Inferring Intraspecific Phylogenies." *Molecular Biology and Evolution* 16(1): 37–48. https://academic.oup.com/mbe/article-lookup/doi/10.1093/oxfordjournals.molbev.a026036 (March 19, 2019).

Bass, Chris, Ian Denholm, Martin S. Williamson, and Ralf Nauen. 2015. "The Global Status of Insect Resistance to Neonicotinoid Insecticides." *Pesticide Biochemistry and Physiology* 121(10): 78–87. https://www.sciencedirect.com/science/article/pii/S0048357515000826 (March 15, 2020).

Beck, Robin M.D. 2008. "A Dated Phylogeny of Marsupials Using a Molecular Supermatrix and Multiple Fossil Constraints." *Journal of Mammalogy* 89(1): 175–89. https://academic.oup.com/jmammal/article/89/1/175/1020874 (October 1, 2021).

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57(1): 289–300.

Bernt, Matthias et al. 2013. "MITOS: Improved de Novo Metazoan Mitochondrial Genome Annotation." *Molecular Phylogenetics and Evolution* 69(2): 313–19.

Besnier, Francois et al. 2014. "Human-Induced Evolution Caught in Action: SNP-Array

Reveals Rapid Amphi-Atlantic Spread of Pesticide Resistance in the Salmon Ecotoparasite Lepeophtheirus Salmonis." *BMC Genomics* 15(1): 1–18.

Bindea, Gabriela et al. 2009. "ClueGO: A Cytoscape Plug-in to Decipher Functionally Grouped Gene Ontology and Pathway Annotation Networks." *Bioinformatics* 25(8): 1091–93. https://academic.oup.com/bioinformatics/article/25/8/1091/324247 (July 4, 2022).

Bintanja, Richard, Roderik S.W. Van De Wal, and Johannes Oerlemans. 2005. "Modelled Atmospheric Temperatures and Global Sea Levels over the Past Million Years." *Nature* 437(7055): 125–28. https://pubmed.ncbi.nlm.nih.gov/16136140/ (May 30, 2022).

Black, Karen H., Michael Archer, Suzanne J. Hand, and Henk Godthelp. 2012. "The Rise of Australian Marsupials: A Synopsis of Biostratigraphic, Phylogenetic, Palaeoecologic and Palaeobiogeographic Understanding." In *Earth and Life: Global Biodiversity, Extinction Intervals and Biogeographic Perturbations Through Time*, Springer Netherlands, 983–1078. https://link.springer.com/chapter/10.1007/978-90-481-3428-1_35 (June 21, 2021).

Bond, Donna M. et al. 2023. "The Admixed Brushtail Possum Genome Reveals Invasion History in New Zealand and Novel Imprinted Genes." *Nature Communications 2023 14:1* 14(1): 1–17. https://www.nature.com/articles/s41467-023-41784-8 (October 18, 2023).

Bouckaert, Remco R. 2010. "DensiTree: Making Sense of Sets of Phylogenetic Trees." *Bioinformatics* 26(10): 1372–73. https://academic.oup.com/bioinformatics/article-abstract/26/10/1372/192963 (September 8, 2021).

Boyle, C. Mary. 1960. "Case of Apparent Resistance of Rattus Norvegicus Berkenhout to Anticoagulant Poisons." *Nature* 188(4749): 517.

Braithwaite, R. W. 1990. "Australia's Unique Biota: Implications for Ecological Processes." *Journal of Biogeography* 17(4–5): 347–54.

Brown, Jason L. et al. 2018. "Paleoclim, High Spatial Resolution Paleoclimate Surfaces for Global Land Areas." *Scientific Data* 5. https://www.nature.com/articles/sdata2018254 (March 18, 2022).

Brown, Thomas M., and Gregory T. Payne. 1988. "Experimental Selection for Insecticide Resistance." *Journal of economic entomology* 81(1): 49–56.

Bryant, Litticia M., and Susan J. Fuller. 2014. "Pleistocene Climate Fluctuations Influence Phylogeographical Patterns in Melomys Cervinipes across the Mesic Forests of Eastern Australia." *Journal of Biogeography* 41(10): 1923–35. https://onlinelibrary.wiley.com/doi/full/10.1111/jbi.12341 (October 5, 2022).

Buaboocha, W., and R. T. Gemmell. 1997. "Development of Lung, Kidney and Skin in the

Brushtail Possum, Trichosurus Vulpecula." *Cells Tissues Organs* 159(1): 15–24. https://www.karger.com/Article/FullText/147960 (July 5, 2022).

Burridge, Christopher Paul. 2012. "Update: Divergence of Island Biotas When They Were Not Always Islands." *Frontiers of Biogeography* 3(4). https://escholarship.org/uc/item/74n463jr (October 17, 2022).

Byrne, M. 2008. "Evidence for Multiple Refugia at Different Time Scales during Pleistocene Climatic Oscillations in Southern Australia Inferred from Phylogeography." *Quaternary Science Reviews* 27(27–28): 2576–85.

de Castro, Érika C.P., Mika Zagrobelny, Márcio Z. Cardoso, and Søren Bak. 2018. "The Arms Race between Heliconiine Butterflies and Passiflora Plants – New Insights on an Ancient Subject." *Biological Reviews* 93(1): 555–73. https://onlinelibrary.wiley.com/doi/full/10.1111/brv.12357 (May 19, 2023).

Chapple, David G., J. Scott Keogh, and Mark N. Hutchinson. 2005. "Substantial Genetic Substructuring in Southeastern and Alpine Australia Revealed by Molecular Phylogeography of the Egernia Whitii (Lacertilia: Scincidae) Species Group." *Molecular Ecology* 14(5): 1279–92. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-294X.2005.02463.x (October 5, 2022).

Chaves-Campos, Johel, Steven G. Johnson, and C. Darrin Hulsey. 2011. "Spatial Geographic Mosaic in an Aquatic Predator-Prey Network." *PLoS ONE* 6(7): R75. http://genomebiology.biomedcentral.com/articles/10.1186/gb-2005-6-9-r75 (March 11, 2020).

Clark, Peter U. et al. 2009. "The Last Glacial Maximum." *Science* 325(5941): 710–14. https://pubmed.ncbi.nlm.nih.gov/19661421/ (May 30, 2022).

Clout, M, and K Ericksen. 2000. "Anatomy of a Disastrous Success: The Brushtail Possum as an Invasive Species." *The brushtail possum-biology, impact and management of an introduced marsupial*: 1–9. internal-pdf://0.0.3.94/10012031131.html.

Clout, M N, and Kris Ericksen. 2000. "Anatomy of a Disastrous Success: The Brushtail Possum as an Invasive Species." In *The Brushtail Possum: Biology, Impact and Management of an Introduced Marsupial*, , 1–9. https://ci.nii.ac.jp/naid/10012031131/ (March 15, 2020).

Coleman, Annette W. 2013. "Analysis of Mammalian RDNA Internal Transcribed Spacers." *PLoS ONE* 8(11). /pmc/articles/PMC3834078/ (February 10, 2022).

Conesa, Ana et al. 2016. "A Survey of Best Practices for RNA-Seq Data Analysis." *Genome Biology* 17(1): 1–19. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0881-8 (July 5, 2022).

Cooper, Christine E et al. 2018. "Geographical Variation in the Standard Physiology of Brushtail Possums ({Trichosurus}): Implications for Conservation Translocations." *Conserv Physiol* 6(1). https://academic.oup.com/conphys/article/6/1/coy042/5075449 (July 1, 2019).

Costa-Silva, Juliana, Douglas Domingues, and Fabricio Martins Lopes. 2017. "RNA-Seq Differential Expression Analysis: An Extended Review and a Software Tool." *PLoS ONE* 12(12): e0190152. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0190152 (July 5, 2022).

Cowan, P. 2016. "Characteristics and Behaviour of Brushtail Possums Initially Moving into a Depopulated Area." *New Zealand Journal of Zoology* 43(3): 223–33.

Cowan, P. E. 1989. "Changes in Milk Composition during Lactation in the Common Brushtail Possum, Trichosurus Vulpecula (Marsupialia: Phalangeridae)." *Reproduction, Fertility and Development* 1(4): 325–35. https://www.publish.csiro.au/rd/rd9890325 (July 5, 2022).

———. 1992. "The Eradication of Introduced Australian Brushtail Possums, Trichosurus Vulpecula, from Kapiti Island, a New Zealand Nature Reserve." *Biological Conservation* 61(3): 217–26. https://www.sciencedirect.com/science/article/pii/000632079291119D (March 15, 2020).

Cowan, P E. 1990. "Brushtail Possum." *The handbook of New Zealand mammals*: 68–89.

Cristiano, Luigi. 2021. "The Pseudogenes of Eukaryotic Translation Elongation Factors (EEFs): Role in Cancer and Other Human Diseases." *Genes & Diseases*.

Dalman, Mark R., Anthony Deeter, Gayathri Nimishakavi, and Zhong Hui Duan. 2012. "Fold Change and P-Value Cutoffs Significantly Alter Microarray Interpretations." *BMC bioinformatics* 13 Suppl 2(2): 1–4. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-S2-S11 (December 19, 2022).

Davis, M. B., and R. G. Shaw. 2001. "Range Shifts and Adaptive Responses to Quaternary Climate Change." *Science* 292(5517): 673–79. https://www.science.org/doi/full/10.1126/science.292.5517.673 (May 30, 2022).

Dawkins, R., and J. R. Krebs. 1979. "Arms Races between and within Species." *Proceedings of the Royal Society of London - Biological Sciences* 205(1161): 489–511.

Deakin, Janine E. et al. 2013. "Towards an Understanding of the Genetic Basis behind 1080 (Sodium Fluoroacetate) Tolerance and an Investigation of the Candidate Gene ACO2."

Australian Journal of Zoology 61(1): 69–77. http://www.publish.csiro.au/?paper=ZO12108 (March 15, 2020).

Denholm, I, and M W Rowland. 1992. *TACTICS FOR MANAGING PESTICIDE RESISTANCE IN ARTHROPODS: Theory and Practice*. www.annualreviews.org (March 12, 2020).

Dermauw, Wannes et al. 2013. "A Link between Host Plant Adaptation and Pesticide Resistance in the Polyphagous Spider Mite Tetranychus Urticae." *Proceedings of the National Academy of Sciences of the United States of America* 110(2): E113–22.

———. 2018. "Does Host Plant Adaptation Lead to Pesticide Resistance in Generalist Herbivores?" *Current Opinion in Insect Science* 26: 25–33.

Drummond, Alexei J., Marc A. Suchard, Dong Xie, and Andrew Rambaut. 2012. "Bayesian Phylogenetics with BEAUti and the BEAST 1.7." *Molecular Biology and Evolution* 29(8): 1969–73. https://academic.oup.com/mbe/article-abstract/29/8/1969/1044583 (September 8, 2021).

Dunlop, Judy et al. 2021. "Industry Environmental Offset Funding Facilitates a Large Multi-Species Fauna Translocation Program." *Pacific Conservation Biology* 102: 98–133. https://www.publish.csiro.au/pc/PC20036 (May 10, 2022).

Eason, Charles. 2002. "Sodium Monofluoroacetate (1080) Risk Assessment and Risk Communication." *Toxicology* 181–182: 523–30.

Eason, Charles, Aroha Miller, Shaun Ogilvie, and Alastair Fairweather. 2011. "An Updated Review of the Toxicology and Ecotoxicology of Sodium Fluoroacetate (1080) in Relation to Its Use as a Pest Control Tool in New Zealand." *New Zealand Journal of Ecology* 35: 1–20. https://www.jstor.org/stable/24060627 (March 13, 2020).

Edger, Patrick P. et al. 2015. "The Butterfly Plant Arms-Race Escalated by Gene and Genome Duplications." *Proceedings of the National Academy of Sciences of the United States of America* 112(27): 8362–66. https://www.pnas.org/doi/abs/10.1073/pnas.1503926112 (May 19, 2023).

Efford, Murray, Bruce Warburton, and Nick Spencer. 2000. "Home-Range Changes by Brushtail Possums in Response to Control." *Wildlife Research* 27(2): 117–27. http://www.publish.csiro.au/WR/WR99005 (March 15, 2020).

Ehrlich, Paul R., and Peter H. Raven. 1964. "Butterflies and Plants: A Study in Coevolution." *Evolution* 18(4): 586. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1558-5646.1964.tb01674.x (March 15, 2020).

El-Merhibi, Adaweyah et al. 2011. "Cytochrome P450 CYP3A in Marsupials: Cloning and

Characterisation of the Second Identified CYP3A Subfamily Member, Isoform 3A78 from Koala (Phascolarctos Cinereus)." *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* 154(4): 367–76.

Elena, Santiago F., and Richard E. Lenski. 2003. "Evolution Experiments with Microorganisms: The Dynamics and Genetic Bases of Adaptation." *Nature Reviews Genetics* 4(6): 457–69.

Endara, María José et al. 2017. "Coevolutionary Arms Race versus Host Defense Chase in a Tropical Herbivore–Plant System." *Proceedings of the National Academy of Sciences of the United States of America* 114(36): E7499–7505.

Endler, John A. 1977. "Geographic Variation, Speciation, and Clines." *Monographs in population biology* 10: 1–246.

Faith, J. Tyler et al. 2017. "Large Mammal Species Richness and Late Quaternary Precipitation Change in South-Western Australia." *Journal of Quaternary Science* 32(6): 760–69. https://onlinelibrary.wiley.com/doi/full/10.1002/jqs.2888 (October 1, 2021).

Felsenstein, J. 1985. "Phylogenies and the Comparative Method." *American Naturalist* 125(1): 1–15.

Felsenstein, Joseph, and Hirohisa Kishino. 1993. "Is There Something Wrong with the Bootstrap on Phylogenies? A Reply to Hillis and Bull." *Systematic Biology* 42(2): 193.

Fick, SE, RJ Hijmans - International journal of climatology, and undefined 2017. 2017. "WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas." *Wiley Online Library* 37(12): 4302–15. https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5086?casa_token=bTtWGlmoSeAAAAAA:M50yU39CG-YMgytWWhb-H7bD5RMri3sszTwKqCTWDbLda8xgwuUYj1tBKsMj952jSGuzlAbdlOYuafBF (September 8, 2021).

Firman, Renée C. et al. 2020. "Extreme and Variable Climatic Conditions Drive the Evolution of Sociality in Australian Rodents." *Current Biology* 30(4): 691-697.e3.

Flannery, T F, W D Turnbull, T H V Rich, and E L Lundelius. 1987. "The Phalangerids (Marsupialia: Phalangeridae) of the Early Pliocene Hamilton Local Fauna, Southwestern Victoria." *Possums and opossums: studies in evolution. Surrey Beatty and Sons and the Royal Zoological Society of New South Wales, Sydney*: 537–46.

Georghiou, George P., and Roni B. Mellon. 1983. "Pesticide Resistance in Time and Space." In *Pest Resistance to Pesticides*, Springer US, 1–46.

Griffitts, Joel S., and Raffi V. Aroian. 2005. "Many Roads to Resistance: How Invertebrates

Adapt to Bt Toxins." *BioEssays* 27(6): 614–24. https://onlinelibrary.wiley.com/doi/full/10.1002/bies.20239 (May 9, 2023).

Gupta, RC. 2015. "Handbook of Toxicology of Chemical Warfare Agents." https://books.google.com/books?hl=fr&lr=&id=MXKDBAAAQBAJ&oi=fnd&pg=PP1&dq=Goncharov+N,+et+al.+2015+Fluoracetate.+Chapter+16,+pages+193–214+in+Gupta+RC+(ed)+Handbook+of+Toxicology+of+Chemical+Warfare+Agents+(2nd+edition).+Academic+Press.+978-0-12-800159-2&ots=L0zowzfqi3&sig=zkRzHn12-_2tCGiRS4zNQ2OIWOM (March 13, 2020).

Harris, M. A. et al. 2004. "The Gene Oncology (GO) Database and Informatics Resource." *Nucleic Acids Research* 32(DATABASE ISS.): D258–61. https://academic.oup.com/nar/article/32/suppl_1/D258/2505186 (May 3, 2023).

Hart, Steven N. et al. 2013. "Calculating Sample Size Estimates for RNA Sequencing Data." *Journal of Computational Biology* 20(12): 970–78. https://pubmed.ncbi.nlm.nih.gov/23961961/ (July 5, 2022).

Hawkins, Nichola J., Chris Bass, Andrea Dixon, and Paul Neve. 2019. "The Evolutionary Origins of Pesticide Resistance." *Biological Reviews* 94(1): 135–55. http://doi.wiley.com/10.1111/brv.12440 (March 15, 2020).

He, Yuxin, and Huanye Wang. 2021. "Terrestrial Material Input to the Northwest Shelf of Australia Through the Pliocene-Pleistocene Period and Its Implications on Continental Climates." *Geophysical Research Letters* 48(17): e2021GL092745. https://onlinelibrary.wiley.com/doi/full/10.1029/2021GL092745 (May 30, 2022).

Helfrich, Philipp et al. 2019. "Treeannotator: Versatile Visual Annotation of Hierarchical Text Relations." In *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, , 1958–63. https://www.aclweb.org/anthology/L18-1308.pdf (September 8, 2021).

Hewitt, Godfrey M. 2004. "Genetic Consequences of Climatic Oscillations in the Quaternary." In *Philosophical Transactions of the Royal Society B: Biological Sciences*, The Royal Society, 183–95. /pmc/articles/PMC1693318/?report=abstract (May 30, 2022).

Hijmans, RJ et al. 2017. "Package 'Dismo.'" *gr.xemacs.org*. https://scholar.google.com/ftp://ftp.gr.xemacs.org/mirrors/CRAN/web/packages/dismo/dismo.pdf (March 18, 2022).

Hocknull, Scott A. et al. 2020. "Extinction of Eastern Sahul Megafauna Coincides with Sustained Environmental Deterioration." *Nature Communications* 11(1): 1–14. https://www.nature.com/articles/s41467-020-15785-w (March 9, 2022).

How, R. A., and S. J. Hillcox. 2000. “Brushtail Possum, Trichosurus Vulpecula, Populations in South-Western Australia: Demography, Diet and Conservation Status.” *Wildlife Research* 27(1): 81–89. https://www.publish.csiro.au/wr/wr98064 (July 5, 2022).

Huelsenbeck, John P., and David M. Hillis. 1993. “Success of Phylogenetic Methods in the Four-Taxon Case.” *Systematic Biology* 42(3): 247–64. https://academic.oup.com/sysbio/article/42/3/247/1629468 (September 30, 2021).

Jackson, R. et al. 2019. “Ecology of a Brushtail Possum (Trichosurus Vulpecula) Population at Castlepoint in the Wairarapa, New Zealand.” *New Zealand Journal of Ecology* 43(2).

Jalili, Vahid et al. 2021. “The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2020 Update.” *Nucleic Acids Research* 48(W1): W395–402. https://academic.oup.com/nar/article/48/W1/W395/5849904 (July 4, 2022).

James, C. H., and C. Moritz. 2000. “Intraspecific Phylogeography in the Sedge Frog Litoria Fallax (Hylidae) Indicates Pre-Pleistocene Vicariance of an Open Forest Species from Eastern Australia.” *Molecular Ecology* 9(3): 349–58. https://onlinelibrary.wiley.com/doi/full/10.1046/j.1365-294x.2000.00885.x (January 10, 2022).

Jarman, Peter. 2006. “Life of Marsupials.” *Austral Ecology* 31(5): 670–670. https://books.google.com/books?hl=fr&lr=&id=gDFeNMhJlPUC&oi=fnd&pg=PP7&dq=life+of+marsupials&ots=e8g6ILivtt&sig=CnZRgadgx3IxwwmrYAZcxgptNlU (July 5, 2022).

Jombart, Thibaut. 2008. “Adegenet: A R Package for the Multivariate Analysis of Genetic Markers.” *Bioinformatics* 24(11): 1403–5.

Jombart, Thibaut, Sébastien Devillard, and François Balloux. 2010. “Discriminant Analysis of Principal Components: A New Method for the Analysis of Genetically Structured Populations.” *BMC Genetics* 11(1): 1–15. https://bmcgenomdata.biomedcentral.com/articles/10.1186/1471-2156-11-94 (February 16, 2022).

Kanehisa, Minoru, and Susumu Goto. 2000. “KEGG: Kyoto Encyclopedia of Genes and Genomes.” *Nucleic Acids Research* 28(1): 27–30. https://academic.oup.com/nar/article/28/1/27/2384332 (May 3, 2023).

Karageorgi, Marianthi et al. 2019. “Genome Editing Retraces the Evolution of Toxin Resistance in the Monarch Butterfly.” *Nature* 574(7778): 409–12. https://pubmed.ncbi.nlm.nih.gov/31578524/ (May 9, 2023).

Kearse, M et al. 2012. “Geneious Basic: An Integrated and Extendable Desktop Software

Platform for the Organization and Analysis of Sequence Data." *Bioinformatics* 28(12): 1647–49. https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bts199.

Kerle, A., G. M. McKay, and G. B. Sharman. 1991. "A Systematic Analysis of the Brushtail Possum, Trichosurus Vulpecula (Kerr, 1792) (Marsupialia: Phalangeridae)." *Australian Journal of Zoology* 39(3): 263–71. https://www.publish.csiro.au/zo/zo9910313 (November 23, 2021).

Kerle, A. 1984. "Variation in the Ecology of Trichosurus: Its Adaptive Significance." *Possums and gliders*: 115–28.

———. 2002. "Possums: The Brushtails, Ringtails and Greater Glider." *Choice Reviews Online* 39(10): 39-5816-39–5816. https://books.google.com/books?hl=fr&lr=&id=YDM0hjAwchUC&oi=fnd&pg=PR7&dq=opossum+Kerle&ots=4gYdItzt2O&sig=T5sZ2OXT-lr669669KDdfxjw7G4 (September 23, 2022).

Kerle, A, and RA How. 2008. "Common Brushtail Possum." In *The Mammals of Australia*, , 274–76.

Kim, Daehwan et al. 2019. "Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype." *Nature Biotechnology 2019 37:8* 37(8): 907–15. https://www.nature.com/articles/s41587-019-0201-4 (July 4, 2022).

Kimura, Motoo. 1977. "Preponderance of Synonymous Changes as Evidence for the Neutral Theory of Molecular Evolution [33]." *Nature* 267(5608): 275–76. https://www.nature.com/articles/267275a0 (April 24, 2023).

King, D. R., A. J. Oliver, and R. J. Mead. 1978. "The Adaptation of Some Western Australian Mammals to Food Plants Containing Fluoroacetate." *Australian Journal of Zoology* 26(4): 699–712.

Koot, Emily M., Mary Morgan-Richards, and Steven A. Trewick. 2022. "Climate Change and Alpine-Adapted Insects: Modelling Environmental Envelopes of a Grasshopper Radiation." *Royal Society Open Science* 9(3). https://royalsocietypublishing.org/doi/abs/10.1098/rsos.211596 (March 23, 2022).

Korthauer, Keegan et al. 2019. "A Practical Guide to Methods Controlling False Discoveries in Computational Biology." *Genome Biology* 20(1): 1–21. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1716-1 (August 16, 2022).

Kozulin, Peter et al. 2022. "Divergent Evolution of Developmental Timing in the Neocortex

Revealed by Marsupial and Eutherian Transcriptomes." *Development (Cambridge)* 149(3). https://journals.biologists.com/dev/article/149/3/dev200212/274319/Divergent-evolution-of-developmental-timing-in-the (July 5, 2022).

Kruskal, Joseph B. 1956. "On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem." *Proceedings of the American Mathematical Society* 7(1): 48–50. https://www.ams.org/proc/1956-007-01/S0002-9939-1956-0078686-7/ (February 16, 2022).

Kuch, Ulrich et al. 2005. "Phylogeography of Australia's King Brown Snake (Pseudechis Australis) Reveals Pliocene Divergence and Pleistocene Dispersal of a Top Predator." *Naturwissenschaften* 92(3): 121–27. https://link.springer.com/article/10.1007/s00114-004-0602-0 (October 5, 2022).

Kung, Janet W.C., Ian S. Currie, Stuart J. Forbes, and James A. Ross. 2010. "Liver Development, Regeneration, and Carcinogenesis." *Journal of Biomedicine and Biotechnology* 2010.

Lamarre, Sophie et al. 2018a. "Optimization of an RNA-Seq Differential Gene Expression Analysis Depending on Biological Replicate Number and Library Size." *Frontiers in Plant Science* 9: 108.

———. 2018b. "Optimization of an RNA-Seq Differential Gene Expression Analysis Depending on Biological Replicate Number and Library Size." *Frontiers in Plant Science* 9: 108.

Lanfear, Robert et al. 2017. "Partitionfinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses." *Molecular Biology and Evolution* 34(3): 772–73. https://academic.oup.com/mbe/article-abstract/34/3/772/2738784 (March 16, 2020).

Langfelder, Peter, and Steve Horvath. 2008. "WGCNA: An R Package for Weighted Correlation Network Analysis." *BMC Bioinformatics* 9(1): 1–13. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559 (December 19, 2022).

Larsson, Johan, and Peter Gustafsson. 2018. "{eulerr}: Area-Proportional Euler and Venn Diagrams with Ellipses." *Proceedings of {{International Workshop}} on {{Set Visualization}} and {{Reasoning}}* 2116: 84–91. https://cran.r-project.org/package=eulerr (December 19, 2022).

Leigh, Jessica W, and David Bryant. 2015. "Popart: Full-Feature Software for Haplotype Network Construction." *Methods in Ecology and Evolution* 6(9): 1110–16.

https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12410 (June 17, 2019).

Lenski, Richard E. 2017. "Convergence and Divergence in a Long-Term Experiment with Bacteria." *The American naturalist* 190(S1): S57–68. http://www.ncbi.nlm.nih.gov/pubmed/28731830 (March 12, 2020).

Leong, Lex Ee Xiang et al. 2017. "Fluoroacetate in Plants - a Review of Its Distribution, Toxicity to Livestock and Microbial Detoxification." *Journal of Animal Science and Biotechnology* 8(1): 55. http://jasbsci.biomedcentral.com/articles/10.1186/s40104-017-0180-6 (March 13, 2020).

Li, Tingting et al. 2009. "Multi-Stage Analysis of Gene Expression and Transcription Regulation in C57/B6 Mouse Liver Development." *Genomics* 93(3): 235–42.

Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. "FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features." *Bioinformatics* 30(7): 923–30. https://academic.oup.com/bioinformatics/article/30/7/923/232889 (July 4, 2022).

Livingstone, P. G., N. Hancox, G. Nugent, and G. W. de Lisle. 2015. "Toward Eradication: The Effect of Mycobacterium Bovis Infection in Wildlife on the Evolution and Future Direction of Bovine Tuberculosis Management in New Zealand." *New Zealand Veterinary Journal* 63: 4–18.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15(12): 1–21. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8 (July 4, 2022).

Lynch, A. Jasmyn J., Robert J.S. Beeton, and Penelope Greenslade. 2019. "The Conservation Significance of the Biota of Barrow Island, Western Australia." *Journal of the Royal Society of Western Australia* 102: 98–133. https://www.researchgate.net/profile/A-Lynch/publication/336855839_The_conservation_significance_of_the_biota_of_Barrow_Island_Western_Australia/links/5dc261f94585151435ec6ff5/The-conservation-significance-of-the-biota-of-Barrow-Island-Western-Australia.p (May 10, 2022).

Maclean, Craig. 2010. "The Population Genetics of Antibiotic Resistance: Integrating Molecular Mechanisms and Treatment Contexts." *nature.com*. https://www.researchgate.net/publication/44608740 (March 15, 2020).

Macnair, M. R. 1991. "Why the Evolution of Resistance to Anthropogenic Toxins Normally Involves Major Gene Changes: The Limits to Natural Selection." *Genetica* 84(3): 213–

19.

Macqueen, Peggy et al. 2010. “Phylogenetics of the Pademelons (Macropodidae: Thylogale) and Historical Biogeography of the Australo-Papuan Region.” *Molecular Phylogenetics and Evolution* 57(3): 1134–48.

Mallet, James. 2008. “Hybridization, Ecological Races and the Nature of Species: Empirical Evidence for the Ease of Speciation.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 363(1506): 2971–86. https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2008.0081 (March 14, 2022).

Marioni, John C. et al. 2008. “RNA-Seq: An Assessment of Technical Reproducibility and Comparison with Gene Expression Arrays.” *Genome Research* 18(9): 1509–17. https://genome.cshlp.org/content/18/9/1509.full (July 12, 2022).

Martin, H. A. 2006. “Cenozoic Climatic Change and the Development of the Arid Vegetation in Australia.” *Journal of Arid Environments* 66(3 SPEC. ISS.): 533–63.

Martin, H A. 1998. “Tertiary Climatic Evolution and the Development of Aridity in Australia.” In *Proceedings-Linnean Society Of New South Wales*, , 115–36.

Martin, Marcel. 2011. “Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads.” *EMBnet.journal* 17(1): 10.

Mcgowran, Brian, and Phil Bock. 2000. “Article in Memoirs of the Association of Australasian Palaeontologists.” https://www.researchgate.net/publication/292144424 (October 1, 2021).

McIlroy, J. C. 1983. “The Sensitivity of the Brushtail Possum (Trichosurus Vulpecula) to 1080 Poison (Sodium Monfluoroacetate).” *New Zealand Journal of Ecology* 6: 125–31. https://www.jstor.org/stable/24052734 (March 15, 2020).

Mead, R. J., A. J. Oliver, D. R. King, and P. H. Hubach. 1985. “The Co-Evolutionary Role of Fluoroacetate in Plant-Animal Interactions in Australia.” *Oikos* 44(1): 55.

Meredith, Robert W., Michael Westerman, and Mark S. Springer. 2009. “A Phylogeny of Diprotodontia (Marsupialia) Based on Sequences for Five Nuclear Genes.” *Molecular Phylogenetics and Evolution* 51(3): 554–71. https://www.sciencedirect.com/science/article/pii/S1055790309000414 (June 24, 2021).

Meredith, Robert W, Michael Westerman, Judd A Case, and Mark S Springer. 2008. “A Phylogeny and Timescale for Marsupial Evolution Based on Sequences for Five Nuclear Genes.” *Journal of Mammalian Evolution* 15(1): 1–36.

Miller, Gifford H. et al. 2005. “Anthropology: Ecosystem Collapse in Pleistocene Australia and a Human Role in Megafaunal Extinction.” *Science* 309(5732): 287–90.

https://www.science.org/doi/abs/10.1126/science.1111288 (March 9, 2022).

Mittermeier, Russell A. 2004. *Hotspots Revisited*. Cemex.

Modepalli, Vengamanaidu et al. 2018. "Gene Expression Profiling of Postnatal Lung Development in the Marsupial Gray Short-Tailed Opossum (Monodelphis Domestica) Highlights Conserved Developmental Pathways and Specific Characteristics during Lung Organogenesis." *BMC Genomics* 19(1): 732. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6173930/ (August 18, 2019).

Montague, T. L. (Thomas L.). 2000. *The Brushtail Possum : Biology, Impact and Management of an Introduced Marsupial*. https://trove.nla.gov.au/work/32729533?q&versionId=46367736 (March 15, 2020).

Müllner, Daniel. 2013. "Fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python." *Journal of Statistical Software* 53(9): 1–18. https://cran.r-project.org/package=fastcluster (December 19, 2022).

Myers, Norman et al. 2000. "Biodiversity Hotspots for Conservation Priorities." *Nature* 403(6772): 853–58. https://www.nature.com/articles/35002501 (November 8, 2021).

Naimi, Babak et al. 2014. "Where Is Positional Uncertainty a Problem for Species Distribution Modelling?" *Ecography* 37(2): 191–203. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0587.2013.00205.x (March 18, 2022).

Neaves, Linda E. et al. 2016. "Phylogeography of the Koala, (Phascolarctos Cinereus), and Harmonising Data to Inform Conservation." *PLoS ONE* 11(9): e0162207. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0162207 (February 11, 2022).

Neher, R. A., B. I. Shraiman, and D. S. Fisher. 2010. "Rate of Adaptation in Large Sexual Populations." *Genetics* 184(2): 467–81. https://academic.oup.com/genetics/article/184/2/467/6077864 (May 30, 2022).

Neve, Paul, Roberto Busi, Michael Renton, and Martin M. Vila-Aiub. 2014. "Expanding the Eco-Evolutionary Context of Herbicide Resistance Research." *Pest Management Science* 70(9): 1385–93.

New, Tim R. 2011. "'In Considerable Variety': Introducing the Diversity of Australia's Insects." *'In Considerable Variety': Introducing the Diversity of Australia's Insects*.

Nogués-Bravo, David et al. 2018. "Cracking the Code of Biodiversity Responses to Past Climate Change." *Trends in Ecology and Evolution* 33(10): 765–76. https://pubmed.ncbi.nlm.nih.gov/30173951/ (March 23, 2022).

Nugent, G., P. Sweetapple, J. Coleman, and P. Suisted. 2000. "Possum Feeding Patterns: Dietary Tactics of a Reluctant Folivore." In *The Brushtail Possum: Biology, Impact and Management of an Introduced Marsupial*, , 10–19.

O'Reilly-Wapstra, Julianne M., and Phil Cowan. 2010. "Native Plant/Herbivore Interactions as Determinants of the Ecological and Evolutionary Effects of Invasive Mammalian Herbivores: The Case of the Common Brushtail Possum." *Biological Invasions* 12(2): 373–87.

Ogilvie, S C et al. 2009. *Uptake of 1080 by Watercress and Puha-Culturally-+lmportant Plants Used for Food By*. http://dspace.lincoln.ac.nz/handle/10182/1389 (March 13, 2020).

Oliver, A. J., and D. R. King. 1979. "Fluoroacetate Tolerance, a Genetic Marker in Some Australian Mammals." *Australian Journal of Zoology* 27(3): 331–47. https://www.publish.csiro.au/zo/zo9790363 (April 22, 2022).

Oliver, Paul M. et al. 2017. "A Novel Hotspot of Vertebrate Endemism and an Evolutionary Refugium in Tropical Australia." *Diversity and Distributions* 23(1): 53–66. https://onlinelibrary.wiley.com/doi/full/10.1111/ddi.12506 (October 5, 2022).

Oliver, Paul M., Mark Adams, and Paul Doughty. 2010. "Molecular Evidence for Ten Species and Oligo-Miocene Vicariance within a Nominal Australian Gecko Species (Crenadactylus Ocellatus, Diplodactylidae)." *BMC Evolutionary Biology* 10(1): 1–11. https://bmcecolevol.biomedcentral.com/articles/10.1186/1471-2148-10-386 (October 5, 2022).

Orzech, Kathryn M, and Mark Nichter. 2008. "From Resilience to Resistance: Political Ecological Lessons from Antibiotic and Pesticide Resistance." www.annualreviews.org (March 12, 2020).

Palumbi, S. R. 2001. "Humans as the World's Greatest Evolutionary Force." *Science* 293(5536): 1786–90.

Paradis, Emmanuel. 2010. "Pegas: An R Package for Population Genetics with an Integrated–Modular Approach." *Bioinformatics* 26(3): 419–20. https://academic.oup.com/bioinformatics/article/26/3/419/215731 (March 26, 2019).

Paranjpe, Monika et al. 2019. "Transcriptomic Analysis of MAP3K1 and MAP3K4 in the Developing Marsupial Gonad." *Sexual Development* 13(4): 195–204. https://www.karger.com/Article/FullText/505799 (July 5, 2022).

Pattabiraman, Nimeshika, Mary Morgan-Richards, Ralph Powlesland, and Steven A. Trewick. 2021. "Unrestricted Gene Flow between Two Subspecies of Translocated Brushtail Possums (Trichosurus Vulpecula) in Aotearoa New Zealand." *Biological Invasions* 24(1):

247–60. https://link.springer.com/article/10.1007/s10530-021-02635-z (May 31, 2022).

Peng, Lai et al. 2012. "RNA Sequencing Reveals Dynamic Changes of MRNA Abundance of Cytochromes P450 and Their Alternative Transcripts during Mouse Liver Development." *Drug Metabolism and Disposition* 40(6): 1198–1209. https://pubmed.ncbi.nlm.nih.gov/22434873/ (December 21, 2022).

Perez, John C., Sathit Pichyangkul, and Vivian E. Garcia. 1979. "The Resistance of Three Species of Warm-Blooded Animals to Western Diamondback Rattlesnake (Crotalus Atrox) Venom." *Toxicon* 17(6): 601–7. https://pubmed.ncbi.nlm.nih.gov/524385/ (May 2, 2023).

Peterson, B., C. C. Bezuidenhout, and J. Van den Berg. 2017. "An Overview of Mechanisms of Cry Toxin Resistance in Lepidopteran Insects." *Journal of Economic Entomology* 110(2): 362–77. https://pubmed.ncbi.nlm.nih.gov/28334065/ (May 10, 2023).

Phillips, M. J., Y. H. Lin, G. L. Harrison, and D. Penny. 2001. "Mitochondrial Genomes of a Bandicoot and a Brushtail Possum Confirm the Monophyly of Australidelphian Marsupials." *Proceedings of the Royal Society B: Biological Sciences* 268(1475): 1533–38.

Poran, Naomie S., Richard G. Coss, and Eli Benjamini. 1987. "Resistance of California Ground Squirrels (Spermophilus Beecheyi) to the Venom of the Northern Pacific Rattlesnake (Crotalus Viridis Oreganus): A Study of Adaptive Variation." *Toxicon* 25(7): 767–77. https://pubmed.ncbi.nlm.nih.gov/3672545/ (May 2, 2023).

Porder, Stephen. 2014. "Coevolution of Life and Landscapes." *Proceedings of the National Academy of Sciences of the United States of America* 111(9): 3207–8. www.pnas.org/cgi/doi/10.1073/pnas.1400954111 (March 9, 2022).

Potter, Sally et al. 2012. "Phylogenetic Relationships of Rock-Wallabies, Petrogale (Marsupialia: Macropodidae) and Their Biogeographic History within Australia." *Molecular Phylogenetics and Evolution* 62(2): 640–52.

Powles, Stephen B., and Qin Yu. 2010. "Evolution in Action: Plants Resistant to Herbicides." *Annual Review of Plant Biology* 61(1): 317–47.

Pracy, L. 1974. "Opposums." *New Zealand Nature Heritage* 3(32): 873–82.

Prideaux, Gavin J. et al. 2007. "Mammalian Responses to Pleistocene Climate Change in Southeastern Australia." *Geology* 35(1): 33–36. http://pubs.geoscienceworld.org/gsa/geology/article-pdf/35/1/33/3532085/i0091-7613-35-1-33.pdf (February 16, 2022).

R Development Core Team. 2021. "R: A Language and Environment for Statistical Computing

(Version 4.1. 2)[Computer Software].” *R Foundation for Statistical Computing*.

Rambaut, Suchard, Xie, and A Drummond. 2015. “Tracer v1.7.” http://tree.bio.ed.ac.uk/software/tracer/ (March 16, 2020).

Reid, Noah M., and Bryan C. Carstens. 2012. “Phylogenetic Estimation Error Can Decrease the Accuracy of Species Delimitation: A Bayesian Implementation of the General Mixed Yule-Coalescent Model.” *BMC Evolutionary Biology* 12(1): 1–11. https://bmcecolevol.biomedcentral.com/articles/10.1186/1471-2148-12-196 (February 11, 2022).

Reznick, David N., and Cameron K. Ghalambor. 2001. “The Population Ecology of Contemporary Adaptations: What Empirical Studies Reveal about the Conditions That Promote Adaptive Evolution.” *Genetica* 112–113(1): 183–98. https://link.springer.com/article/10.1023/A:1013352109042 (May 30, 2022).

Ritchie, Matthew E. et al. 2015. “Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies.” *Nucleic Acids Research* 43(7): e47. /pmc/articles/PMC4402510/ (December 19, 2022).

Rix, Michael G et al. 2015. “Biogeography and Speciation of Terrestrial Fauna in the South-Western Australian Biodiversity Hotspot.” *Biological Reviews* 90(3): 762–93.

Robbins, Robert K. et al. 2021. “A Switch to Feeding on Cycads Generates Parallel Accelerated Evolution of Toxin Tolerance in Two Clades of Eumaeus Caterpillars (Lepidoptera: Lycaenidae).” *Proceedings of the National Academy of Sciences of the United States of America* 118(7): e2018965118. https://www.pnas.org/doi/abs/10.1073/pnas.2018965118 (May 2, 2023).

Roberts, Karen Kristine, Mina Bassarova, and Michael Archer. 2008. “Oligo-Miocene Ringtail Possums of the Genus Paljara (Pseudocheiridae: Marsupialia) from Queensland, Australia.” *Geobios* 41(6): 833–44. https://www.sciencedirect.com/science/article/pii/S0016699508000806 (September 30, 2021).

Roberts, R. G. et al. 2001. “New Ages for the Last Australian Megafauna: Continent-Wide Extinction about 46,000 Years Ago.” *Science* 292(5523): 1888–92. https://www.science.org/doi/abs/10.1126/science.1060264 (March 9, 2022).

Robinson, Weldon B. 1953. “Coyote Control with Compound 1080 Stations in National Forests.” *Journal of Forestry* 51(12): 880–85.

Robles, José A. et al. 2012. “Efficient Experimental Design and Analysis Strategies for the Detection of Differential Expression Using RNA-Sequencing.” *BMC Genomics* 13(1): 1–

14. https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-13-484 (July 4, 2022).

Ronquist, Fredrik, and John P Huelsenbeck. 2003. "MrBayes 3: Bayesian Phylogenetic Inference under Mixed Models." *Bioinformatics* 19(12): 1572–74. https://academic.oup.com/bioinformatics/article-abstract/19/12/1572/257621 (March 16, 2020).

Rosenheim, Jay A. et al. 1996. "Biochemical Peradaptations, Founder Events, and the Evolution of Resistance in Arthropods." *Journal of Economic Entomology* 89(2): 263–73.

Ross, James G., Kathryn Bicknell, and G.J. Hickling. 1999. "COST-EFFECTIVE CONTROL OF 1080 BAIT-SHY POSSUMS."

Roush, R, and BE Tabashnik. 2012. "Pesticide Resistance in Arthropods." https://books.google.com/books?hl=fr&lr=&id=ooHjBwAAQBAJ&oi=fnd&pg=PA7&dq=pesticide+resistance+review+&ots=b4HTRVe3DT&sig=gatc_jXjdzJN2icrjAZpQkbtNgo (March 12, 2020).

Rowe, Ashlee H., and Matthew P. Rowe. 2008. "Physiological Resistance of Grasshopper Mice (Onychomys Spp.) to Arizona Bark Scorpion (Centruroides Exilicauda) Venom." *Toxicon* 52(5): 597–605. https://pubmed.ncbi.nlm.nih.gov/18687353/ (May 2, 2023).

Rowe, Kevin C. et al. 2008. "Pliocene Colonization and Adaptive Radiations in Australia and New Guinea (Sahul): Multilocus Systematics of the Old Endemic Rodents (Muroidea: Murinae)." *Molecular Phylogenetics and Evolution* 47(1): 84–101.

Saltré, Frédérik et al. 2016. "Climate Change Not to Blame for Late Quaternary Megafauna Extinctions in Australia." *Nature Communications* 7(1): 1–7. https://www.nature.com/articles/ncomms10511 (March 9, 2022).

Scheuerl, Thomas, Johannes Cairns, Lutz Becks, and Teppo Hiltunen. 2019. "Predator Coevolution and Prey Trait Variability Determine Species Coexistence." *Proceedings of the Royal Society B: Biological Sciences* 286(1902).

Seliskar, Matej, and Damjana Rozman. 2007. "Mammalian Cytochromes P450-Importance of Tissue Specificity." *Biochimica et Biophysica Acta - General Subjects* 1770(3): 458–66. https://pubmed.ncbi.nlm.nih.gov/17097232/ (July 5, 2022).

Seyednasrollah, Fatemeh, Asta Laiho, and Laura L. Elo. 2013. "Comparison of Software Packages for Detecting Differential Expression in RNA-Seq Studies." *Briefings in Bioinformatics* 16(1): 59–70. https://academic.oup.com/bib/article/16/1/59/240754 (December 19, 2022).

Shiao, Yih-Horng. 2021. *Discoveries and Biological Implications of Mammalian 45S RDNA*

Variants and Non-Structural RRNAs. Preprints. https://www.preprints.org/manuscript/202102.0613/v1 (February 11, 2022).

Smith, Kathleen K., and Anna L. Keyte. 2020. "Adaptations of the Marsupial Newborn: Birth as an Extreme Environment." *Anatomical Record* 303(2): 235–49. https://onlinelibrary.wiley.com/doi/full/10.1002/ar.24049 (March 6, 2023).

Smoot, Michael E. et al. 2011. "Cytoscape 2.8: New Features for Data Integration and Network Visualization." *Bioinformatics* 27(3): 431–32. https://academic.oup.com/bioinformatics/article/27/3/431/321742 (July 4, 2022).

Sniderman, J. M. K. et al. 2016. "Pliocene Reversal of Late Neogene Aridification." *Proceedings of the National Academy of Sciences of the United States of America* 113(8): 1999–2004. www.pnas.org/cgi/doi/10.1073/pnas.1520188113 (May 30, 2022).

Sniderman, J. M.K. et al. 2019. "Vegetation and Climate Change in Southwestern Australia During the Last Glacial Maximum." *Geophysical Research Letters* 46(3): 1709–20. https://onlinelibrary.wiley.com/doi/full/10.1029/2018GL080832 (October 1, 2021).

Sochorová, Jana et al. 2021. "Analyses of the Updated 'Animal Rdna Loci Database' with an Emphasis on Its New Features." *International Journal of Molecular Sciences* 22(21): 11403. https://www.mdpi.com/1422-0067/22/21/11403/htm (February 11, 2022).

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30(9): 1312–13. https://academic.oup.com/bioinformatics/article-abstract/30/9/1312/238053 (March 16, 2020).

Steel, Mike, and Andy McKenzie. 2001. "Properties of Phylogenetic Trees Generated by Yule-Type Speciation Models." *Mathematical Biosciences* 170(1): 91–112.

Strasburg, Jared L., Michael Kearney, Craig Moritz, and Alan R. Templeton. 2007. "Combining Phylogeography with Distribution Modeling: Multiple Pleistocene Range Expansions in a Parthenogenetic Gecko from the Australian Arid Zone." *PLoS ONE* 2(8): e760. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000760 (October 5, 2022).

Szücs, M. et al. 2017. "Rapid Adaptive Evolution in Novel Environments Acts as an Architect of Population Range Expansion." *Proceedings of the National Academy of Sciences of the United States of America* 114(51): 13501–6. www.pnas.org/cgi/doi/10.1073/pnas.1712934114 (May 30, 2022).

Tao, Wenjing et al. 2018. "Transcriptome Display during Tilapia Sex Determination and Differentiation as Revealed by RNA-Seq Analysis." *BMC Genomics* 19(1): 1–12.

https://link.springer.com/articles/10.1186/s12864-018-4756-0 (July 5, 2022).

Taylor, Andrea C. et al. 2004. "High Microsatellite Diversity and Differential Structuring among Populations of the Introduced Common Brushtail Possum, Trichosurus Vulpecula, in New Zealand." *Genetical Research* 83(2): 101–11.

van Thiel, Jory et al. 2022. "Convergent Evolution of Toxin Resistance in Animals." *Biological Reviews* 97(5): 1823–43. https://onlinelibrary.wiley.com/doi/full/10.1111/brv.12865 (May 1, 2023).

Thompson, Jim A., and Peter J.S. Fleming. 1994. "Evaluation of the Efficacy of 1080 Poisoning of Red Foxes Using Visitation to Non-Toxic Baits as an Index of Fox Abundance." *Wildlife Research* 21(1): 91–104.

Thuiller, Wilfried et al. 2016. "Package 'Biomod2.'" *Species distribution modeling within an ensemble forecasting framework.*

Thuiller, Wilfried, Damien Georges, and Robin Engler. 2015. "Package 'Biomod2': Ensemble Platform for Species Distribution Modeling."

Todd, Erica V., Michael A. Black, and Neil J. Gemmell. 2016. "The Power and Promise of RNA-Seq in Ecology and Evolution." *Molecular Ecology* 25(6): 1224–41. https://onlinelibrary.wiley.com/doi/full/10.1111/mec.13526 (July 12, 2022).

Trelstad, R. L. 1973. "The Developmental Biology of Vertebrate Collagens." *Journal of Histochemistry and Cytochemistry* 21(6): 521–28.

Triggs, S. J., and W. Q. Green. 1989. "Geographic Patterns of Genetic Variation in Brushtail Possums Trichosurus Vulpecula and Implications for Pest Control." *New Zealand Journal of Ecology* 12: 1–10. https://science.sciencemag.org/content/293/5536/1786.short?casa_token=_2rXnNJtJmwAAAAA:4ONmMBnglQOi7ZQqbP3A0z22dV7pzPhvDSFT_1h9wVi12Vjh7yP-rGi4ieBRnRrR7QzNRQWqjQBPp1g (March 13, 2020).

Turney, Chris S.M. et al. 2008. "Late-Surviving Megafauna in Tasmania, Australia, Implicate Human Involvement in Their Extinction." *Proceedings of the National Academy of Sciences of the United States of America* 105(34): 12150–53. www.pnas.org/cgi/content/full/ (March 9, 2022).

Twigg, L. E. et al. 1996. "Fluoroacetate Content of Some Species of the Toxic Australian Plant Genus, Gastrolobium, and Its Environmental Persistence." *Natural Toxins* 4(3): 122–27. http://doi.wiley.com/10.1002/19960403NT4 (March 15, 2020).

———. 2003. "Sensitivity of Some Australian Animals to Sodium Fluoroacetate (1080): Additional Species and Populations, and Some Ecological Considerations." *Australian*

Journal of Zoology 51(5): 515–31. https://www.publish.csiro.au/zo/zo03040 (March 23, 2022).

Twigg, L. E., and D. R. King. 1991. "The Impact of Fluoroacetate-Bearing Vegetation on Native Australian Fauna: A Review." *Oikos* 61(3): 412. https://www.jstor.org/stable/3545249?casa_token=wxSjPb9PE-QAAAAA:AqA216Efy0uOSzZJwBTgcg7Eo64RP0a9Hmm8Ua_rf4Faa-Qp-yot16PIkohRXPufbktoxRP5zU0RpZ0KIa9YOsEXHe5Co014rR-2rRyJe0hgP43FRiJi (March 15, 2020).

Ujvari, Beata et al. 2015. "Widespread Convergence in Toxin Resistance by Predictable Molecular Evolution." *Proceedings of the National Academy of Sciences of the United States of America* 112(38): 11911–16. www.pnas.org/cgi/doi/10.1073/pnas.1511706112 (March 15, 2020).

Umbrello, Linette S., Raphael K. Didham, Ric A. How, and Joel A. Huey. 2020. "Multi-Species Phylogeography of Arid-Zone Sminthopsinae (Marsupialia: Dasyuridae) Reveals Evidence of Refugia and Population Expansion in Response to Quaternary Change." *Genes 2020, Vol. 11, Page 963* 11(9): 963. https://www.mdpi.com/2073-4425/11/9/963/htm (February 11, 2022).

Valavi, Roozbeh, Jane Elith, José J. Lahoz-Monfort, and Gurutzeta Guillera-Arroita. 2019. "BlockCV: An r Package for Generating Spatially or Environmentally Separated Folds for k-Fold Cross-Validation of Species Distribution Models." *Methods in Ecology and Evolution* 10(2): 225–32. https://www.biorxiv.org/content/10.1101/357798.full (September 8, 2021).

VanDerWal, Jeremy et al. 2014. "Package 'SDMTools.'" *R package*.

Vignali, S, AG Barras, … R Arlettaz - Ecology and, and undefined 2020. 2020. "SDMtune: An R Package to Tune and Evaluate Species Distribution Models." *Wiley Online Library* 10(20): 11488–506. https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.6786 (September 8, 2021).

Wang, Sho Ya, and Ging Kuo Wang. 2017. "Single Rat Muscle Na+ Channel Mutation Confers Batrachotoxin Autoresistance Found in Poison-Dart Frog Phyllobates Terribilis." Proceedings of the National Academy of Sciences of the United States of America 114(39): 10491–96. https://www.pnas.org/doi/abs/10.1073/pnas.1707873114 (September 25, 2023).

Williams, Kristen J et al. 2011. "Forests of East Australia: The 35th Biodiversity Hotspot." In *Biodiversity Hotspots*, , 295–310.

de Wit, Cynthia A., and Björn R. Weström. 1987. "Venom Resistance in the Hedgehog, Erinaceus Europaeus: Purification and Identification of Macroglobulin Inhibitors as Plasma Antihemorrhagic Factors." *Toxicon* 25(3): 315–23. https://pubmed.ncbi.nlm.nih.gov/3590212/ (May 2, 2023).

Wood, R.J. 1981. "Insecticide Resistance: Genes and Mechanisms." *The Genetic Consequences of Man Made Change.*: Pp 53-96. http://agris.fao.org/agris-search/search.do?recordID=US201302634817 (March 12, 2020).

Worth, J. R.P. et al. 2017. "Habitat Type and Dispersal Mode Underlie the Capacity for Plant Migration across an Intermittent Seaway." *Annals of Botany* 120(4): 539–49. https://academic.oup.com/aob/article/120/4/539/4049543 (April 25, 2022).

Yang, Ziheng. 2007. "PAML 4: Phylogenetic Analysis by Maximum Likelihood." *Molecular Biology and Evolution* 24(8): 1586–91. https://academic.oup.com/mbe/article/24/8/1586/1103731 (March 10, 2023).

Yang, Ziheng, and Joseph R. Bielawski. 2000. "Statistical Methods for Detecting Molecular Adaptation." *Trends in Ecology and Evolution* 15(12): 496–503. https://pubmed.ncbi.nlm.nih.gov/11114436/ (May 3, 2023).

Yatsu, Ryohei et al. 2016. "RNA-Seq Analysis of the Gonadal Transcriptome during Alligator Mississippiensis Temperature-Dependent Sex Determination and Differentiation." *BMC Genomics* 17(1): 1–13. https://link.springer.com/articles/10.1186/s12864-016-2396-9 (July 5, 2022).

Young, Matthew, Maintainer Nadia Davidson, Anthony Hawkins, and G O biocViews Sequencing. 2013. "Package 'Goseq.'"

Yu, Guangchuang. 2021. "Enrichplot: Visualization of Functional Enrichment Result." https://yulab-smu.top/biomedical-knowledge-mining-book/.

Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. "ClusterProfiler: An R Package for Comparing Biological Themes among Gene Clusters." *Omics: a journal of integrative biology* 16(5): 284–87.

Yuan, Michael L., and Ian J. Wang. 2018. "Sodium Ion Channel Alkaloid Resistance Does Not Vary with Toxicity in Aposematic Dendrobates Poison Frogs: An Examination of Correlated Trait Evolution." PLoS ONE 13(3). https://pubmed.ncbi.nlm.nih.gov/29534110/ (September 25, 2023).

Zachos, J et al. 2001. "Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present." *Science* 292(5517): 686–93. https://science.sciencemag.org/content/292/5517/686.abstract (October 1, 2021).

Zhang, Zong Hong et al. 2014. "A Comparative Study of Techniques for Differential Expression Analysis on RNA-Seq Data." *PLoS ONE* 9(8): e103207. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0103207 (July 5, 2022).

Zuur, Alain F., Elena N. Ieno, and Chris S. Elphick. 2010. "A Protocol for Data Exploration to Avoid Common Statistical Problems." *Methods in Ecology and Evolution* 1(1): 3–14. https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/J.2041-210x.2009.00001.X (March 18, 2022).

Chapter 2: Time-calibrated phylogeny and ecological niche models indicate Pliocene aridification drove intraspecific diversification of brushtail possums in Australia.

Abstract

Major aridification events in Australia during the Pliocene may have had significant impact on the distribution and structure of widespread species. To explore the potential impact of Pliocene and Pleistocene climate oscillations we estimated the timing of population fragmentation and past connectivity of the currently isolated but morphologically similar subspecies of the widespread brushtail possum (*Trichosurus vulpecula*). We use ecological niche modelling (ENM) with the current fragmented distribution of brushtail possum to estimate the environmental envelope of this marsupial. We projected the ENM on models of past climatic conditions in Australia to infer the potential distribution of brushtail possums over six million years. D-loop haplotypes were used to describe population structure. From shotgun sequencing we assembled whole mitochondrial DNA genomes and estimated timing of intraspecific divergence. Our projections of ENMs suggest current possum populations were unlikely to have been in contact during the Pleistocene. Although lowered sea level during glacial periods enabled colonisation of Tasmania, climate fluctuation during this time would not have facilitated gene flow. The most recent common ancestor of sampled intraspecific diversity dates to the early Pliocene when continental aridification caused significant changes to Australian ecology and *Trichosurus vulpecula* distribution was likely fragmented. Phylogenetic analysis revealed that the subspecies *T. v. hypoleucus* (koomal; southwest), *T. v. arnhemensis* (langkurr; north) and *T. v. vulpecula* (bilda; southeast) correspond to distinct mitochondrial lineages. Despite little phenotypic differentiation, *Trichosurus vulpecula* populations probably experienced little gene flow with one another since the Pliocene, supporting the recognition of several subspecies and explaining their adaptations to the regional plant assemblages on which they feed.

Keywords: Australia, climate cycles, ecological niche modelling, marsupials, molecular dating, possums, subspecies

Note for the reader:

This chapter was published in 2022 in Ecology and Evolution. Parts including nuclear markers are an addition not included in the published manuscript.

Reference: Carmelet-Rescan, D., Morgan-Richards, M., Pattabiraman, N., & Trewick, S. A. (2022). Time-calibrated phylogeny and ecological niche models indicate Pliocene aridification

drove intraspecific diversification of brushtail possums in Australia. Ecology and Evolution, 12(12), e9633.

Background

Genetic signatures of range expansion, gene flow and former barriers provide information that can help our understanding of past and future biological response to environmental change. Phylogeographic studies of Australian animals reveal several different responses to past climatic oscillations. Some studies show that, similar to the Northern hemisphere, Pleistocene climate cycles were major drivers of species phylogeographical patterns (Bryant & Fuller, 2014; Strasburg et al., 2007). But the majority of phylogeographic studies of Australian taxa suggest older climate fluctuations explain current genetic structure (Kuch et al., 2005; P. M. Oliver et al., 2010, 2017), especially among marsupials (Macqueen et al., 2010; Potter et al., 2012; Rowe et al., 2008). A rich fossil record of Australian megafauna allows calibration of changes in faunal composition (Hocknull et al., 2020; Miller et al., 2005; Roberts et al., 2001; Saltré et al., 2016; Turney et al., 2008) and it has been suggested that natural climate cycling during the Pleistocene had little impact on mammal assemblages (Prideaux et al., 2007). To determine whether current patterns of genetic structure within a single species date to events that occurred more than 2 million years ago we examined the phylogeography of the Australian brushtail possum *Trichosurus vulpecula*.

Among Australia's endemic marsupial fauna, the brushtail possum *Trichosurus vulpecula* is one of the most widespread with a geographic range spanning the continent of nearly 7.7 million km^2. Populations of brushtail possum exist at both the east and west extremities of Australia, some 4000km apart, and the species spans the temperate south to the tropical north (Figure 1). Brushtail possums are arboreal herbivores and across their geographic range they interact with regional floras in ecologically distinct regions such as Jarrah and Kerri eucalyptus forest in south-west Australia, tropical rainforests of Queensland and temperate Tasmanian vegetation (Kerle, 1984). Despite this wide geographic and ecological range, the existence of numerous Aboriginal names (Abbott 2012), and subspecies names for geographic races, no reliable diagnostic traits distinguishing brushtail possum populations are apparent from skull morphometrics, allozymes or chromosomes (A. Kerle et al., 1991). The six subspecies of *Trichosurus vulpecula* are recognised based on geographic location, size and fur colour (How

and Kerle 1995), and show considerable physiological variation across Australia (Cooper et al., 2018). Brushtail possum populations differ in their exposure to plant chemical defences as they forage on plants known to contain compounds toxic to mammals including *Erythophleum*, *Acacia*, *Eucalyptus* and *Gastrolabium*, and as a result have evolved different tolerances to toxins. For example, brushtail possums in South-west Australia (koomal) have an LD50 160 times higher than their cousins (bilda) in east Australia (Twigg et al., 1996) when exposed to potent toxin fluoroacetate (known as 1080; Leong et al., 2017). This emergence of physiological adaptations specific to regional flora demonstrates the coevolutionary interaction resulting in genomic divergence of spatial populations (Mead et al., 1985; A. J. Oliver & King, 1979), which implies a sustained interaction counter to the impression drawn from morphology.

Although *Trichosurus vulpecula* occurs widely in Australia, population density is uneven; there are large areas in the west and south of the country where the species is not recorded (Figure 2a) and this could help explain lineage/subspecies distinctions. Lack of gene flow between brushtail possum populations in different habitats would allow fixing of local adaptations (Mallet, 2008). We predicted that past climate variation resulted in range shifts among brushtail possum populations across the Australian continent that resulted in a connection of habitat and populations that are now isolated.-Using modern records of brushtail possum occurrence we infer the environmental envelope of *Trichosurus vulpecula* and extrapolate past distributions based on models of past climate. We then use independent time-calibrated phylogenetic analysis to test for the expected correlation between the timing of niche fragmentation and lineage splitting using a lineage-specific rate of molecular evolution based on a fossil-calibrated analysis.

Methods

Six subspecies of brushtail possum *Trichosurus vulpecula* are currently recognised (A. Kerle et al., 1991), of which five are represented in our data: langkurr *T. v. arnhemensis*; koomal *T. v. hypoleucus*; bilda *T. v. vulpecula*; *T. v. johnstonii; T. v. fuliginosus* (Figure 1). Only *T. v. eburacensis* from Cape York is unsampled here.

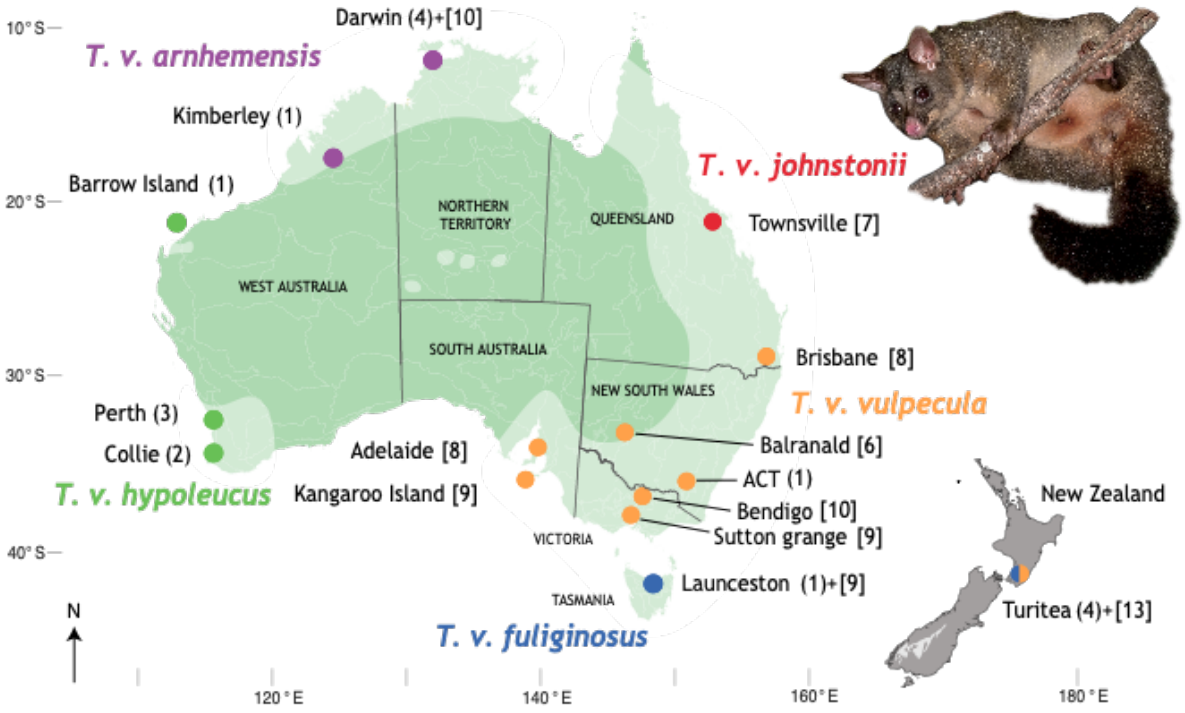


Figure 1: Sampling locations in Australian and New Zealand of *Trichosurus vulpecula* brushtail possums used for phylogenetic analyses. Spot colours coded by putative subspecies identification. Number of samples providing mitogenome sequences for molecular dating (in parentheses) and the number of samples providing D-loop sequence [in square brackets]. In lighter green is the putative current distribution of brushtail possum in Australia (Kerle, 2002). Grey brushtail possum inset credit: Tony Jewell.

Niche Modelling

To model the environmental envelope of *Trichosurus vulpecula* we obtained occurrence data for the species in Australia from the Atlas of living Australia (ala.org.au) which provided us with data from 80 different datasets (Appendix C2) including museum records from regions where they are not present anymore (dating as early as 1857) (Kerle, 2002). Filtering of duplicate records at a 0.1˚ latitude and longitude resolution resulted in a dataset of 6297 records. Niche modelling approaches typically require both presence and absence data, but true absence information is difficult to verify. In this analysis, we used an alternative to true absence records by generating pseudo-absence data using the surface range envelop algorithm (SRE) from R package 'Biomod2' (Thuiller et al., 2015). 3000 pseudo-absence points were generated in locations that have contrasting environmental conditions to the presence locations (Figure 2a). Defining and projecting the climatic envelope of *Trichosurus vulpecula* required current and past climatic data (Current, Mid-Holocene, Last Glacial Maximum, Last interglacial period, Mid-pliocene warm period and Pliocene M2) for Australia and New Zealand (where translocated brushtail possums are thriving), and these were obtained from the Worldclim and Paleoclim databases (Brown et al., 2018; Fick et al., 2017). Multivariate environmental

similarity surfaces analysis (MESS) was computed using the 'dismo' R package (Hijmans et al., 2017) to check for the presence of non-analogous conditions in all scenarios of projection and to select only appropriate predictors that would not require models to extrapolate into novel climates. Multicollinearity of predictors was assessed using variance inflation factors (VIF, Zuur et al., 2010) retaining predictors with VIF smaller than 10 using 'usdm' R package (Naimi et al., 2014). To investigate potential local differences and encompass Australia's broad climatic diversity and subspecies delimitations the presence-absence data were separated into four spatial blocks using the R package 'BlockCV' (Valavi et al., 2019; Figure 2a, *T. v. fuliginosus* and *T. v. vulpecula* were integrated in the same block, and *T. v. johnsoni* with *T. v. eburacensis*). The four spatial blocks were iteratively removed from the ecological niche models. We used a genetic algorithm from the R package 'SDMtune' (Vignali et al., 2020) to optimize the models hyperparameter configuration and obtain more robust models (Gradient Boosting Machine (GBM), Random Forest (RF), artificial neural network (ANN) and Maximum Entropy (Maxent); Appendix C2 Table 1). The models were then computed using 'Biomod2' in R (Thuiller et al., 2016) with 80% of the input data used to calibrate the models, with the remaining 20% used to test them. Presences were weighted inversely proportional to the sampling abundance in order to give more importance to under-sampled areas. VarImport was set to 3, allowing three permutations to estimate variable importance. Each model was run on each of the four different blocks four times, and model accuracy of each run was assessed using receiver operating characteristic (ROC) and true skill statistic (TSS). An ensemble model was then generated from the qualifying subset of runs with ROC score > 0.80. The importance of each variable was calculated to assess the main drivers (or their proxies) of the alternative models and response curves were plotted to visualize the effect of the variables on the predicted habitat suitability. Projections of the ensemble model were then computed on current and past climate conditions using ensemble mean weights model (EMmw); cut-off values were determined by maximizing sensitivity (the proportion of observed presence correctly predicted) and specificity (the proportion of observed pseudo-absence correctly predicted) of the model. Range change statistics were computed using the R package 'SDMTools' (VanDerWal et al., 2014).

Acquisition of DNA sequence data

Tissue samples of *Trichosurus vulpecula* were provided by the Australian National Wildlife at Canberra (ANWC) and the Australian Biological Tissue Collection (ABTC) at South Australia Museum. New Zealand samples were donated from pest-control trapping in the Manawatu

region, North island (Turitea) and added to this study in order to increase the sample size of the population from Tasmania and Victoria, which is their origin (Pracy, 1974). Total genomic DNA was extracted from 106 specimens using the GeneAid Tissue DNA Isolation Kit following the manufacturers guidelines (Figure 1), and quantity and quality checked using PerkinElmer LabChip® GX Touch HT.

We examined the genealogy and population diversity of *Trichosurus vulpecula* across Australia with mtDNA D-loop sequences (Neaves et al., 2016; Umbrello et al., 2020) (Figure 1). Polymerase Chain Reaction (PCR) amplification targeted a 730 bp fragment of the mtDNA D-loop with PCR primers Tcan_218f and Tvul_1023r designed for *Trichosurus,* using conditions described previously (Pattabiraman et al., 2022). Amplification products were sequenced using BigDye® chemistry (Perkin Elmer) on an ABI3730 DNA analyser. Sequences were edited and aligned using the software GENEIOUS PRIME (v. 2021.1.1, BioMatters; (Kearse et al., 2012)).

Whole genomic DNA from two short-eared possums (*Trichosurus caninus*) and a representative subsample of 17 *Trichosurus vulpecula* (Figure 1, Appendix C2 Table 3) specimens were selected for shotgun sequencing to provide raw data for assembly of whole mitochondrial genomes. High-throughput DNA sequencing on the BGI DNBSEQ-G50 platform was used to generate up to 100 million 100bp paired-end reads per sample. These were paired and mapped to a reference mitochondrial genome (Genbank ID: NC_003039, (Phillips et al., 2001)) in GENEIOUS PRIME (v. 2021.1.1, BioMatters; (Kearse et al., 2012)), using the GENEIOUS internal mapper and at least five iterations. Once a draft had been assembled, the data were remapped to the consensus sequence and the result annotated by comparison to a reference and independently via the MITOS-webservice (Bernt et al., 2013), and verified by translation of putative coding regions. These mitochondrial sequences were combined to produce an alignment of 19 DNA sequences of 15,462 bp (Genbank accessions ON400078-ON400096, see Appendix C2 Table 3). D-loop sequences from the 17 *Trichosursus vulpecula* were extracted from the mitogenomes and included to the following analysis using this marker.

Population statistics

For visualization of evolutionary relationships among the sequence variants of D-loop median-joining and minimum spanning haplotype networks (Bandelt et al., 1999; Kruskal, 1956) were inferred using POPART (Leigh & Bryant, 2015) and R package 'pegas' (Paradis, 2010).

Discriminant Analysis of Principle Components (DAPC : Jombart et al., 2010) was perform on the D-loop alignment using the 'adegenet' package in R (Jombart, 2008) in order to describe the different population clusters. DAPC uses Principal Component analysis to select the components holding the most variance followed and then applies Discriminant Analysis that navigates among this variability to provide the clearest distinction between groups. We retained components that defined a total of 90% of the variance and we used subspecies as the prior grouping factor for the analysis.

Phylogenetic analysis

The newly assembled mtDNA genomes from 17 *Trichosurus vulpecula* were aligned with two mtDNA genomes from *T. caninus*. Phylogenetic relationships among these were inferred using maximum likelihood and Bayesian inference methods using RAXML 8.2.12 (Stamatakis, 2014) and MRBAYES v3.2.7 (Ronquist & Huelsenbeck, 2003). First, the data were divided into four partitions separating RNA (12S,16S and tRNA) and each codon position of protein coding genes. PARTITIONFINDER v2.1.1 (Lanfear et al., 2017) was then used to select the best fitting substitution model for the data for each partitions under corrected Akaike Information Criterion (AICc). This analysis determined the best substitution models to be GTR+G for all partitions and we used RAXML to compute the maximum likelihood analyses with the default tree search approach using simultaneous nearest neighbour interchange method and a neighbour joining tree as starting tree to estimate the ML tree topologies with 1,000 bootstrap replicates (Felsenstein, 1985; Felsenstein & Kishino, 1993; Huelsenbeck & Hillis, 1993). Bayesian inference analyses was performed using unlinked parameters for each partition and allowing branch lengths to vary proportionately across partitions. The analysis consisted of two independent runs each with two simultaneous Markov Chain Monte Carlo (MCMC) chains of 30 million generations sampled every 1,000 generations. Convergence and stationarity of runs and burn-in period (first 25%) were determined using TRACER v1.7 (Rambaut et al., 2015).

Time calibration

We created an alignment of the mitochondrial genome from 14 species representing Dasyuromorphia and Diprotodontia the two biggest Australian marsupials' orders (NCBI Genbank; Appendix C2 Table 5) and including both *Trichosurus vulpecula* and *T. caninus.* A molecular dating analysis was performed with this alignment to generate branch rate estimates that could be applied to dating analysis of intraspecific *Trichosurus* lineages. The data were partitioned into tRNA and rRNA, D-loop, CDS 1st and 2nd codon positions, and CDS 3rd codon positions using BEAUTI (Drummond et al., 2012) to input the priors and parameters for the BEAST model. Substitution models for each partition were determined using the "BeastModel" algorithm. Model parameters included an optimised relaxed clock to allow the substitution rates to vary among the different species. A Yule model was implemented as it is appropriate for a phylogeny of mammal species as it assumes lineages split (constant speciation rate) without extinctions (Reid & Carstens, 2012; Steel & McKenzie, 2001). For time calibration of this phylogeny we added priors estimated in a previous analyses of marsupial evolution using fossil calibrations (Meredith et al., 2009). That analysis of a DNA sequence from five nuclear genes (protein-coding portions of ApoB, BRCA1, IRBP, Rag1, and vWF) representing nine placental lineages and 22 marsupials included 32 hard time constraints based on both the fossil record and previous phylogenetic analyses (Meredith et al., 2008).

We applied mean times and 95% highest posterior density (HDP; to accommodate error around past estimates) to the nodes in our analysis corresponding to the *Australelphia*, *Macropodiformes* + *Phalangeroidea*, and *Phalangeridae* lineages with corresponding normal probability distributions, 63.0 Ma (55.6-70.0), 45.3 Ma (39.0-52.0) and 17.4 Ma (13.6-21.7) respectively. The Markov chain Monte Carlo (MCMC) chain length was set to 100 million generations and sampled every 1000 generations and the model was then run through BEAST v2.6.5 (Drummond et al., 2012). Convergence of chains and ESS of the beast run were checked using TRACER (Rambaut et al., 2015), using the ESS score (>200) to assess that all the parameters properly converged. This provided node calibration for the split between *Trichosurus vulpecula and T. caninus* in a similar analysis of data representing 17 *T. vulpecula* individuals and two *T. caninus*. For this we implemented a coalescent constant population model in BEAST with an optimized relaxed clock to determine the mutation rate of each partition in the species', and a strict clock was then set to produce the final dated phylogeny. We used TREEANNOTATOR (Helfrich et al., 2019) to compile the trees from the Beast runs into

a maximum clade credibility tree using a standard burn-in of the first 10% trees and median node heights. DENSITREE and FIGTREE were used to visualize trees (Bouckaert, 2010).

Results

Ecological niche modelling

After filtering presence data, we had 5757 locality records of brushtail possum across Australia (Figure 2a). Mapping these records revealed five regions where living possums are documented in Australia, separated by areas of apparently uninhabited land (or sea). We retained five bioclimatic variables (following MESS and VIF analysis) that could be used over the region of study and over the four regions and six climatic conditions. Other bioclimatic variables were discarded to avoid extrapolating the models to never explored values and to improve robustness of models and projections (Appendix C2 Figure 1). Of the 64 runs 55 had a ROC score greater than 0.80 and were retained for ensemble modelling (Appendix C2 Table 2). The ensemble model showed a good fit to the current range of *Trichosurus vulpecula*, with a cut-off value of 635.5 that maximized specificity (>98%) but still had a very good sensitivity of > 93%. Plotting the projected ENM used this cut-off value (635.5) and 70% of this value to represent the available niche space (Figure 2b). The modelled potential distribution (where suitable climate currently exists) closely matched the current occupancy of brushtail possums with the addition of potentially suitable niche in central-west and central-east Australia where only a few presences were recorded, mainly corresponding to museum samples. In general, it appears that the current distribution of brushtail possums is probably constrained by the climate variables or their proxies in our models. Estimates of the importance of each the climatic variables included in the model and their response curves revealed that the most influential parameters were precipitation of the driest month and precipitation of the warmest quarter in the MaxEnt model and precipitation of the wettest month for the other models (GBM, ANN and RF) or their proxies (Appendix C2 Figure 2).

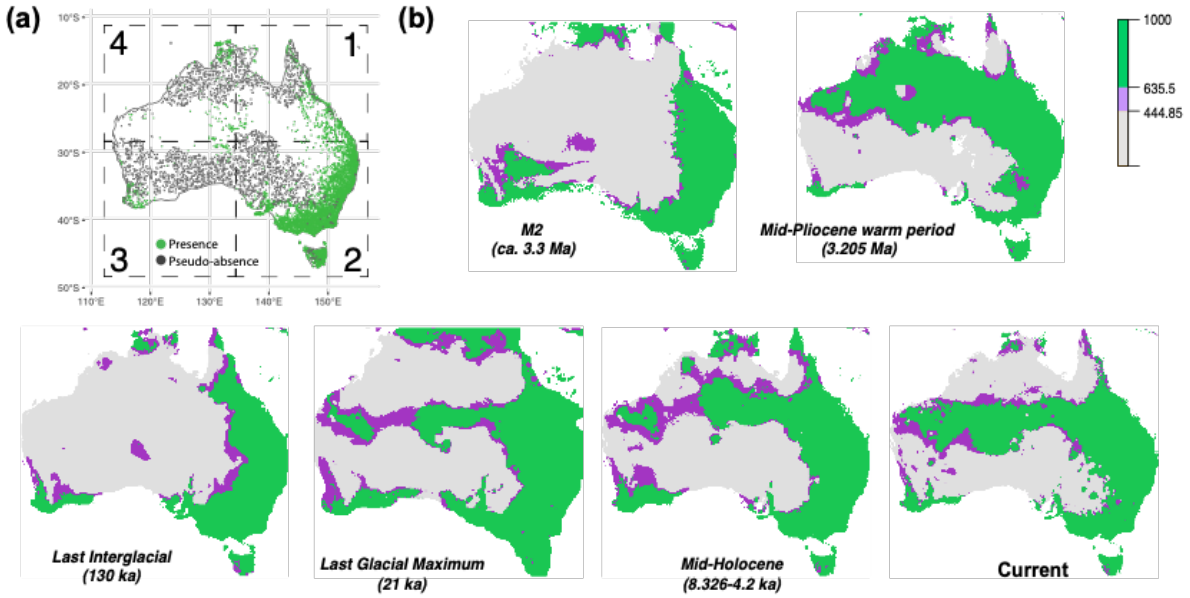


Figure 2: a) Presence-absence points of *Trichosurus vulpecula* in Australia used for environmental niche modelling, with presence records from the Atlas of living Australia database and pseudo-absence generated using the biomod2 SRE algorithm. Numbered squares represent the spatial blocks defined by blockCV and run separately to be combined in the ensemble modelling. **b)** Weighted mean ensemble ecological niche modelling projections on Current and past climate data (Mid Holocene, Last Glacial Maximum, Last Interglacial period, Mid-Pliocene warm period and Pliocene M2). Pixel of colours represent ensemble modelling scores; green, scores above the cutoff maximizing sensitivity (635.5); purple, scores >70% of this cutoff value (444.85); grey, low scores signifying low probability of presence.

Projections of ENM on to models of past climatic conditions suggest wider potential brushtail possum distribution during the LGM (Figure 2b). An expanded distribution would have facilitated gene flow between populations. Suitable habitat for brushtail possums probably expanded during cold glacial periods and retracted during warm interglacials. The implied expansion during the LGM of suitable niche space corresponds to increased land surface as a result of lower sea levels and would have increased the opportunity for connection between previously isolated populations. Our niche modelling suggests habitat connection during the LGM between Tasmanian forests and southern Australia, and between available niche-space in northern and eastern Australia. However, the habitat of *T. v. hypoleucus* in southwestern Australia remained isolated from all other suitable regions during the LGM (Figure 2b). We suggest that the availability of habitat would have been similar during all previous climate oscillation of the late Pleistocene (interglacial/glacial cycle), limiting contact between eastern and western populations. Our mid-Holocene and Last interglacial projections infer a similar distribution of suitable niche-space to current conditions. The projection of our possum ENM onto the climate model for the Mid-Pliocene warm period reveals potential niche-space in

Northern Territory central Australia and north of Western Australia, contiguous with potential habitat in eastern Australia (Figure 2b). In contrast, the projection of our possum ENM on to the climate model for 3.3 million years ago (Pliocene M2 period) suggests a wider availability of potential habitat in the south east and a possible habitat connection between eastern and western niche-space on the south coast.

Haplotypic diversity

Aligned D-loop (Appendix C2 Table 4) sequence data comprised 561 base pairs for 106 specimens and revealed 44 haplotypes differing from one another by up to 7% (uncorrected). The haplotype network revealed clusters broadly concordant with sampling locations (Figure 3a). These clusters also appear in the DAPC analysis (Figure 3b) and correlated with the taxa *T. v. hypoleucus*, *T. v. arnhemensis* and *T. v. johnstonii,* and clustering of the south eastern samples (*T. v. vulpecula* and *T. v. fulginosus*). Largest mtDNA distances were between *T. v. arnhemensis* in the north and T. *v. hypoleucus* in the southwest. The Tasmanian (*T. v. fulginosus*) haplotypes nest among the haplotype diversity from New South Wales, Victoria and South Australia *T. v. vulpecula* but are not shared. As expected, the haplotypes obtained from New Zealand brushtail possums are most similar to those of *T. v. fulginosus* and *T. v. vulpecula*, but were unique within our sampling.

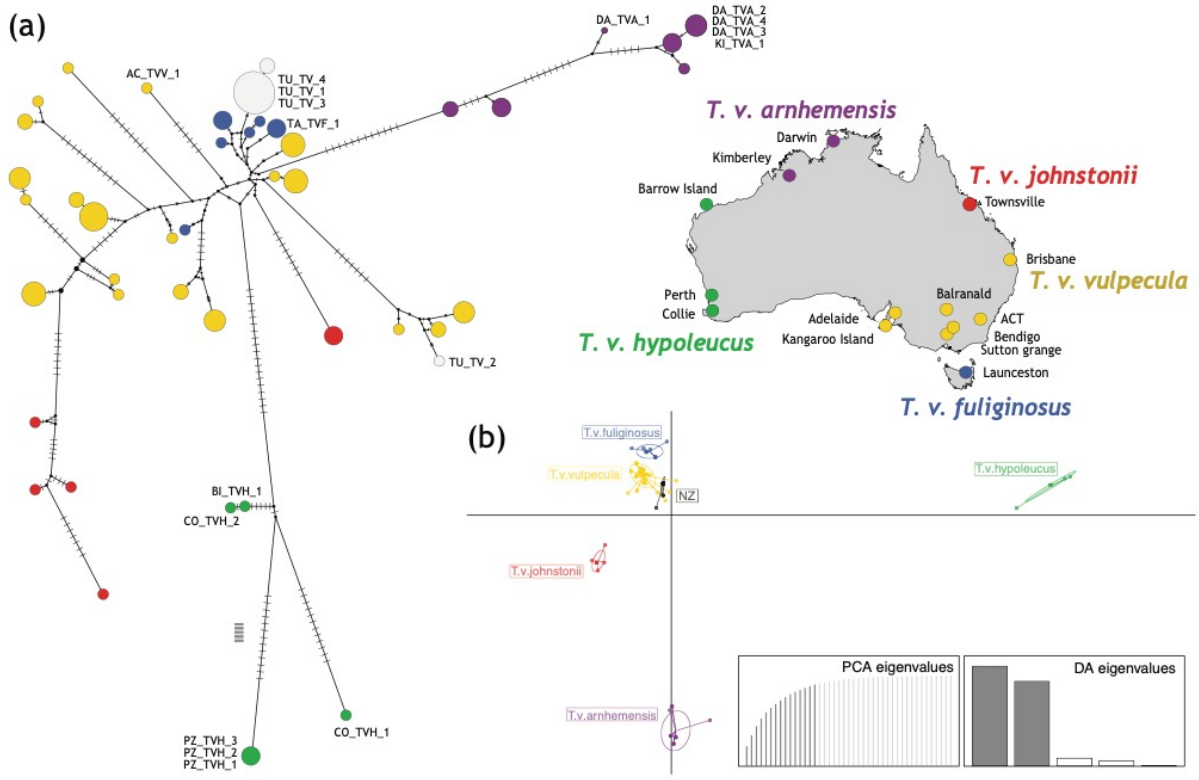


Figure 3: a) Resolved Minimum Spanning haplotype network of mitochondrial D-loop sequences of *Trichosurus vulpecula* subspecies from Australia and New Zealand. Samples used for phylogenetic and molecular clock analysis are indicated with specimen codes. Network nodes are scaled by number of individuals sharing a haplotype and coloured by population origin. Number of steps between haplotypes is represented by dashes on the branches. **b**) Scatterplot of the two main components of the DAPC analysis of D-loop alignment from *Trichosurus vulpecula* in Australia and New Zealand with subspecies as prior grouping. Prior Principal Components Analysis and Discriminant Analysis eigenvalues and selected components are also displayed. Each dot is an individual place with two coordinates on the discriminant analysis two main component. The discriminant analysis uses genetic data to separate groups, distance between groups on the discriminant analysis components corelate with the genetic distance of the groups.

Whole mitochondrial sequences

The mitochondrial genome of *Trichosurus vulpecula* was 14,701 bp, containing the expected 37 genes (13 protein-coding, 2 rRNA and 22 tRNA). We aligned the 17 whole mtDNA sequences with two assembled from the related possum species *Trichosurus caninus* (Figure 1). Our phylogenetic analyses resolved the same strongly supported topology (Appendix C2 Figure 3) with three well supported lineages within our sampling of brushtail possums (Maximum likelihood bootstrap score >99 and Bayesian posterior probabilities >0.99). Samples from southwest of Australia (Perth, Collie) and Barrow Island that represent the subspecies *T. v. hypoleucus* form a lineage sister to the rest of the mtDNA diversity. Possums from Kimberley and Darwin in Northern Australia (*T. v. arnhemensis*) are sister to those from Victoria, Tasmania, and New Zealand (*T. v. fulginosus and T. v. vulpecula*).

Molecular dating

We calibrated a phylogenetic tree using 16 species of Australian marsupial under a Yule model and three calibration points (based on previous estimates derived from fossil-calibrated analyses). The best substitution models for the analysis determined by the beastModel algorithm was TN93, for both calibrated trees. DNA substitution rates from the optimised molecular clock analyses in BEAST yielded site rates of 1.44E-03 (±1.8E-07) for the RNA partition, 1.21E-03 (±1.2E-07) for the partition including the nucleotides 1 and 2 of each codon of the CDSs and 1.26E-02 (±7.9E-06) for the partition including the third nucleotides of each codon of the CDSs. This analysis had a high ESS score, and the node dates were consistent with our calibration except for the Phalangeridae that appear slightly older than previously reported, on our calibrated phylogeny (22.44 Ma (19.42–25.53); Figure 4a). This analysis allowed us to estimate the most recent common ancestor of the *T. vulpecula* and *T. caninus* mitochondrial lineage lived about 5.22 million years ago with 95% highest posterior density

ranging between 4.10 Ma and 6.48 Ma (Figure 4a). We used this estimate of most recent common ancestor (MRCA) for our second calibrated tree with the *T. vulpecula* lineages evolving under a coalescent model, thus allowing us to estimate the timing of the divergence of the intraspecific diversity sampled. DNA substitution rates from the optimised molecular clock analyses in BEAST yielded site rates of 1.6E-03 (±7.28E-8) for the RNA partition, 1.93E-03 (±2.02E-07) for the partition including the first and second codons of the CDSs and 1.45E-02 (±2.2E-06) for the partition including the third codon of the CDSs, those rates were then used with a strict clock for the final run. This analysis produced a high ESS score and showed consistency with the root calibration. *Trichosurus v. hypoleucus* from south-west Australia is sister to all the *T. vulpecula* diversity we sampled. The most recent common ancestor of *T. v. hypoleucus* with the rest of *T. vulpecula* diversity existed around 3.5 Ma (95%HDP 2.5–4.6Ma). More recently the common ancestor between *T. v. arnhemensis* and *T. v. vulpecula* and *T. v. fulginosus* lived around 2.5 Ma (95%HDP 1.77–3.33 Ma). Our estimate of the timing of the most recent common ancestor of the two sub-lineages sampled from *T. v. fulginosus* and *T. v. vulpecula* is dated to 0.55 Ma (95%HDP 0.35–0.78Ma; Figure 4b).

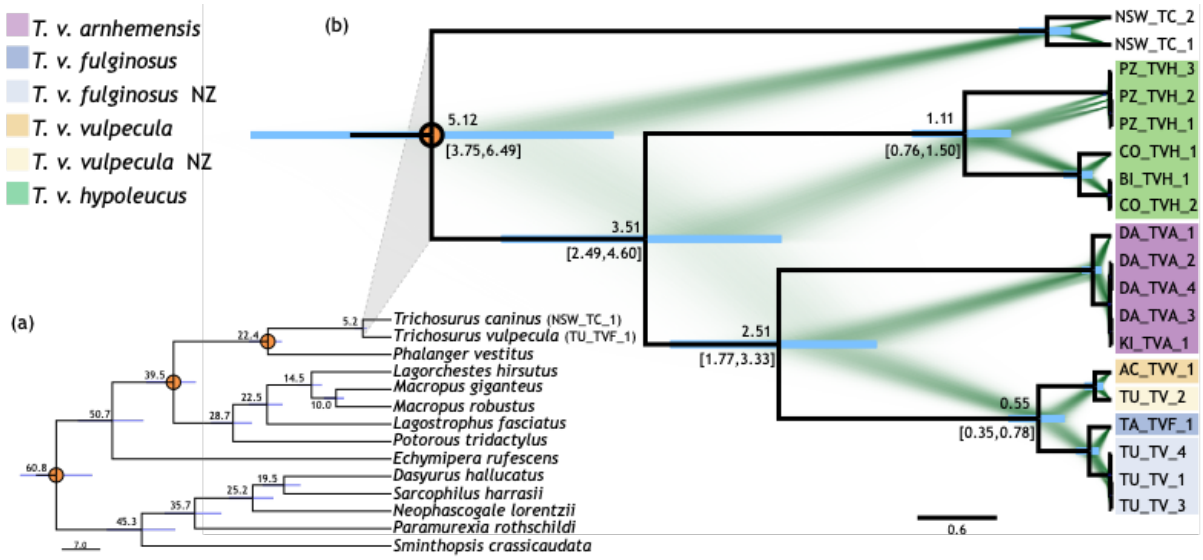


Figure 4: a) Molecular clock calibrated phylogeny to estimate most recent common ancestor of *Trichosurus vulpecula* and the short-eared possum *T. caninus* using whole mitochondrial genomes. Calibration nodes (orange circles) based on dates from fossil-calibrated analyses (Meredith et al., 2009). Maximum clade credibility tree using calibrated Beast analysis under Yule model, nodes corresponding to the divergence of lineages are dated in million years and the bars represents the 95% HDP interval. **b)** Phylogenetic hypothesis for brushtail possum *Trichosurus vulpecula* using whole mitochondrial genomes. Maximum clade credibility tree using calibrated Beast analysis under coalescent model overlapped with a representation of the different trees from the MCMC chain made using DensiTree. The nodes corresponding to the divergence of lineages are dated in million years and the 95% HDP interval are indicated and represented by the bars. Names of the sampled subspecies are indicated, and possum codes are coloured accordingly.

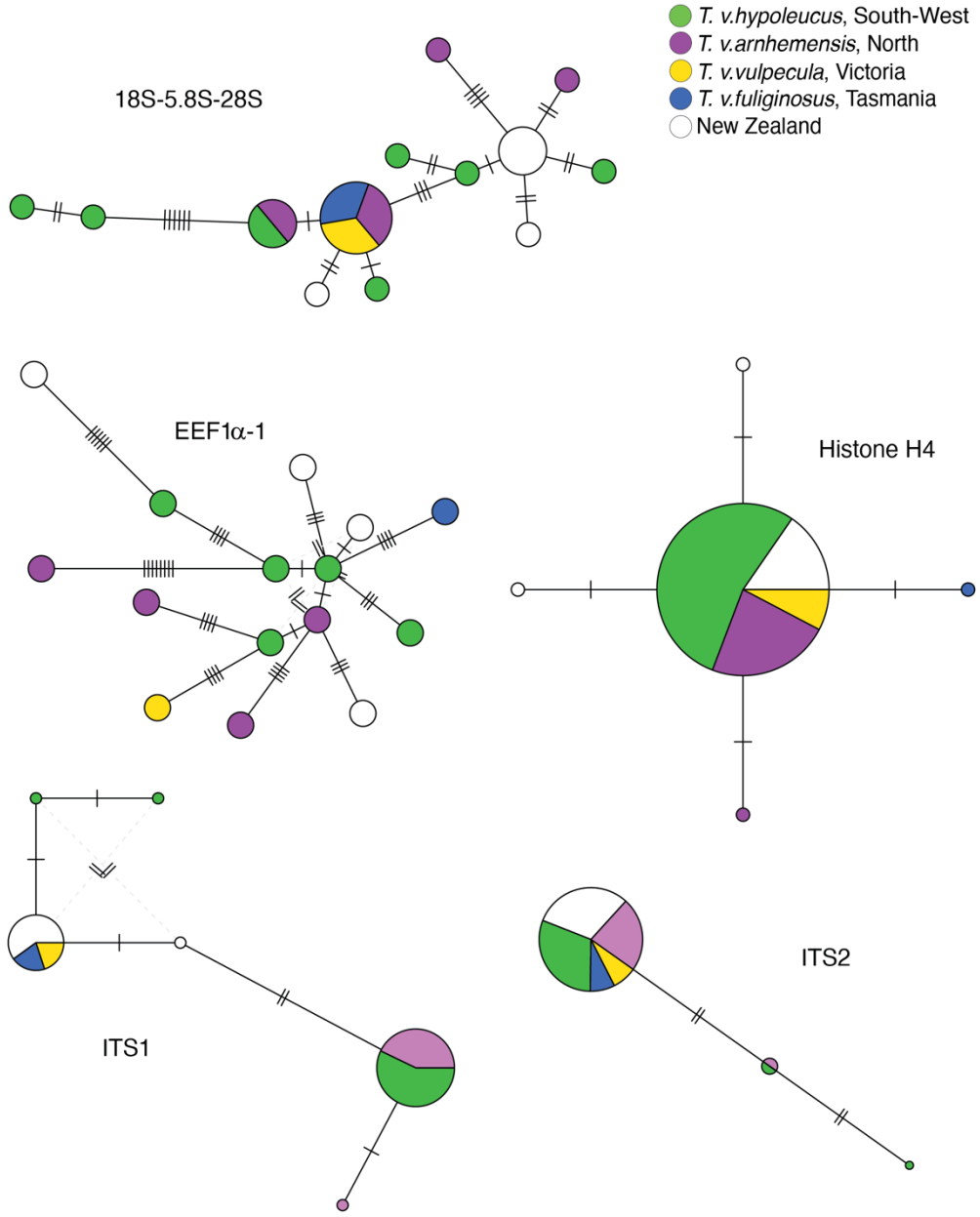


Figure 5: Minimum spanning haplotype network of nuclear loci, nodes are scale by number of individuals sharing the haplotype and coloured by population of origin. Number of steps is represented by dashes on the links.

Nuclear data analysis

Using shotgun DNA sequence data from 17 possum specimens we assembled five markers that exist as multicopy nuclear loci (18S-5.8S-28S, ITS1, ITS2, Histone H4, EEF1α-1, Appendix C2 Table 6). All these nuclear markers contained little sequence variation as revealed by the minimum spanning networks (Figure 5). Individual possums from different locations shared the same DNA sequences for all markers except EEF1α-1. Sequence variability indicate that

these multicopy nuclear loci carry insufficient variability to be phylogenetically informative. During the read alignment process, EEF1α-1 displayed a relatively high proportion of reads with alternative DNA sequences within individuals (Appendix C2 Figure 4 and Table 6) suggesting the existence of pseudogenes obscuring the potential phylogenetic information. Similarly, the 45S rRNA cassette (including ITS1 and ITS2) exhibited some variation within individuals. The short read sequence data from *T. vulpecula* contained multiple apparently paralogous sequences for the 45S rRNA cassette, and within individuals, mutations represented different proportions of their stacked reads (Appendix C2 Table 6), however, concerted evolution and selection have maintained a single 28S, 18S and 5.8S sequence per individual.

Discussion

In sexually reproducing animals the fixation of adaptive traits in a population depends on the strength of natural selection and degree of gene flow (Mallet, 2008; Neher et al., 2010). While population size, range and gene flow can be influenced by the emergence of adaptive traits (Reznick & Ghalambor, 2001; Szücs et al., 2017), climate and other abiotic factors have a strong influence on population structure (Davis & Shaw, 2001; Hewitt, 2004) and thus speciation (Endler, 1977). As a result patterns of biodiversity and population structure are intrinsically linked to changes in Earth's climate (Byrne, 2008; Koot et al., 2022; Nogués-Bravo et al., 2018). Here we tested the origin of intraspecific division in the Australian brushtail possum that experiences very different habitats across its spatial range. The signatures of Pliocene divergence have been observed in a number of widespread Australian animals (Macqueen et al., 2010; P. M. Oliver et al., 2017; Potter et al., 2012; Rowe et al., 2008). Here we infer Pliocene fragmentation has shaped the current brushtail possums population structure, but also suggest that Pleistocene glacial cycles have had an impact on their current distribution (Bryant & Fuller, 2014; Chapple et al., 2005).

Regional biota across the vast landscape of Australia have distinct characteristics such that southwest Australia is a recognised biodiversity hotspot of global significance with endemic plants that account for at least 1.4% of the world's plant species (Myers et al., 2000; Rix et al., 2015), and forests of east Australia are also recognised as globally significant for their distinct biodiversity (Williams et al., 2011). Indeed, seven major ecoregions are recognised across the continent, each having distinct environmental attributes (Mittermeier, 2004), and a high level of endemism among plant and animal assemblages that suggests protracted regional

coevolution (Braithwaite, 1990; Firman et al., 2020; New, 2011; Porder, 2014). Despite the very wide environmental range of the brushtail possum and the likely influence of biotic interactions (from predators and from the destruction of their habitat) limiting their current distribution (Abbott, 2012), our ecological niche models for *Trichosurus vulpecula* had both high sensitivity and specificity. The model suggested that regions of western and central Australian may have been suitable for brushtail possums before European settlements (Abbott, 2012; Kerle, 2002). During the last glacial maximum (LGM), precipitation was lower than today (Faith et al., 2017; Sniderman et al., 2019), but we inferred extension of available niche-space for *T. vulpecula* due to the global reduction in sea level that exposed low elevation land around Australia (Figure 2b). The distribution of D-loop haplotypes confirm that Tasmanian *T. v. fulginosus* is nested within the mtDNA diversity of *T. v. vulpecula* across the Bass Strait in southeast Australia. Bass Strait today has a maximum depth of about 70m, and lowered global sea level of about 120m (Bintanja et al., 2005; Clark et al., 2009) during glacial stages of the Pleistocene would have resulted in a land connection across (Worth et al., 2017). Our inferred niche model confirms that the climate was welcoming for *Trichosurus vulpecula* in Tasmania at the LGM (Figure 2b), and our calibrated phylogenetic analysis estimated a relatively recent common ancestor of lineages sampled from *T. v. fulginosus* and *T. v. vulpecula* at 550,000 years ago. This implies gene flow between Tasmania and southeastern Australian populations during earlier glacial periods (e.g. marine isotope stages 7 and 9) and the isolation of the populations could be explained by several factors, this could imply they diverged sufficiently to not exchange genes or that getting across the Bassian Plain was actually pretty difficult, and the Tasmanian lineage could be better considered 'chance dispersal', rather than a 'vicariant isolation' (Burridge, 2012).

Our projected niche model suggests that the connection of potential possum habitat would have been incomplete during the Pleistocene. In particular, our models suggest little if any suitable habitat in South Australia, even near the south coast. This, coupled with the inhospitable arid environment that prevailed across central Australia throughout the Pleistocene, would have severely limited or prevented gene flow between the western population (*Trichosurus v. hypoleucus*) and those to the east and north of Australia. The inferred distribution of suitable habitat during the LGM and presumably previous glacial phases suggests that the intraspecific diversity of brushtail possums might have originated well before the Pleistocene. Our ecological niche models suggest that the initial fragmentation of *Trichosurus vulpecula* occurred during the Pliocene. Our molecular dating analysis using whole mitochondrial

genome data suggests that the most recent common ancestor of the sampled populations of brushtail possum in Australia existed in the middle of the Pliocene (about 3.5 million years ago; 95% HDP interval 2.4, 4.6 Ma).

On-the-other-hand, nuclear loci failed to resolve the same deep brushtail possum lineages as seen in the mtDNA. Unfortunately, multiple DNA sequences found for ITS1, ITS2 and EEF1α-1 will obscure phylogenetic signal unless the different putative variants can be separated. The existence of multiple copies in the genome of EEF1α-1 and ITSs, and pseudogenes of EEF1α-1 are documented in humans and a number of other organisms (Alanagreh et al., 2017; Coleman, 2013; Cristiano, 2021; Shiao, 2021; Sochorová et al., 2021). It is also possible the lack of phylogenetic signal in these markers is due to gene flow that may have been limited but not absent during the Pleistocene climate oscillations. Limited gene flow would not remove the signal of a common mitochondrial ancestor 2-4 million years ago. A sampling of possums from central Australia would be valuable for ascertaining whether habitat availability allowed gene flow among northern, central, and southern populations during the Pleistocene. The signature of admixture in the mtDNA data from *T. v. johnstonii* suggests unsampled diversity is likely to exist, along with evidence of interbreeding. Fragmentation of populations during the Pliocene provides an explanation of the deep mitochondrial divergence with brushtail possums.

The phylogenetic results of this study would benefit from additional samples, to include all six subspecies and samples from the central and eastern Australian populations. we have inferred phylogenies that are gene trees not subspecies trees, but there is no evidence that data from nuclear markers would change inferences about mtDNA lineage divergence. The concordant results from niche modelling and mtDNA phylogenetic analyses allows us to formulate an evolutionary scenario for brushtail possum in Australia (Figure 6). The divergence between *T. caninus* and *T. vulpecula* corresponds to approximately the end of the Miocene, a period that is marked by a dry and cold climate in Australia (Zachos et al., 2001). At the end of the Miocene it is thought that rainforest persisted only in small patches in southern and south-eastern Australia, with wet and open sclerophyll forest gradually replacing the rainforest (Black et al., 2012; Martin, 1998). This period of fragmented habitat might have been suitable for allopatric divergence of arboreal possums in their respective forest patches. Indeed it appears that a number of modern marsupial groups (e.g. macropodine kangaroos, dasyurids, peramelids)

underwent extensive radiations during the late Miocene, probably in response to this climate change (Beck, 2008; Meredith et al., 2009). Although late Miocene fossil deposits are rare in Australia, *Trichosurus* fossils have been found in the Riversleigh deposit (Archer et al., 2006; Roberts et al., 2008) confirming their presence in eastern Australia at that time.

If the late Miocene can be considered a time of reduced niche-space for Australian possums, then the early Pliocene provided more favourable conditions for brushtail possum population expansion. The short-lived Pliocene reversal of aridification yielded warmer, wetter conditions across most of Australia (Sniderman et al., 2016). Pollen records show that rainforest as well as wet sclerophyll forests expanded in peripheral regions (Black et al., 2012; Martin, 2006; Mcgowran & Bock, 2000) including semiarid southern Australia (Sniderman et al., 2016). Development of these types of habitats at this time plausibly enabled expansion of *T. vulpecula's* range along around the drier heartland of Australia, so that they reached the west from the south and/or the north. Fossils of a related species, *Trichosurus hamiltonensis* (Flannery et al., 1987) found near Hamilton, Victoria indicates that suitable habitat for brushtail possums was available there at that time.

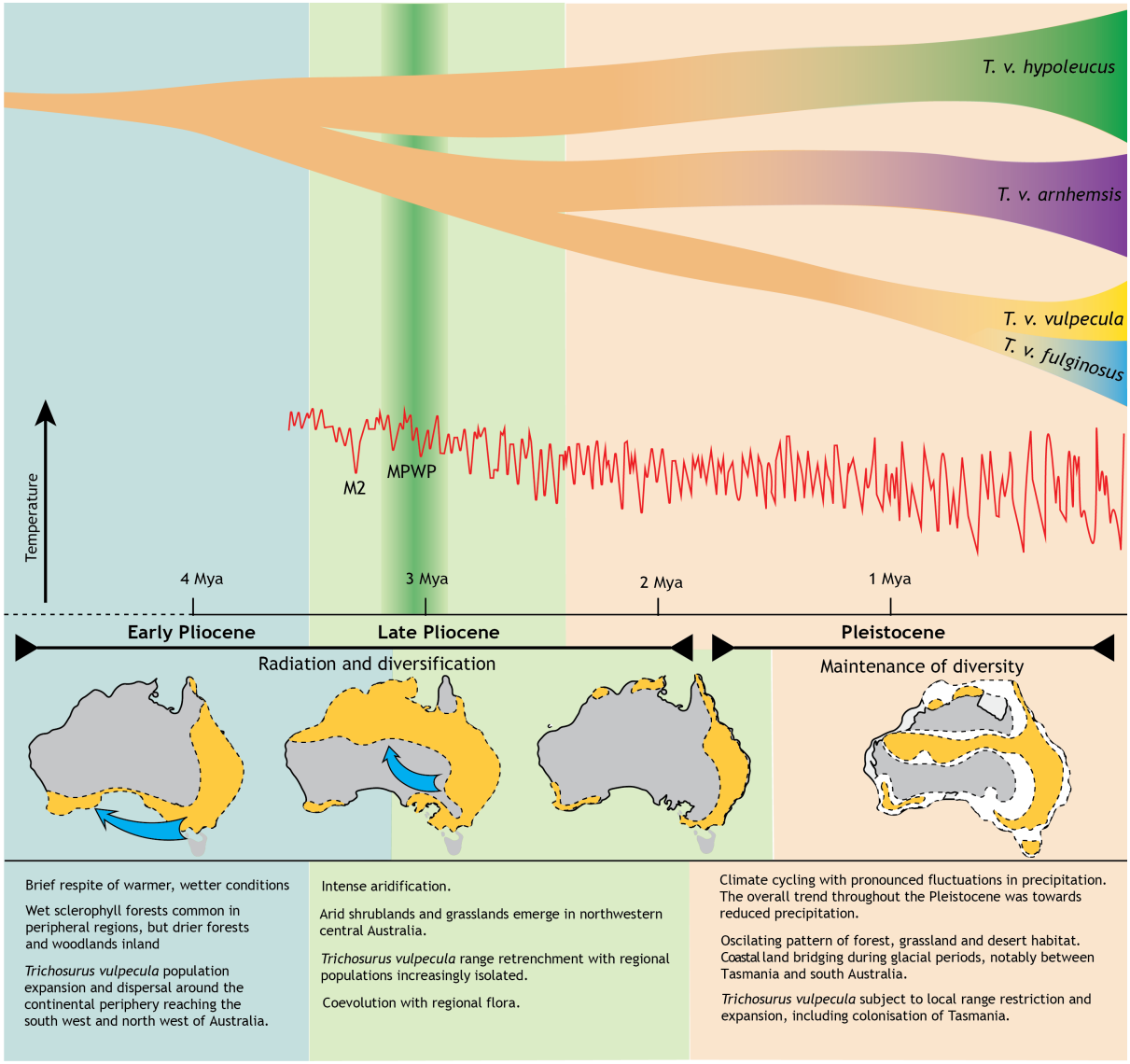


Figure 6: Hypothesis of intraspecific *Trichosurus vulpecula* differentiation in Australia inferred from calibrated phylogeny and niche modelling analyses. The main lineage splitting during the Pliocene corresponds with strong continental aridification, while Pleistocene glacial periods enabled exchange between Tasmania and south Australia due to lowered sea level. Aridification associated with glacial period also restricted forest habitat that may have forced more intense plant-herbivore coevolution.

The Late Pliocene saw increasingly dry conditions but much of Australia probably remained forested. The vegetation is thought to have consisted of sclerophyll open canopy forest but with arid shrublands and grasslands in north-western and central Australia. This gradual change toward drier conditions and the expansion of arid ecosystems would very likely have reduced the available habitat for *T. vulpecula.* As habitat became more fragmented, brushtail possum populations would have become increasingly fragmented and experienced reduced gene flow, perhaps resulting in the current subspecies corresponding to the mtDNA lineage divergence we observe on our genetic data (Figures 3, 6). Dating of most recent common ancestors

observe within other Australian species such as the eastern dwarf tree frog *Litoria fallax* and blue-tongued skink *Tiliqua rugosa* also corresponded to the Pliocene (Ansari et al., 2019; James & Moritz, 2000). By the end of the Pliocene, although wetter on average than today the modern climate of Australia was largely established (Black et al., 2012; Martin, 2006).

Barrow Island population (northwest Australia), which has been referred to as both *T. v. arhenemensis* (langkurr) and *T. v. hypoleucus* (koomal; Dunlop et al., 2021; Lynch et al., 2019) is interesting. Despite our ENM inference of little suitable habitat on the western coast of Australia the molecular data suggest that the Barrow Island possums are part of the southwest mitochondrial lineage (koomal, *T. v. hypoleucus*). We do not know if the shared mtDNA lineage is due to human mediated translocation or natural gene flow between northwestern and southwestern populations before European settlement when their distribution was wider as suggested by historical records (Abbott, 2012). Inferences of a wetter early Pliocene conditions and fluctuating Pleistocene vegetation in this area (He & Wang, 2021; Sniderman et al., 2019), oceanic amelioration of coastal habitat and periods of lowered sea level could plausibly combine to provide a dispersal corridor.

Overall, we find that fragmentation of populations during the Pliocene provides an explanation for the deep mitochondrial divergence within brushtail possums. Despite little phenotypic differentiation, *Trichosurus vulpecula* populations probably experienced little gene flow with one another since the Pliocene, supporting the recognition of several subspecies and explaining their local adaptations to wide range of climate zones they tolerate and the regional plant assemblages on which they feed.

Acknowledgements

Samples of Australian possums were generously provided by the Australian Biological Tissue Collection at the South Australia Museum (Stephen Donnellan) and the Australian National Wildlife Collection CSIRO. We also thanks OSPRI New Zealand and Predator Free New Zealand for the financing.

Data Accessibility Statement

Sequence data used this study are accessible on GenBank (accession numbers in text and supporting information). Script and presence data of niche modelling analysis and intermediary files of the different phylogenetic analysis can be downloaded on https://figshare.com/s/ab4f5426ee7064f1d7f2.

Funding

Financial support for this research came from OSPRI New Zealand Ltd (Richard Curtis) and Predator Free 2050 Ltd (Dan Tompkins).

Competing Interests Statement

The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report. We certify that the submission is original work and is not under review at any other publication.

Author Contributions section: DCR, MMR and SAT conceived the ideas; and SAT collected data with the help from collaborators; DCR and NP analysed the data; DCR, MMR and SAT contributed to the writing of the final manuscript.

Reference

Abbott, I. (2012). Original distribution of Trichosurus vulpecula (Marsupialia : Phalangeridae) in Western Australia, with particular reference to occurrence outside the southwest. *Journal of the Royal Society of Western Australia*, *95*, 83–93. https://www.rswa.org.au/publications/Journal/95(2)/Abbottpp.83-93.pdf

Ansari, M. H., Cooper, S. J. B., Schwarz, M. P., Ebrahimi, M., Dolman, G., Reinberger, L., Saint, K. M., Donnellan, S. C., Bull, C. M., & Gardner, M. G. (2019). Plio-Pleistocene diversification and biogeographic barriers in southern Australia reflected in the phylogeography of a widespread and common lizard species. *Molecular Phylogenetics and Evolution*, *133*, 107–119. https://doi.org/10.1016/j.ympev.2018.12.014

Archer, M., Arena, D. A., Bassarova, M., Beck, R. M. D., Black, K., Boles, W. E., Brewer, P., Cooke, B. N., Crosby, K., Gillespie, A., Godthelp, H., Hand, S. J., Kear, B. P., Louys, J., Morrell, A., Muirhead, J., Roberts, K. K., Scanlon, J. D., Travouillon, K. J., & Wroe, S.

(2006). Current status of species-level representation in faunas from selected fossil localities in the riversleigh world heritage area, northwestern queensland. *Alcheringa*, *30*(4), 1–17. https://doi.org/10.1080/03115510609506851

Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Beck, R. M. D. (2008). A dated phylogeny of marsupials using a molecular supermatrix and multiple fossil constraints. *Journal of Mammalogy*, *89*(1), 175–189. https://doi.org/10.1644/06-MAMM-A-437.1

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., & Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, *69*(2), 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Bintanja, R., Van De Wal, R. S. W., & Oerlemans, J. (2005). Modelled atmospheric temperatures and global sea levels over the past million years. *Nature*, *437*(7055), 125–128. https://doi.org/10.1038/nature03975

Black, K. H., Archer, M., Hand, S. J., & Godthelp, H. (2012). The rise of Australian marsupials: A synopsis of biostratigraphic, phylogenetic, palaeoecologic and palaeobiogeographic understanding. In *Earth and Life: Global Biodiversity, Extinction Intervals and Biogeographic Perturbations Through Time* (pp. 983–1078). Springer Netherlands. https://doi.org/10.1007/978-90-481-3428-1_35

Bouckaert, R. R. (2010). DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics*, *26*(10), 1372–1373. https://doi.org/10.1093/bioinformatics/btq110

Braithwaite, R. W. (1990). Australia's unique biota: implications for ecological processes. *Journal of Biogeography*, *17*(4–5), 347–354. https://doi.org/10.2307/2845364

Brown, J. L., Hill, D. J., Dolan, A. M., Carnaval, A. C., & Haywood, A. M. (2018). Paleoclim, high spatial resolution paleoclimate surfaces for global land areas. *Scientific Data*, *5*. https://doi.org/10.1038/sdata.2018.254

Bryant, L. M., & Fuller, S. J. (2014). Pleistocene climate fluctuations influence phylogeographical patterns in Melomys cervinipes across the mesic forests of eastern Australia. *Journal of Biogeography*, *41*(10), 1923–1935. https://doi.org/10.1111/jbi.12341

Burridge, C. P. (2012). update: Divergence of island biotas when they were not always islands. *Frontiers of Biogeography*, *3*(4). https://doi.org/10.21425/F5FBG12439

Byrne, M. (2008). Evidence for multiple refugia at different time scales during Pleistocene climatic oscillations in southern Australia inferred from phylogeography. *Quaternary Science Reviews*, *27*(27–28), 2576–2585. https://doi.org/10.1016/j.quascirev.2008.08.032

Chapple, D. G., Keogh, J. S., & Hutchinson, M. N. (2005). Substantial genetic substructuring in southeastern and alpine Australia revealed by molecular phylogeography of the Egernia whitii (Lacertilia: Scincidae) species group. *Molecular Ecology*, *14*(5), 1279–1292. https://doi.org/10.1111/J.1365-294X.2005.02463.X

Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., Mitrovica, J. X., Hostetler, S. W., & McCabe, A. M. (2009). The Last Glacial Maximum. *Science*, *325*(5941), 710–714. https://doi.org/10.1126/science.1172873

Cooper, C. E., Withers, P. C., Munns, S. L., Geiser, F., & Buttemer, W. A. (2018). Geographical variation in the standard physiology of brushtail possums ({Trichosurus}): implications for conservation translocations. *Conserv Physiol*, *6*(1). https://doi.org/10.1093/conphys/coy042

Davis, M. B., & Shaw, R. G. (2001). Range shifts and adaptive responses to quaternary climate change. In *Science* (Vol. 292, Issue 5517, pp. 673–679). American Association for the Advancement of Science. https://doi.org/10.1126/science.292.5517.673

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*(8), 1969–1973. https://doi.org/10.1093/molbev/mss075

Dunlop, J., Smith, A., Burbidge, A. H., Thomas, N., Hamilton, N. A., & Morris, K. (2021). Industry environmental offset funding facilitates a large multi-species fauna translocation program. *Pacific Conservation Biology*, *102*, 98–133. https://doi.org/10.1071/PC20036

Endler, J. A. (1977). Geographic variation, speciation, and clines. *Monographs in Population Biology*, *10*, 1–246. https://doi.org/10.2307/2418859

Faith, J. T., Dortch, J., Jones, C., Shulmeister, J., & Travouillon, K. J. (2017). Large mammal species richness and late Quaternary precipitation change in south-western Australia. *Journal of Quaternary Science*, *32*(6), 760–769. https://doi.org/10.1002/jqs.2888

Felsenstein, J. (1985). Phylogenies and the comparative method. *American Naturalist*, *125*(1), 1–15. https://doi.org/10.1086/284325

Felsenstein, Joseph, & Kishino, H. (1993). Is There Something Wrong with the Bootstrap on Phylogenies? A Reply to Hillis and Bull. *Systematic Biology*, *42*(2), 193. https://doi.org/10.2307/2992541

Fick, S., climatology, R. H.-I. journal of, & 2017, undefined. (2017). WorldClim 2: new 1-km

spatial resolution climate surfaces for global land areas. *Wiley Online Library*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Firman, R. C., Rubenstein, D. R., Moran, J. M., Rowe, K. C., & Buzatto, B. A. (2020). Extreme and Variable Climatic Conditions Drive the Evolution of Sociality in Australian Rodents. *Current Biology*, *30*(4), 691-697.e3. https://doi.org/10.1016/j.cub.2019.12.012

Flannery, T. F., Turnbull, W. D., Rich, T. H. V, & Lundelius, E. L. (1987). The phalangerids (Marsupialia: Phalangeridae) of the early Pliocene Hamilton Local Fauna, southwestern Victoria. *Possums and Opossums: Studies in Evolution. Surrey Beatty and Sons and the Royal Zoological Society of New South Wales, Sydney*, 537–546.

He, Y., & Wang, H. (2021). Terrestrial Material Input to the Northwest Shelf of Australia Through the Pliocene-Pleistocene Period and Its Implications on Continental Climates. *Geophysical Research Letters*, *48*(17), e2021GL092745. https://doi.org/10.1029/2021GL092745

Helfrich, P., Rieb, E., Abrami, G., Lücking, A., & Mehler, A. (2019). Treeannotator: Versatile visual annotation of hierarchical text relations. *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, 1958–1963. https://www.aclweb.org/anthology/L18-1308.pdf

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1442), 183–195. https://doi.org/10.1098/rstb.2003.1388

Hijmans, R., Phillips, S., Leathwick, J., Circles, J. E.-, & 2017, U. (2017). Package "dismo." In *gr.xemacs.org*. https://scholar.google.com/ftp://ftp.gr.xemacs.org/mirrors/CRAN/web/packages/dismo/dismo.pdf

Hocknull, S. A., Lewis, R., Arnold, L. J., Pietsch, T., Joannes-Boyau, R., Price, G. J., Moss, P., Wood, R., Dosseto, A., Louys, J., Olley, J., & Lawrence, R. A. (2020). Extinction of eastern Sahul megafauna coincides with sustained environmental deterioration. *Nature Communications*, *11*(1), 1–14. https://doi.org/10.1038/s41467-020-15785-w

Huelsenbeck, J. P., & Hillis, D. M. (1993). Success of Phylogenetic Methods in the Four-Taxon Case. *Systematic Biology*, *42*(3), 247–264. https://doi.org/10.1093/SYSBIO/42.3.247

James, C. H., & Moritz, C. (2000). Intraspecific phylogeography in the sedge frog Litoria fallax (Hylidae) indicates pre-Pleistocene vicariance of an open forest species from eastern Australia. *Molecular Ecology*, *9*(3), 349–358. https://doi.org/10.1046/J.1365-294X.2000.00885.X

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 1–15. https://doi.org/10.1186/1471-2156-11-94

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kerle, A. (1984). Variation in the ecology of Trichosurus: its adaptive significance. *Possums and Gliders*, 115–128.

Kerle, A. (2002). Possums: the brushtails, ringtails and greater glider. *Choice Reviews Online*, *39*(10), 39-5816-39–5816. https://doi.org/10.5860/choice.39-5816

Kerle, A., McKay, G. M., & Sharman, G. B. (1991). A systematic analysis of the brushtail possum, Trichosurus vulpecula (Kerr, 1792) (Marsupialia: Phalangeridae). *Australian Journal of Zoology*, *39*(3), 263–271. https://doi.org/10.1071/ZO9910313

Koot, E. M., Morgan-Richards, M., & Trewick, S. A. (2022). Climate change and alpine-adapted insects: modelling environmental envelopes of a grasshopper radiation. *Royal Society Open Science*, *9*(3). https://doi.org/10.1098/rsos.211596

Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, *7*(1), 48–50. https://doi.org/10.1090/s0002-9939-1956-0078686-7

Kuch, U., Keogh, J. S., Weigel, J., Smith, L. A., & Mebs, D. (2005). Phylogeography of Australia's king brown snake (Pseudechis australis) reveals Pliocene divergence and Pleistocene dispersal of a top predator. *Naturwissenschaften*, *92*(3), 121–127. https://doi.org/10.1007/s00114-004-0602-0

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, *34*(3), 772–773. https://doi.org/10.1093/molbev/msw260

Leigh, J. W., & Bryant, D. (2015). popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Leong, L. E. X., Khan, S., Davis, C. K., Denman, S. E., & McSweeney, C. S. (2017). Fluoroacetate in plants - a review of its distribution, toxicity to livestock and microbial detoxification. In *Journal of Animal Science and Biotechnology* (Vol. 8, Issue 1, p. 55). BioMed Central Ltd. https://doi.org/10.1186/s40104-017-0180-6

Lynch, A. J. J., Beeton, R. J. S., & Greenslade, P. (2019). The conservation significance of the biota of barrow island, western australia. *Journal of the Royal Society of Western Australia*, *102*, 98–133. https://www.researchgate.net/profile/A-Lynch/publication/336855839_The_conservation_significance_of_the_biota_of_Barrow_Island_Western_Australia/links/5dc261f94585151435ec6ff5/The-conservation-significance-of-the-biota-of-Barrow-Island-Western-Australia.p

Macqueen, P., Seddon, J. M., Austin, J. J., Hamilton, S., & Goldizen, A. W. (2010). Phylogenetics of the pademelons (Macropodidae: Thylogale) and historical biogeography of the Australo-Papuan region. *Molecular Phylogenetics and Evolution*, *57*(3), 1134–1148. https://doi.org/10.1016/J.YMPEV.2010.08.010

Mallet, J. (2008). Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 363, Issue 1506, pp. 2971–2986). The Royal SocietyLondon. https://doi.org/10.1098/rstb.2008.0081

Martin, H. A. (2006). Cenozoic climatic change and the development of the arid vegetation in Australia. *Journal of Arid Environments*, *66*(3 SPEC. ISS.), 533–563. https://doi.org/10.1016/j.jaridenv.2006.01.009

Martin, H. A. (1998). Tertiary climatic evolution and the development of aridity in Australia. *Proceedings-Linnean Society Of New South Wales*, *119*, 115–136.

Mcgowran, B., & Bock, P. (2000). *Article in Memoirs of the Association of Australasian Palaeontologists*. https://www.researchgate.net/publication/292144424

Mead, R. J., Oliver, A. J., King, D. R., & Hubach, P. H. (1985). The Co-Evolutionary Role of Fluoroacetate in Plant-Animal Interactions in Australia. *Oikos*, *44*(1), 55. https://doi.org/10.2307/3544043

Meredith, R. W., Westerman, M., Case, J. A., & Springer, M. S. (2008). A phylogeny and timescale for marsupial evolution based on sequences for five nuclear genes. *Journal of Mammalian Evolution*, *15*(1), 1–36.

Meredith, R. W., Westerman, M., & Springer, M. S. (2009). A phylogeny of Diprotodontia (Marsupialia) based on sequences for five nuclear genes. *Molecular Phylogenetics and Evolution*, *51*(3), 554–571. https://doi.org/10.1016/j.ympev.2009.02.009

Miller, G. H., Fogel, M. L., Magee, J. W., Gagan, M. K., Clarke, S. J., & Johnson, B. J. (2005). Anthropology: Ecosystem collapse in pleistocene Australia and a human role in megafaunal extinction. *Science*, *309*(5732), 287–290. https://doi.org/10.1126/science.1111288

Mittermeier, R. A. (2004). *Hotspots revisited*. Cemex.

Myers, N., Mittermeler, R. A., Mittermeler, C. G., Da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. https://doi.org/10.1038/35002501

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography*, *37*(2), 191–203. https://doi.org/10.1111/j.1600-0587.2013.00205.x

Neaves, L. E., Frankham, G. J., Dennison, S., FitzGibbon, S., Flannagan, C., Gillett, A., Hynes, E., Handasyde, K., Helgen, K. M., Tsangaras, K., Greenwood, A. D., Eldridge, M. D. B., & Johnson, R. N. (2016). Phylogeography of the Koala, (Phascolarctos cinereus), and harmonising data to inform conservation. *PLoS ONE*, *11*(9), e0162207. https://doi.org/10.1371/journal.pone.0162207

Neher, R. A., Shraiman, B. I., & Fisher, D. S. (2010). Rate of adaptation in large sexual populations. *Genetics*, *184*(2), 467–481. https://doi.org/10.1534/genetics.109.109009

New, T. R. (2011). 'In Considerable Variety': Introducing the Diversity of Australia's Insects. *'In Considerable Variety': Introducing the Diversity of Australia's Insects*. https://doi.org/10.1007/978-94-007-1780-0

Nogués-Bravo, D., Rodríguez-Sánchez, F., Orsini, L., de Boer, E., Jansson, R., Morlon, H., Fordham, D. A., & Jackson, S. T. (2018). Cracking the Code of Biodiversity Responses to Past Climate Change. In *Trends in Ecology and Evolution* (Vol. 33, Issue 10, pp. 765–776). Trends Ecol Evol. https://doi.org/10.1016/j.tree.2018.07.005

Oliver, A. J., & King, D. R. (1979). Fluoroacetate tolerance, a genetic marker in some australian mammals. *Australian Journal of Zoology*, *27*(3), 331–347. https://doi.org/10.1071/ZO9790363

Oliver, P. M., Adams, M., & Doughty, P. (2010). Molecular evidence for ten species and Oligo-Miocene vicariance within a nominal Australian gecko species (Crenadactylus ocellatus, Diplodactylidae). *BMC Evolutionary Biology*, *10*(1), 1–11. https://doi.org/10.1186/1471-2148-10-386

Oliver, P. M., Laver, R. J., De Mello Martins, F., Pratt, R. C., Hunjan, S., & Moritz, C. C. (2017). A novel hotspot of vertebrate endemism and an evolutionary refugium in tropical

Australia. *Diversity and Distributions*, *23*(1), 53–66. https://doi.org/10.1111/DDI.12506

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*, *26*(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696

Pattabiraman, N., Morgan-Richards, M., Powlesland, R., & Trewick, S. A. (2022). Unrestricted gene flow between two subspecies of translocated brushtail possums (Trichosurus vulpecula) in Aotearoa New Zealand. *Biological Invasions*, *24*(1), 247–260. https://doi.org/10.1007/s10530-021-02635-z

Phillips, M. J., Lin, Y. H., Harrison, G. L., & Penny, D. (2001). Mitochondrial genomes of a bandicoot and a brushtail possum confirm the monophyly of australidelphian marsupials. *Proceedings of the Royal Society B: Biological Sciences*, *268*(1475), 1533–1538. https://doi.org/10.1098/rspb.2001.1677

Porder, S. (2014). Coevolution of life and landscapes. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 111, Issue 9, pp. 3207–3208). https://doi.org/10.1073/pnas.1400954111

Potter, S., Cooper, S. J. B., Metcalfe, C. J., Taggart, D. A., & Eldridge, M. D. B. (2012). Phylogenetic relationships of rock-wallabies, Petrogale (Marsupialia: Macropodidae) and their biogeographic history within Australia. *Molecular Phylogenetics and Evolution*, *62*(2), 640–652. https://doi.org/10.1016/j.ympev.2011.11.005

Pracy, L. (1974). Opposums. *New Zealand Nature Heritage*, *3*(32), 873–882.

Prideaux, G. J., Roberts, R. G., Megirian, D., Westaway, K. E., Hellstrom, J. C., & Olley, J. M. (2007). Mammalian responses to Pleistocene climate change in Southeastern Australia. *Geology*, *35*(1), 33–36. https://doi.org/10.1130/G23070A.1

Rambaut, Suchard, Xie, & Drummond, A. (2015). *Tracer v1.7*. http://tree.bio.ed.ac.uk/software/tracer/

Reid, N. M., & Carstens, B. C. (2012). Phylogenetic estimation error can decrease the accuracy of species delimitation: A Bayesian implementation of the general mixed Yule-coalescent model. *BMC Evolutionary Biology*, *12*(1), 1–11. https://doi.org/10.1186/1471-2148-12-196

Reznick, D. N., & Ghalambor, C. K. (2001). The population ecology of contemporary adaptations: What empirical studies reveal about the conditions that promote adaptive evolution. *Genetica*, *112–113*(1), 183–198. https://doi.org/10.1023/A:1013352109042

Rix, M. G., Edwards, D. L., Byrne, M., Harvey, M. S., Joseph, L., & Roberts, J. D. (2015). Biogeography and speciation of terrestrial fauna in the south-western Australian biodiversity hotspot. *Biological Reviews*, *90*(3), 762–793.

https://doi.org/10.1111/brv.12132

Roberts, K. K., Bassarova, M., & Archer, M. (2008). Oligo-Miocene ringtail possums of the genus Paljara (Pseudocheiridae: Marsupialia) from Queensland, Australia. *Geobios*, *41*(6), 833–844. https://www.sciencedirect.com/science/article/pii/S0016699508000806

Roberts, R. G., Flannery, T. F., Ayliffe, L. K., Yoshida, H., Olley, J. M., Prideaux, G. J., Laslett, G. M., Baynes, A., Smith, M. A., Jones, R., & Smith, B. L. (2001). New ages for the last Australian megafauna: Continent-wide extinction about 46,000 years ago. *Science*, *292*(5523), 1888–1892. https://doi.org/10.1126/science.1060264

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572–1574. https://doi.org/10.1093/bioinformatics/btg180

Rowe, K. C., Reno, M. L., Richmond, D. M., Adkins, R. M., & Steppan, S. J. (2008). Pliocene colonization and adaptive radiations in Australia and New Guinea (Sahul): Multilocus systematics of the old endemic rodents (Muroidea: Murinae). *Molecular Phylogenetics and Evolution*, *47*(1), 84–101. https://doi.org/10.1016/J.YMPEV.2008.01.001

Saltré, F., Rodríguez-Rey, M., Brook, B. W., Johnson, C. N., Turney, C. S. M., Alroy, J., Cooper, A., Beeton, N., Bird, M. I., Fordham, D. A., Gillespie, R., Herrando-Pérez, S., Jacobs, Z., Miller, G. H., Nogués-Bravo, D., Prideaux, G. J., Roberts, R. G., & Bradshaw, C. J. A. (2016). Climate change not to blame for late Quaternary megafauna extinctions in Australia. *Nature Communications*, *7*(1), 1–7. https://doi.org/10.1038/ncomms10511

Sniderman, J. M. K., Hellstrom, J., Woodhead, J. D., Drysdale, R. N., Bajo, P., Archer, M., & Hatcher, L. (2019). Vegetation and Climate Change in Southwestern Australia During the Last Glacial Maximum. *Geophysical Research Letters*, *46*(3), 1709–1720. https://doi.org/10.1029/2018GL080832

Sniderman, J. M. K., Woodhead, J. D., Hellstrom, J., Jordan, G. J., Drysdale, R. N., Tyler, J. J., & Porch, N. (2016). Pliocene reversal of late Neogene aridification. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(8), 1999–2004. https://doi.org/10.1073/pnas.1520188113

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Steel, M., & McKenzie, A. (2001). Properties of phylogenetic trees generated by yule-type speciation models. *Mathematical Biosciences*, *170*(1), 91–112. https://doi.org/10.1016/S0025-5564(00)00061-4

Strasburg, J. L., Kearney, M., Moritz, C., & Templeton, A. R. (2007). Combining phylogeography with distribution modeling: Multiple Pleistocene range expansions in a parthenogenetic gecko from the Australian arid zone. *PLoS ONE*, *2*(8), e760. https://doi.org/10.1371/journal.pone.0000760

Szücs, M., Vahsen, M. L., Melbourne, B. A., Hoover, C., Weiss-Lehman, C., Hufbauer, R. A., & Schoener, T. W. (2017). Rapid adaptive evolution in novel environments acts as an architect of population range expansion. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(51), 13501–13506. https://doi.org/10.1073/pnas.1712934114

Thuiller, W., Georges, D., & Engler, R. (2015). *Package 'biomod2': ensemble platform for species distribution modeling* (R package version 3.3-7).

Thuiller, W., Georges, D., Engler, R., Breiner, F., Georges, M. D., & Thuiller, C. W. (2016). Package 'biomod2.' *Species Distribution Modeling within an Ensemble Forecasting Framework*.

Turney, C. S. M., Flannery, T. F., Roberts, R. G., Reid, C., Fifield, L. K., Higham, T. F. G., Jacobs, Z., Kemp, N., Colhoun, E. A., Kalin, R. M., & Ogle, N. (2008). Late-surviving megafauna in Tasmania, Australia, implicate human involvement in their extinction. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(34), 12150–12153. https://doi.org/10.1073/pnas.0801360105

Twigg, L. E., King, D. R., Bowen, L. H., Wright, G. R., & Eason, C. T. (1996). Fluoroacetate content of some species of the toxic Australian plant genus, gastrolobium, and its environmental persistence. *Natural Toxins*, *4*(3), 122–127. https://doi.org/10.1002/19960403NT4

Umbrello, L. S., Didham, R. K., How, R. A., & Huey, J. A. (2020). Multi-Species Phylogeography of Arid-Zone Sminthopsinae (Marsupialia: Dasyuridae) Reveals Evidence of Refugia and Population Expansion in Response to Quaternary Change. *Genes 2020, Vol. 11, Page 963*, *11*(9), 963. https://doi.org/10.3390/GENES11090963

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2019). blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, *10*(2), 225–232. https://doi.org/10.1111/2041-210X.13107

VanDerWal, J., Falconi, L., Januchowski, S., Shoo, L., Storlie, C., & VanDerWal, M. J. (2014). Package 'SDMTools.' *R Package*.

Vignali, S., Barras, A., … R. A.-E. and, & 2020, undefined. (2020). SDMtune: An R package

to tune and evaluate species distribution models. *Wiley Online Library*, *10*(20), 11488–11506. https://doi.org/10.1002/ece3.6786

Williams, K. J., Ford, A., Rosauer, D. F., De Silva, N., Mittermeier, R., Bruce, C., Larsen, F. W., & Margules, C. (2011). Forests of East Australia: The 35th Biodiversity Hotspot. In *Biodiversity Hotspots* (pp. 295–310). https://doi.org/10.1007/978-3-642-20992-5_16

Worth, J. R. P., Holland, B. R., Beeton, N. J., Schönfeld, B., Rossetto, M., Vaillancourt, R. E., & Jordan, G. J. (2017). Habitat type and dispersal mode underlie the capacity for plant migration across an intermittent seaway. *Annals of Botany*, *120*(4), 539–549. https://doi.org/10.1093/aob/mcx086

Zachos, J., Pagani, H., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. In *Science* (Vol. 292, Issue 5517, pp. 686–693). https://doi.org/10.1126/science.1059412

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, *1*(1), 3–14. https://doi.org/10.1111/j.2041-210x.2009.00001.x

DRC 16

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	David Carmelet-Rescan
Name/title of Primary Supervisor:	Steve A. Trewick
In which chapter is the manuscript /published work:	Chapter 2

Please select one of the following three options:

(●) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

Carmelet-Rescan, D., Morgan-Richards, M., Pattabiraman, N., & Trewick, S. A. (2022). Time-calibrated phylogeny and ecological niche models indicate Pliocene aridification drove intraspecific diversification of brushtail possums in Australia. Ecology and Evolution, 12(12), e9633.

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:

- The percentage of the manuscript/published work that was contributed by the candidate:

- Describe the contribution that the candidate has made to the manuscript/published work:

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	David Carmelet-Rescan Digitally signed by David Carmelet-Rescan Date: 2023.05.21 14:41:29 +02'00'
Date:	21-May-2023
Primary Supervisor's Signature:	Steve Trewick Digitally signed by Steve Trewick Date: 2023.06.05 17:19:28 +12'00'
Date:	6-May-2023

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

GRS Version 5 – 13 December 2019
DRC 19/09/10

Chapter 3: Differential expression analysis and exploration of the molecular pathways to brushtail possums' development.

Abstract

This chapter presents a study of gene expression changes during development in the brushtail possum, a marsupial with high conservation importance in Australia and New Zealand, using RNA sequencing. The analysis focused on identifying differentially expressed genes between adult and juvenile possums, male and female possums, and different fur colour phenotypes. The methodology included preparing strand-specific RNA-seq libraries and using statistical models to test for significant differences in gene expression. The study identified 475 genes with significant changes in transcriptional levels during development, with some genes associated with tissue development, cell cycle, and extracellular matrix. The findings provide insights into the mechanisms associated with possum development and have implications for understanding phenotypic variation and adaptation in non-model organisms. Cytochrome P450 genes were found to be downregulated in juveniles compared to adult possums, potentially indicating their association with the diet of the possums. The study also highlights the usefulness of the new genome of *Trichosurus vulpecula* for RNA-seq differential expression analysis and identifying specific genes with strong expression differences.

Introduction

Introduction to Differential Expression analysis

Differential expression (DE) analysis consists of comparing expression levels of genes among two or more groups of samples. Usually the groups of samples differ in a pre-defined way such as male or female, treated or control, adult or juvenile. Gene expression studies are an important first step to understanding the genetics behind regulatory mechanisms involved in adaptation. Determining which set of genes are differentially expressed is now possible in species without previous genomic resources using RNA sequencing and is becoming more accessible with falling sequencing costs (Todd, Black, and Gemmell 2016). RNA-seq consists of using ultra-high-throughput sequencing techniques on extracted messenger RNA to generate several million short-reads to be mapped on to a reference genome/transcriptome. The number of reads mapped on each gene is then used to infer expression levels within the individual and thus investigate differential gene expression (Lamarre et al. 2018b; Marioni et al. 2008).

Differential expression analysis from RNA-Seq data has emerged as a reliable and cost-effective way to study and investigate the expression pattern of some specific conditions

(Costa-Silva, Domingues, and Lopes 2017; Todd, Black, and Gemmell 2016; Zhang et al. 2014). The basic principle to quantifying gene expression is to count for each gene the number of reads from RNA-Seq output that map onto the genome or transcriptome of the studied species. The number is corrected using the mapping and genomes parameters (number of reads, genome size, length of the gene) and fitted to statistical models (i.e. negative binomial) to test the significance of differences between groups (i.e. tissue types or treatments). The negative binomial distribution is adapted for experimental design with few biological replicates and has the advantage of dealing with the overdispersion problem (Anders and Huber 2010). The advantage over other approaches (e.g. microarrays) of using RNA-Seq data is that it does not need prior knowledge to be performed, making it suitable for non-model organisms.
The recent availability of a complete genome of the brushtail possum (*Trichosurus vulpecula*) facilitates research of DEGs (differentially expressed genes) which is the next important step toward understanding this marsupial. Here I test the efficacy of DE analysis using the newly assembled and annotated reference genome for the brushtail possum, with the aim of developing pipelines necessary to investigate phenotypic variation in brushtail possums that have high conservation importance in Australia and are a major pest species in New Zealand.

Like other marsupials, brushtail possum development consists of a very short gestation (Smith and Keyte 2020) and very long lactation, the early pouch life can also be considered an external gestation. For brushtail possums, the internal gestation lasts around 17 days while lactation lasts around 200 days. RNA-Seq differential expression (DE) analysis in marsupials during their development is the tool of choice to understand the mechanisms associated with this extreme transition (Kozulin et al. 2022; Modepalli et al. 2018).

In this chapter I compare gene expression levels in liver samples taken from adult and juvenile brushtail possums. I hypothesize that juvenile possums are investing resources in their growth and, therefore, expect to see up-regulation of genes involved in cell division and organ and tissue development. I will also investigate differential gene expression in the liver of male and female possums but given the role of this organ, I do not expect to detect evidence of sexual differentiation. Similarly, I will compare gene expression between three different fur colour phenotypes in possums and do not expect to detect significant differential gene expression as fur phenotypes freely interbreed in New Zealand and are not associated with any specific group or population. Finally, I investigate how many individuals are required and how to optimize

analysis parameters for differential gene expression analysis in order to yield a reliable set of genes from small samples.

In the context of our overarching investigation into toxin resistance mechanisms in brushtail possums, this chapter delves into the intricate web of gene expression dynamics within their liver tissue. By analysing developmental, sex-related, and fur colour-dependent expression patterns, I aim to uncover a comprehensive list of genes that influence these phenotypic traits but also will enable the separation of genes potentially associated with toxin resistance in the next chapter from those that may be correlated with other factors, such as developmental processes. Moreover, the exploration of sample size and p-value thresholds sheds light on methodological nuances, ultimately enhancing my ability to discern the multifaceted genetic factors contributing to the possums' adaptability to environmental toxins in further chapter.

Materials and Methods

A set of 23 liver samples came from freshly killed possums trapped in Otago, New Zealand as part of a pest-control programme, according to guidelines issued by the NAEAC, Occasional Paper No 2, 2009 (ISBN 978-0-478-33858-4). The tissues represented 15 adults and 7 juveniles (defined as 26 to 102 days postpartum) and one furred juvenile. The metadata for each sample are given in Table 1. The liver samples were submerged in RNAlater (Invitrogen) and placed on ice until storage at +4 ℃ for 24 hours, followed by long-term storage at -80 ℃.

Total RNA from the samples was extracted using TRIzol Reagent (Invitrogen) kit and treated with DNase to remove DNA using the Turbo DNA-*free* kit (Invitrogen), the procedures were carried out according to the manufacturer's protocol. The concentration of DNase-treated RNA was determined using the Qubit RNA High Sensitivity Kit with the Qubit Fluorometer (Invitrogen).

Poly(A) enriched, strand-specific RNA-Seq libraries were prepared using the NEXTFLEX Rapid Directional RNA-Seq Kit 2.0 and NEXTFLEX Poly(A) Beads (2.0) (Perkin Elmer) according to the protocol with the following modifications: 400 ng of DNase-treated RNA as input, fragmentation time of 12 minutes, 0.48 μM NEXTFLEX RNA-Seq 2.0 Unique Dual Index Barcodes (Perkin Elmer) and 12 cycles of PCR. The Unique Dual Index for each sample given in Table 1. The final libraries were pooled in equimolar amounts and run on 5x lanes of

the HiSeq 2000 (Illumina) at Otago Genomics Facility to generate 100-bp SE (single end) reads.

The sequencing data were uploaded to the Galaxy web platform (Jalili et al. 2021), and the quality of the sequencing reads was assessed using FastQC (Galaxy Version 0.73, Andrews & others, 2010). Reads were then trimmed based on quality using Cutadapt (Galaxy Version 3.7, Martin, 2011) with a custom 3' adapter sequence: AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC, and parameters to remove reads ≤ 20 bp in length and reads with a quality score of ≤ 20. Sequences were then aligned to the brushtailed possums (*Trichosurus vulpecula*) genome (Genebank: mTriVul1.pri - GCA_011100635.1) using HISAT2 (Galaxy Version 2.1.0, Kim et al., 2019), a spliced read aligner, with default parameters and '--rna-strandness R' to specify strand-specific information. Approximately 96-98% of reads aligned (83-92% aligned 1 time, 7-14% aligned >1 time) outputting a .bam file of the mapped reads for each sample.

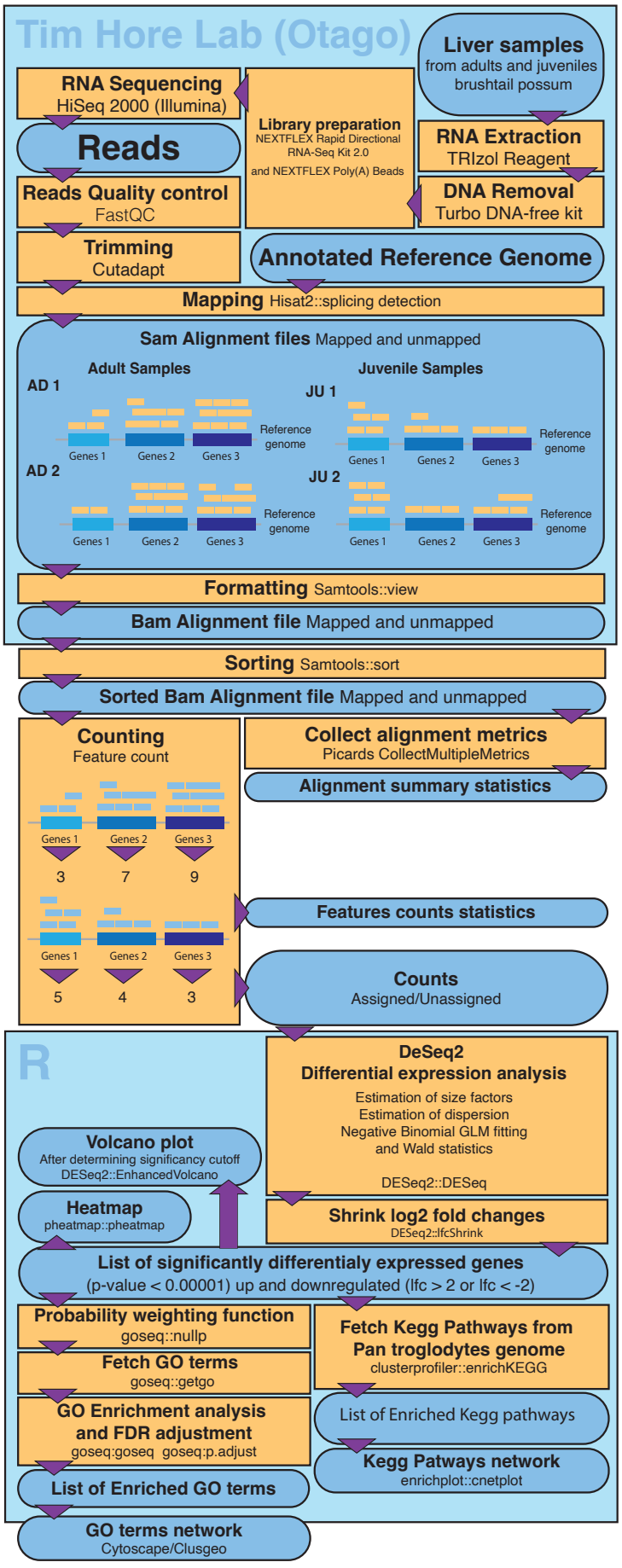
The steps described above were performed by members of the Hore lab at Otago University, New Zealand as part of a project to assemble and report a complete brushtail possum genome *Trichosurus vulpecula* (GenBank: GCA_011100635.1 Bond et al. 2023). Tissue samples were collected by Kyle Richardson, Tim Hore, Mel Laird and Sol Wogan. RNA extractions, DNase treatment and quantifications were performed by Donna Bond and Mel Laird. RNAseq libraries were made by Donna Bond who also processed sequences to generate .bam files.

The number of DNA reads mapped to each gene was reported using the featureCounts software (Liao, Smyth, and Shi 2014), providing a table of the number of mapped reads for each gene that will form the base data for differential expression analysis.

To compute the differential expression analysis the R/Bioconductor package DeSeq2 was used (Love, Huber, and Anders 2014). The DeSeq2 function estimates the size factor and the dispersion before fitting the data to a negative binomial generalized model and computing the Wald statistic for significance testing. Previous studies showed that parametric models are an appropriate approach when replicates are few (Kim et al. 2019) and that DeSeq2 responds to higher read depth by assigning smaller p-values to transcripts with small fold-change (i.e. low differences between counts observed from two groups) (Robles et al. 2012). Significant differentially expressed genes (DEGs) were selected using a threshold of FDR-adjusted p-

values lower than 0.00001 and a log fold change of expression level above 2 or below -2 (corrected using shrinkage estimation) to improve stability and interpretability (Love, Huber, and Anders 2014). Prior analyses were performed using principal component analysis, clustered heatmap of the significant DEGs and volcano plot to visualize those differences using the DeSeq2 R package.

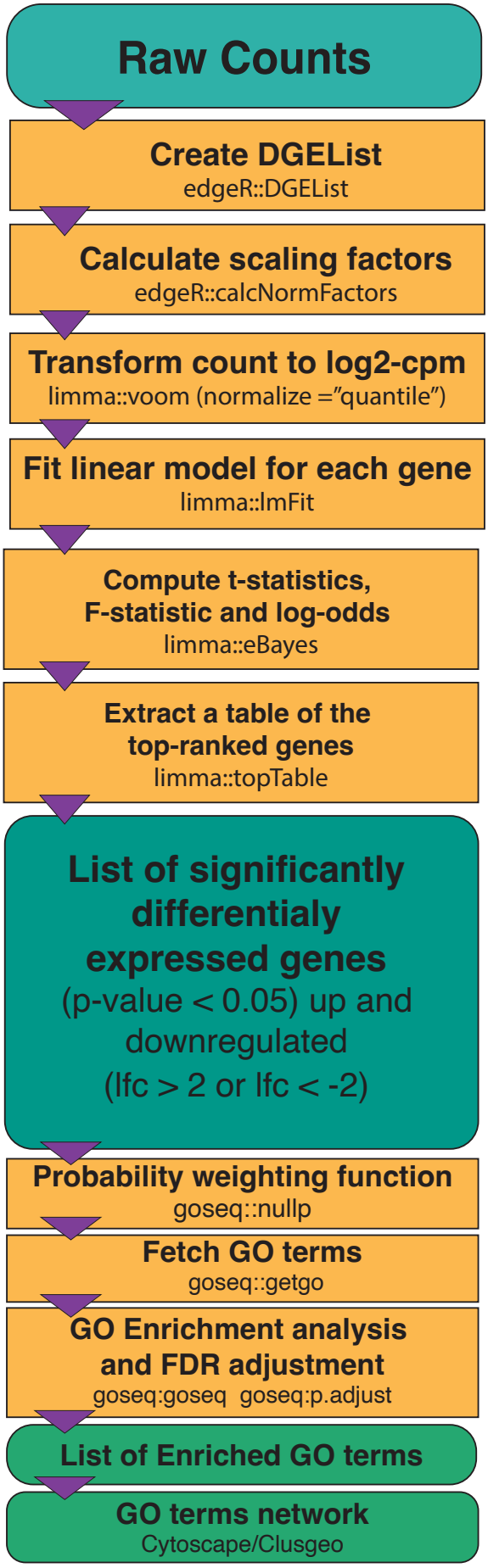
To explore the impact of individual possum developmental stage (age) on the expression of the identified DEGs, each sample was assigned to an age group (adult n=15, juvenile n=8) and the expression level of the significant DEGs was reported among each group. After considering the results of this analysis the sample S-24 was reclassified into the age class; “furred juvenile”. A similar analysis was performed on 14 adult specimens to compare gene expression between males and females and to compare gene expression between three fur colour phenotypes (Table 1).



Since the *Trichosurus vulpecula* genome is not included in the database of gene ontologies and Kegg pathways I used the available genome that returns the most identical genes (> 83% of found genes) of *Pan troglodytes* to select the GO terms and Kegg pathways associated with those genes. GO enrichment analysis was also computed on the significantly differentially expressed genes using goseq and clusterProfiler packages (Young et al. 2013; Yu et al. 2012) to determine the over-represented GO terms including FDR control. Those GO terms and their parenting relations were represented in a node network using the software Cytoscape and the plug-in Clusgeo (Bindea et al. 2009; Smoot et al. 2011).

Figure 1: Flowchart of the complete method to compute differential expression analysis on RNASeq data from liver of brushtail possums (*Trichosurus vulpecula*) using DeSeq2. Including steps realized by Hore laboratory at Otago University.

KEGG enrichment analysis was done using the enrichKEGG function of the "clusterProfiler" R package (Yu et al. 2012). This analysis includes an FDR control. The significantly enriched pathways alongside the associated



under-expressed and overexpressed genes were then plotted using the "enrichplot" R package (Yu 2021). A flowchart detailing the method of the study is represented in Figure 1.

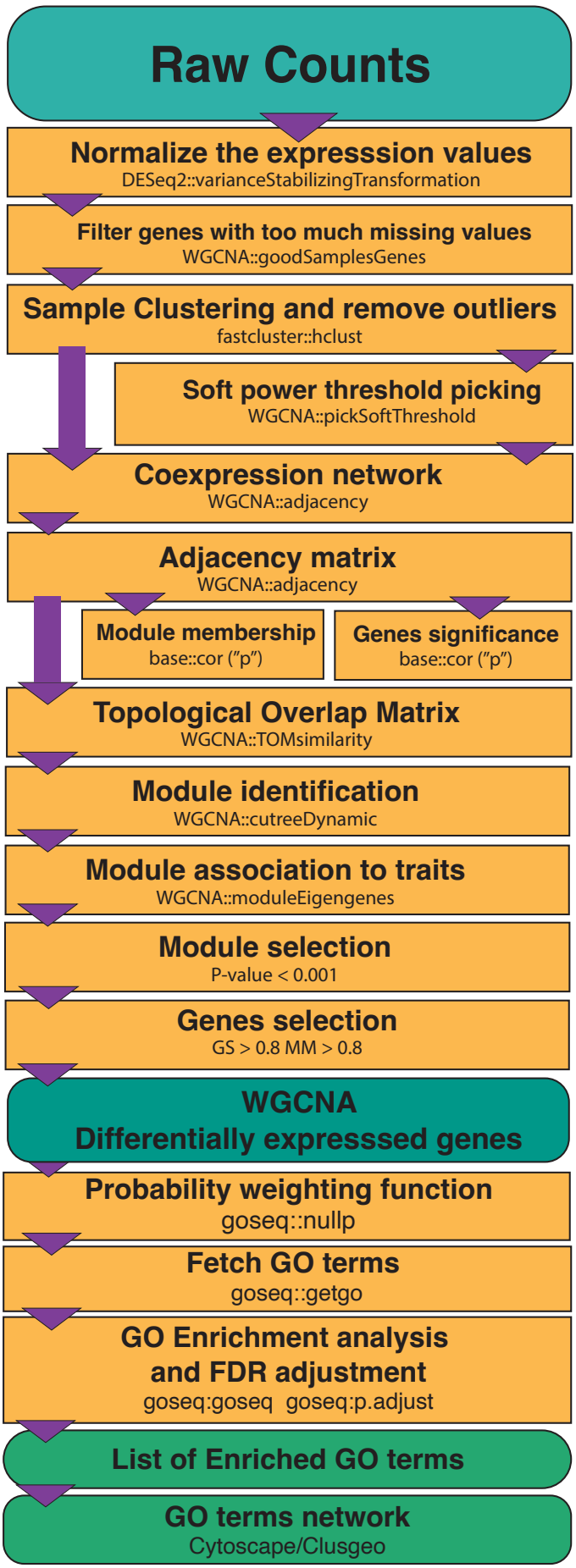
LIMMA

The data were also analysed using a different approach, linear modelling with the Limma R package (Ritchie et al. 2015). The first step is to calculate the scaling factor for each gene from the count data according to the library size and then transform the counts to normalized log2-counts per million. The next step is to fit multiple linear models and then compute associated statistics used to select the significantly differentially expressed genes. The p-value cutoff was set to 0.05, those choices are justified in the discussion. GO enrichment analysis was performed the same on Limma DEGs as on DESeq2 ones. A flowchart detailing the method of the study is represented in Figure 2.

Figure 2: Flowchart of method to compute differential expression analysis on count data from RNASeq sequences from liver of brushtail possums (*Trichosurus vulpecula*) using Limma.

Weighted correlation network analysis (WGCNA)

A search for differentially expressed genes was carried out utilizing weighted correlation network analysis using the WGCNA R package (Langfelder and Horvath 2008). Genes with more than 50% missing data were filtered out reducing the number



of studied genes to 15674 and expression values were normalized using DESeq2 variance stabilisation.

The function "hclust" from the package fastcluster (Müllner 2013) was used on the expression results to cluster samples and exclude samples that deviated based on cluster height value (> 150). The construction of the co-expression network was realized using Pearson's correlation and then the adjacency matrix using the function $a_{mn} = |c_{mn}|^b$ (a_{mn}: adjacency between gene m and gene n, c_{mn}: Pearson's correlation, b: soft-power threshold). The soft-power threshold is picked using the lowest power for which the scale-free topology fit index reaches 0.80. Noise and spurious association effects were minimized by transforming the adjacency matrix into Topological Overlap Matrix and then calculating the associated dissimilarity. Hierarchical clustering on the TOM-based dissimilarity is then performed to produce a hierarchical clustering dendrogram of genes.

Figure 3: Flowchart of method to compute differential expression analysis on count data from RNASeq sequences from liver of brushtail possums (*Trichosurus vulpecula*) using WGCNA.

The dendrogram modules were identified using Dynamic Tree Cut then similar modules were merged based on the co-expression similarity of entire modules. Module association with the traits of interest (maturity and sex) were quantified by performing principal components analysis of each module, the first components of each module are called the module eigengenes (MEs).

Pearson correlation between the traits and the MEs is then calculated with their associated P-value. Modules with a significant p-value ($P< 0.01$) were selected to look for significantly differentially expressed genes. For each gene, Gene Significance (GS) was computed as the correlation between each individual gene of the modules and the trait and module membership (MM) as the correlation between the gene and the module expression profile. Significant genes were then determined using a threshold of $MM > 0.8$ and $GS > 0.8$ in the significant modules. A flowchart of the pipeline of the analysis can be found in figure 3. Similarly, to DESeq2 and Limma, a GO terms enrichment analysis was performed.

Finally, to visualise the difference and common DEGs and GOs between the different packages I used proportional Venn plots using the R package "eulerr" (Larsson and Gustafsson 2018).

How much does sample size limit DEGs discovery and confidence?

To investigate the influence of sample size on the number of genes identified as differentially expressed, the differential expression and the gene ontology enrichment analyses were run again using different subsets of samples and different threshold p-values. A flowchart representing three different algorithms is available in Appendix C3 Figure 1. These analyses were done using the same tools and data as the primary analysis (DESeq2) on R (R Development Core Team 2021).

1) The first approach used between two and eight randomly selected samples in each group (juveniles and adults). For example, three randomly selected juveniles and three randomly selected adults. The selected subset of samples was then analysed in a differential expression and gene ontology enrichment analysis as described above (with similar p-value and fold change thresholds). The significant DEGs and significantly enriched GO terms were then compared to the one issued from the analysis of the whole dataset. The proportion of common significant DEGs and significantly enriched GO

terms are then retained. For each sample subset size, the algorithm was run twenty times (total number of runs = 140).

2) A modification of this analysis used a sample subset size that could take any of all possible values (juveniles 2 to 9, adults 2 to 14), the significantly enriched GO terms are not recorded and the number of iterations for each combination is reduced to ten (total number of runs = 1040). This analysis intend on looking at the impact of the increase sample size of one category.
3) Finally, the influence of the p-value threshold was studied using the same model with subset sizes from two to nine random samples and the threshold of significance for the DEGs varying between four different p-values (p-value < 0.01, 0.001, 0.0001 and 0.00001). Ten iterations per combination were used (total number of runs = 320).

Results

RNA-Seq analysis was performed on adults and juvenile brushtail possums (*Trichosurus vulpecula*) to explore gene expression changes during development. A strict quality control analysis for each sample was conducted to confirm the quality of the short DNA sequences (reads). After strict quality control, more than 13 million reads per sample of clean data were retained, for a total of at least 1.3 Gbps (Table 1).

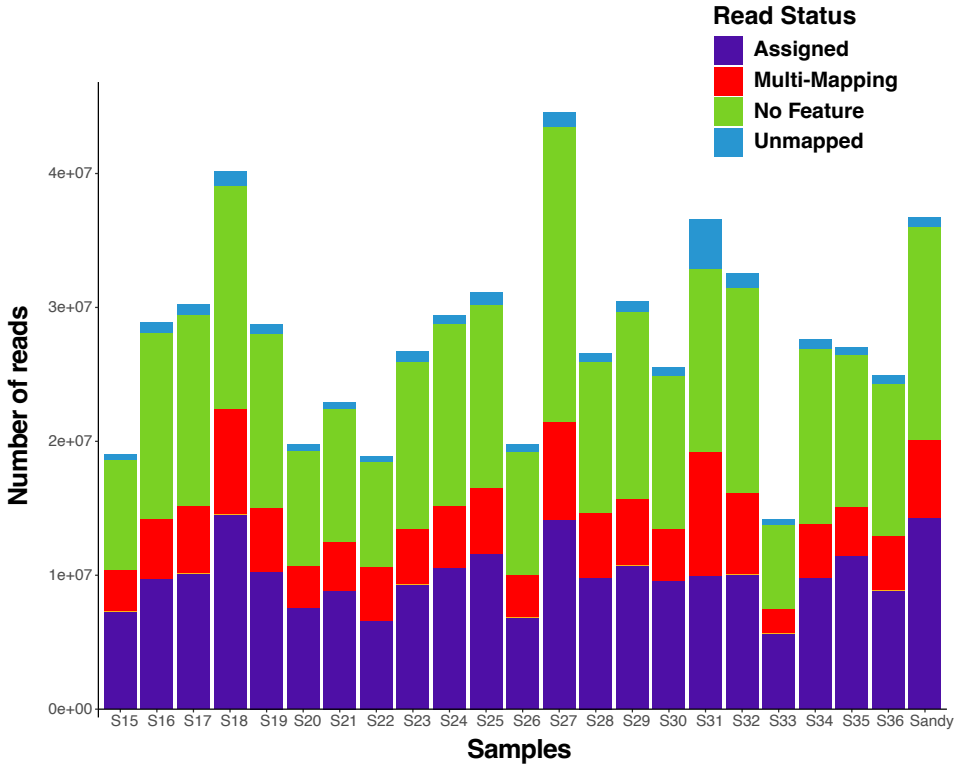


Figure 4: Number of reads and their assignment status for each of 23 brushtail possum (*Trichosurus vulpecula)* RNASeq samples. Read status indicates reads that were mapped: once on an annotated genetic region (Assigned), several times on the genome (Multi-mapping), to regions without any annotated feature (No feature) or unable to map on the genome (Unmapped).

Table 1: Samples of brushtail possums (*Trichosurus vulpecula)* used to study differential expression of genes in the liver. Data output quality and mapping quality for the examined samples, number of reads in million, number of base pairs in Gb, the proportion of aligned reads and proportion of aligned base pairs with at least a Q20 quality. In addition, maturity, fur colour, age (dpp = day-postpartum) and sex are also displayed.

Sample	Reads (M)	Base pairs (Gb)	Mapped reads (%)	Q20 bases (%)	Maturity	Colour	Age	Sex
S15	16.97	1.71	97.44	92.35	Young	-	45 dpp	Male
S16	26.08	2.63	96.99	92.43	Adult	Grey	Adult	Male
S17	26.98	2.72	97.04	92.00	Adult	Black	Adult	Male
S18	34.98	3.53	97.00	90.91	Adult	Red	Adult	Female
S19	25.61	2.58	97.37	92.00	Adult	Red	Adult	Male
S20	17.68	1.78	97.43	92.41	Adult	Grey	Adult	Female
S21	20.49	2.07	97.46	92.65	Young	-	Furred juvenile	Male
S22	16.07	1.62	97.49	90.77	Young	-	69 dpp	Male
S23	24.10	2.43	96.80	92.38	Adult	Grey	Adult	Female
S24	26.32	2.66	97.50	92.63	Adult	Grey	Adult	Male
S25	27.89	2.81	96.72	92.50	Adult	Grey	Adult	Female
S26	17.65	1.78	97.02	92.30	Adult	Black	Adult	Male
S27	39.96	4.03	97.26	91.89	Adult	Grey	Adult	Male
S28	23.22	2.34	97.40	91.88	Young	-	99 dpp	Female
S29	27.26	2.75	97.24	92.12	Adult	Red	Adult	Female
S30	22.94	2.31	97.25	92.94	Young	-	74 dpp	Female
S31	29.98	2.77	87.79	88.76	Young	-	102 dpp	Male
S32	28.54	2.88	96.33	91.33	Adult	Grey	Adult	Female
S33	13.02	1.31	96.84	91.51	Adult	Grey	Adult	Male
S34	25.09	2.53	97.30	92.74	Adult	Red	Adult	Male
S35	24.76	2.50	97.59	93.32	Adult	Grey	Adult	Female
S36	22.22	2.24	97.06	92.08	Young	-	84 dpp	Female
Sandy	33.00	3.33	97.91	92.30	Young	-	26 dpp	Male

Mapping

The mapping process aligned at least 87% of the reads and at least 88% of the bases from those reads are higher quality than Q20 (call accuracy of 99%; Table 1). A total of 15,662 genes were detected across all our samples (pouch young and adults) from the assembled genome of New Zealand *Trichosurus vulpecula* (Genebank: mTriVul1.pri - GCA_011100635.1) which used 92% of the total number of genes in the reference genome. Feature count provides a summary of the alignments allowing further assessment of the extraction and alignment quality. The first is the number of reads that were assigned to genes, which ranges between 27% and 42%. Secondly, the amount of mapped but unassigned reads ranges between 37% and 49% of total reads (Figure 4).

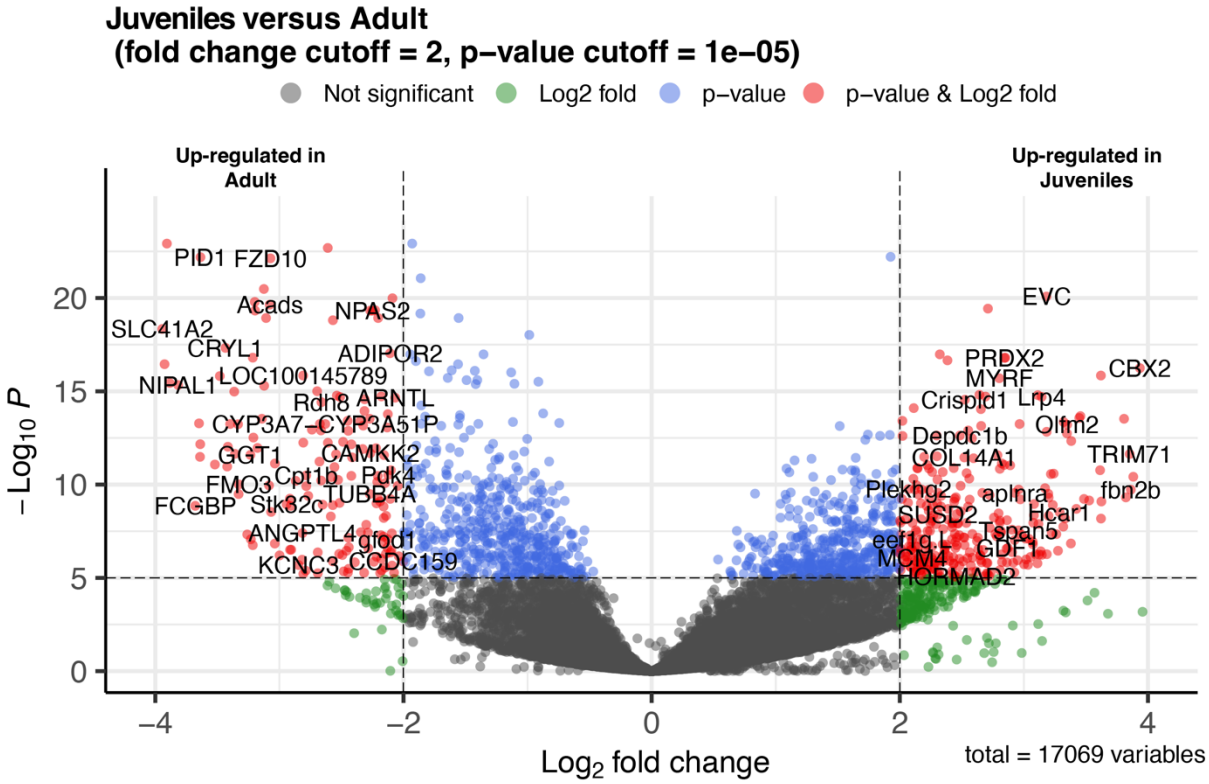


Figure 5: Visualization of differentially expressed genes (DEGs) from brushtail possums (Trichosurus vulpecula) liver samples using volcano plots in R package "DESeq2". The plot compared the DEGs between adults and juveniles. The representations are as follows: x-axis, log2 fold change; y-axis, -log10 of a p-value. The p-values < 0.00001 are in blue dots, and logFC ≥ 2 and logFC ≤ –2 are in green dots; the significant DEGs satisfying both values are in red dots and indicated with gene names. Black dots indicate the remaining genes present in the array that were not significantly changed. The genes that are upregulated in the juveniles are on the right panel, and downregulated genes are on the left panel of the plot. Some DEGs genes display values outside of the intervals plotted above, refer to the genes table (Appendix C3 Table 1).

Differential expression analysis of juveniles

Importantly after differential expression analysis using DESeq2 package, normalisation, shrinkage and FDR analysis, 475 genes were classified to carry significant changes at transcriptional levels, of these genes (more than two times fold difference with an FDR adjusted p-value less than 0.00001) during the development in *Trichosurus vulpecula* (Appendix C3 Table 1). Among those genes, 302 were up-regulated and 173 were down-regulated in juveniles compared to adults (Figure 5).

Principal component analysis and hierarchical clustering (Figure 6-7) on differentially expressed genes clusters the samples into two groups, only adult samples and the eight juveniles plus one individual labelled as an adult (sample S-24).

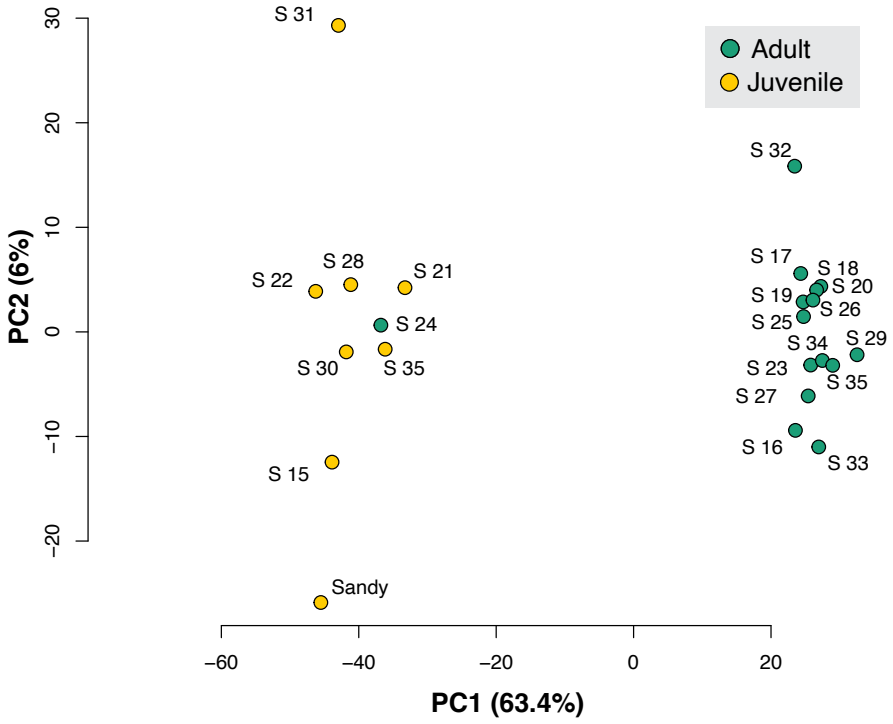


Figure 6: Plot of the first two principal components of the PCA applied to the normalized count data of the RNASeq data from the liver of brushtail possums. Samples are coloured by maturity state and the eigenvalue proportion of each axis is indicated below them.

Close examination of the heatmap depicting expression of each sample (Figure 7), it is apparent that some samples within a group displayed slightly different patterns of expression. Samples S31 displayed some strongly up-regulated and down-regulated genes compared to the other juvenile samples. In contrast samples S21, S24 and S32 include less marked differential expression for the same set of genes. The complete list of differentially expressed genes found with DESeq2 can be found in Appendices (Appendix C3 Table 1).

Many genes in the liver of possums were differentially expressed between juveniles and adults (Figures 8 and 9). Some gene expression levels appear to diminish with development stage (age) (ex: ASP and WISP1), others increase with age (ex: cyb5r2 and TPM2), but more complex patterns can also be seen with the expression level of CYM increasing with days postpartum but dropping among adult samples and ACADSB showing the opposite pattern.

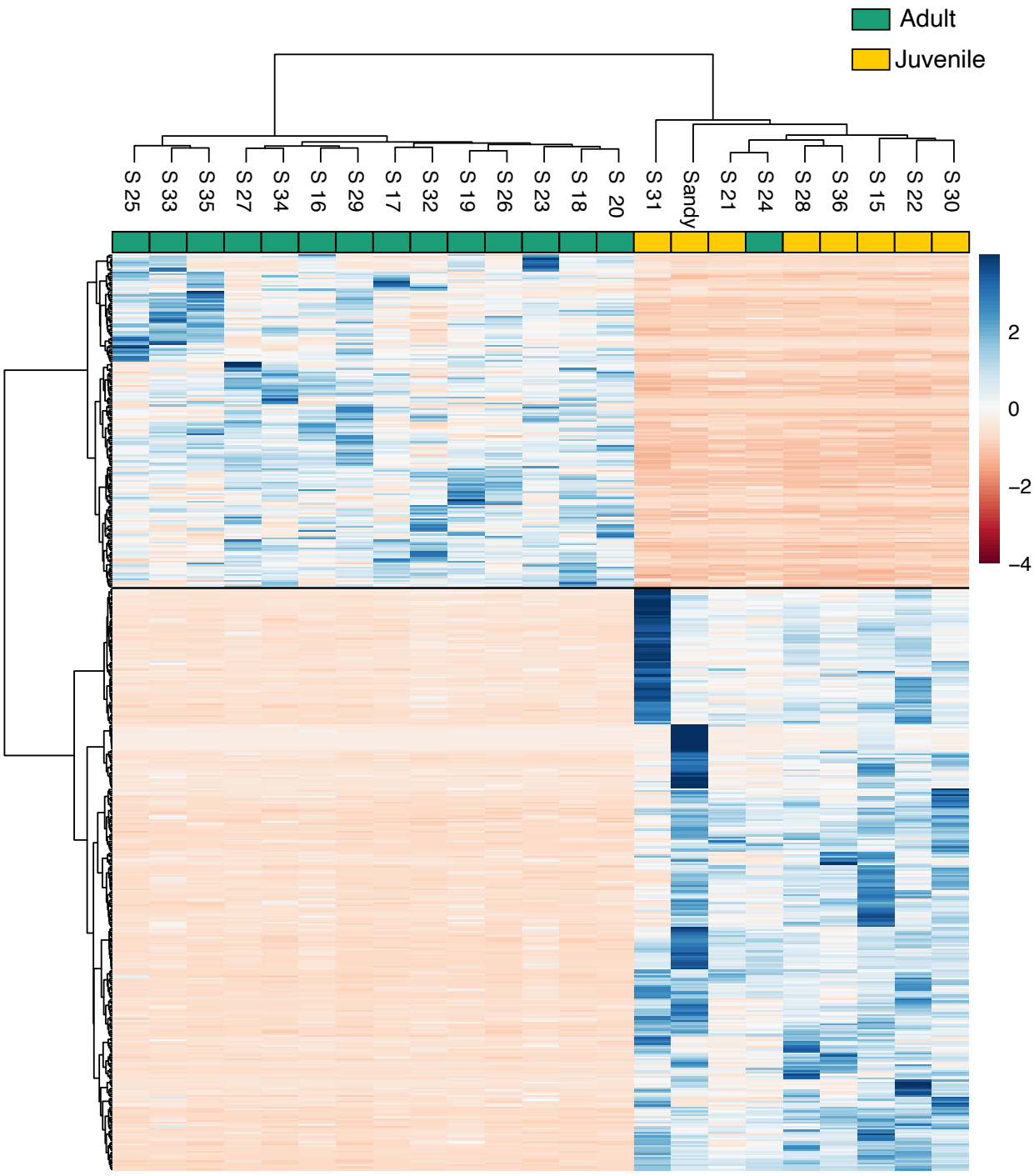


Figure 7: Heatmap and clustering based on a hierarchy of the differentially expressed genes (p.value < 0.00001) from RNASeq data from liver samples from brushtail possums. n = 1402 DEGs. Upregulated genes in juveniles are represented in blue, and downregulated genes are represented in red. Samples maturity is indicated with coloured squares (Adults are in green and Juveniles are yellow).

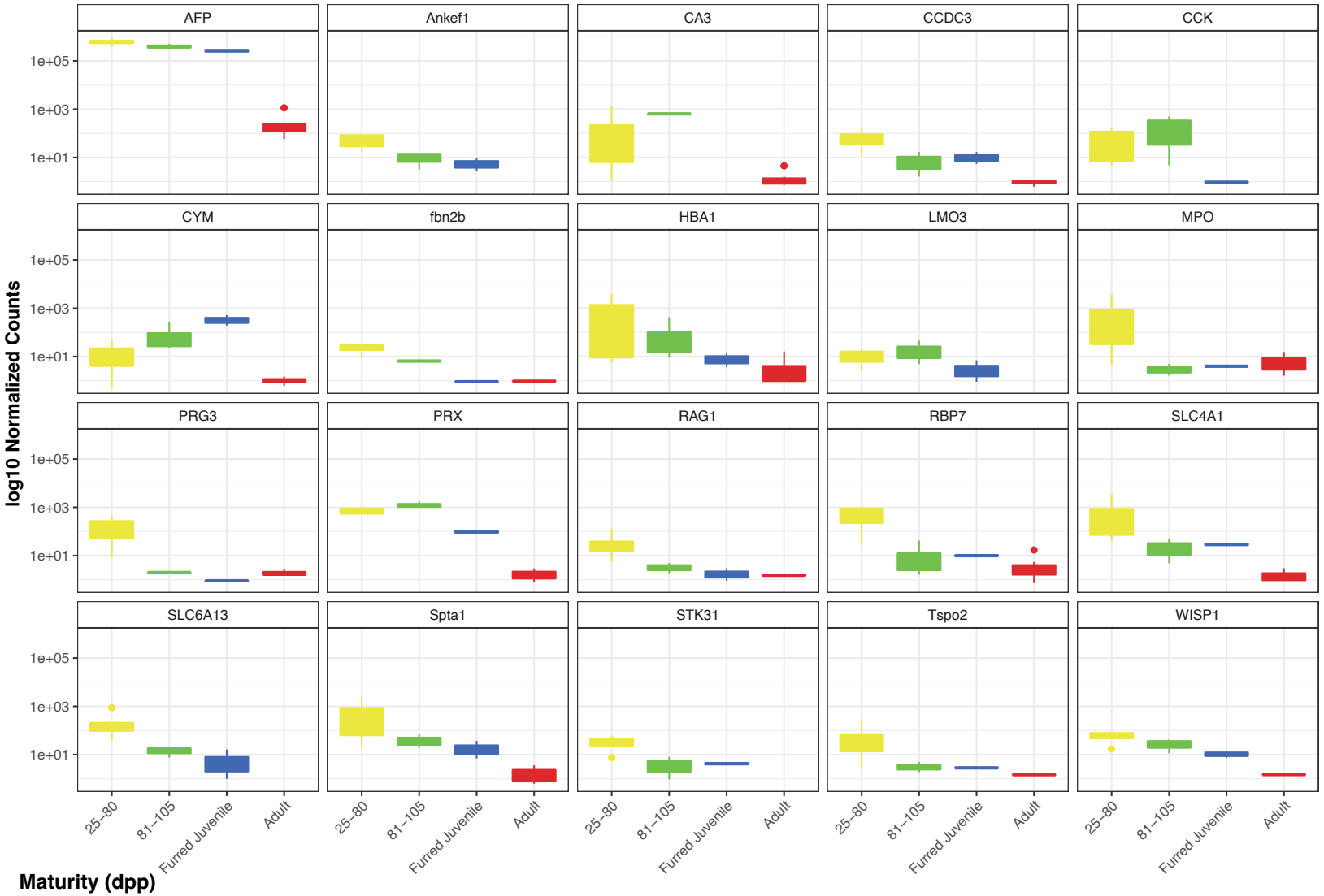


Figure 8: Boxplot of the normalized count for the first twenty up-regulated DEGs (differentially expressed genes) between juveniles and adult brushtail possums (*Trichosurus vulpecula*) liver RNASeq samples throughout four different maturity groups (numbers expressed in days postpartum).

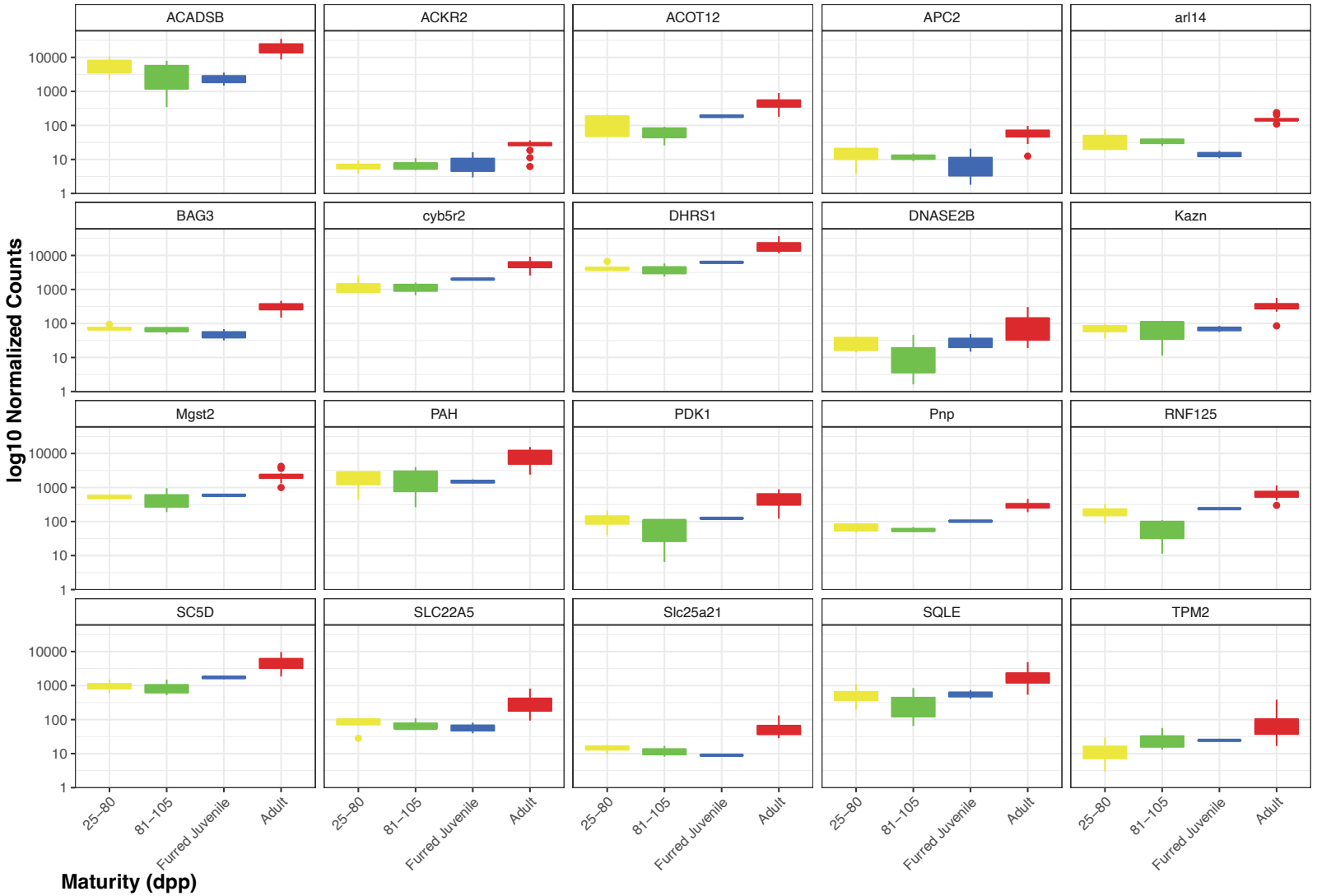


Figure 9: Boxplot of the normalized count for the first twenty down-regulated DEGs (differentially expressed genes) between juveniles and adult brushtail possums (*Trichosurus vulpecula*) liver RNASeq samples throughout four different maturity groups (numbers expressed in days postpartum).

Looking at specific genes in the list of DEGs we can point at the genes from the cytochrome P450s family that shows six genes significantly downregulated with strong log fold change, those genes are associated with the metabolism of endogen molecules (Figure 10).

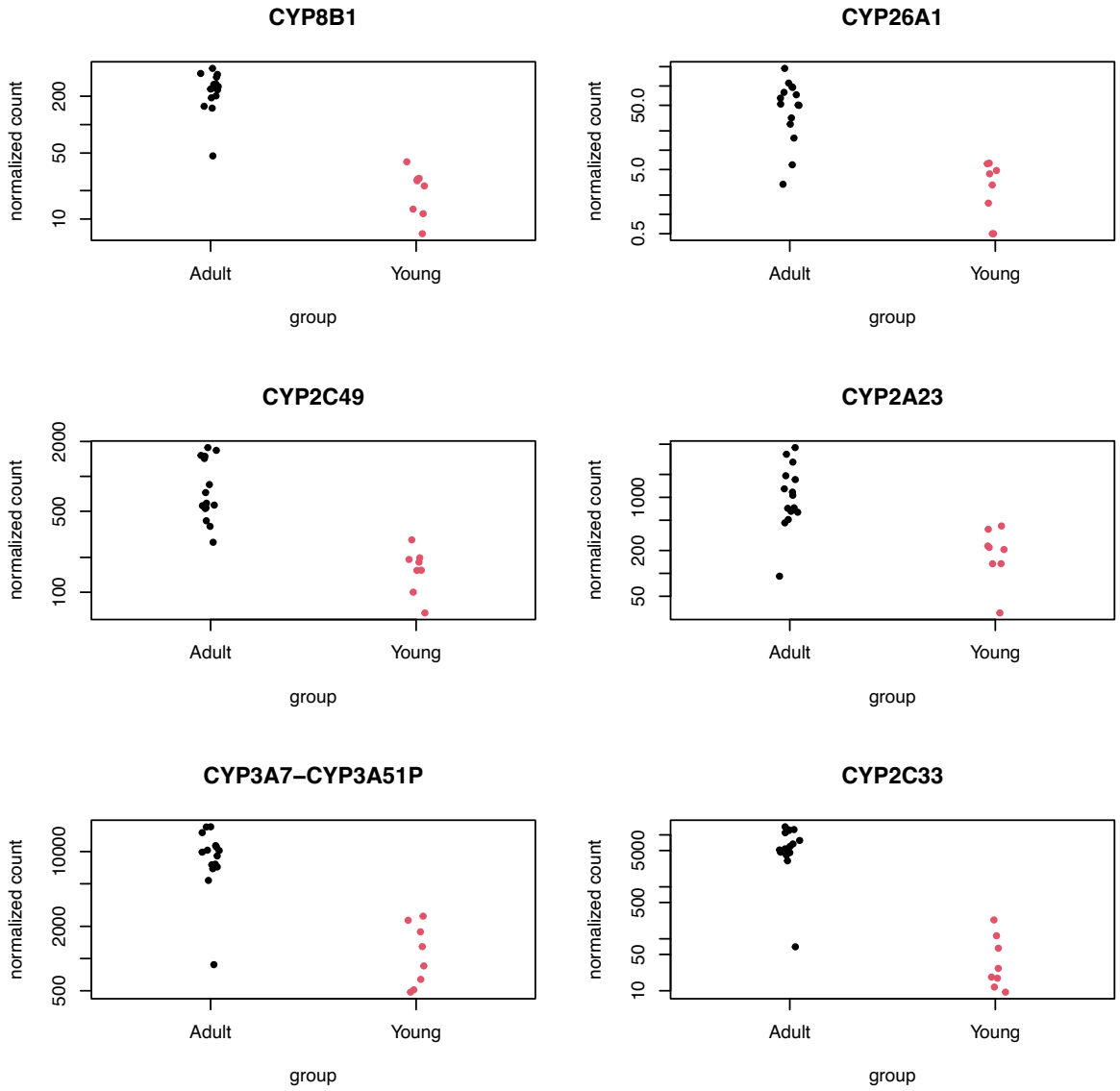


Figure 10: Normalized counts of the 6 significantly differently expressed Cytochrome P450 genes between juveniles and adults brushtail possums (*Trichosurus vulpecula*) liver RNASeq samples. Adults are coloured in black and juveniles in red.

Gene ontology and Kegg pathway of juveniles

To visualize the enriched GO terms for the differentially expressed genes the software Cytoscape and the plug-in ClueGo were used (Figure 11). The 302 significantly up-regulated genes in juvenile brushtail possums were enriched for biological processes with 31 genes

directly related to “mitotic cell cycle process” (GO:1903047) and 16 genes linked to “sister chromatid segregation” (GO:0000819). Other functional groups appear in the enriched GOs among up-regulated genes such as tissue development and organisation with “extracellular matrix organization” (GO:003198) and “collagen fibril organisation” (GO:0030199) with 8 genes and “anatomical structure development” (GO:0048856) with 109 genes.

The 173 down-regulated genes are enriched in GOs associated with catabolic and metabolic processes, for example, “carboxylic acid catabolic process” (GO:0046395) with 9 genes, “cellular lipid catabolic process (GO:0044242) with 11 genes and “organic hydroxy compound metabolic process” (GO:1901615) with 15 genes (Figure 11). The complete list of enriched GO terms from DESeq2 DEGs can be found in Appendices (Appendix C3 Table 2 and 3).

KEGG pathway enrichment analysis adds information on the metabolic background of the development in brushtail possums, overexpressed genes are associated with the cell cycle (12 genes), protein digestion and absorption (12 genes) and ECM-receptor interaction (7 genes) (Figure 12). On another hand, under-expressed genes do not seem to carry any significantly expressed pathways.

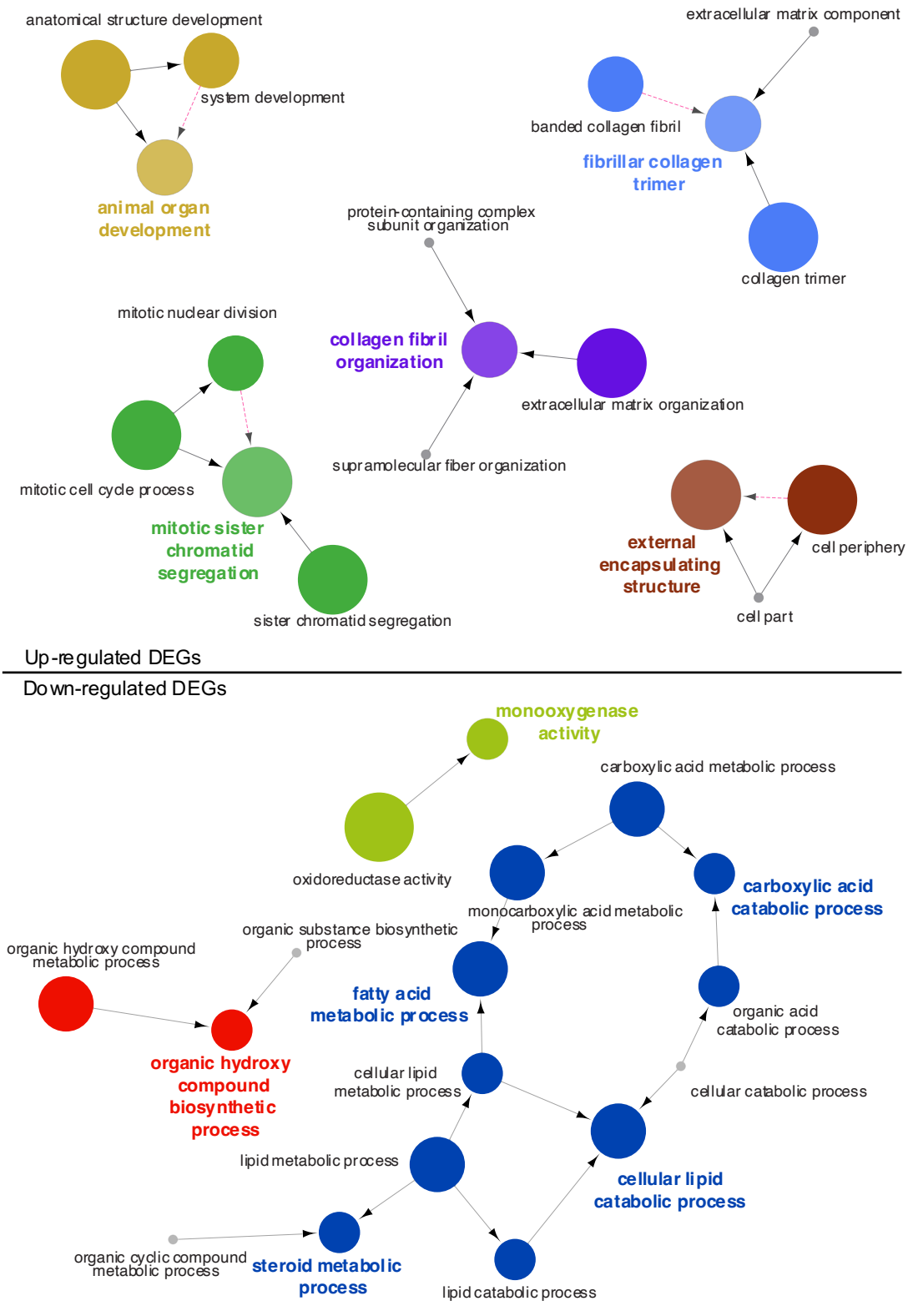


Figure 11: Enriched molecular functions, cellular component and biological processes associated with the DEGs (up-regulated and down-regulated) between juvenile and adult brushtail possums (*Trichosurus vulpecula*) RNASeq liver samples determined using DESeq2 R-package. Related GO terms are connected. Coloured nodes indicate the significantly enriched GO terms (p-value ≤ 0.05), node size is constrained by p-value and node colour code indicates the functional class that they are involved. Small grey nodes are parent GO terms not significantly enriched.

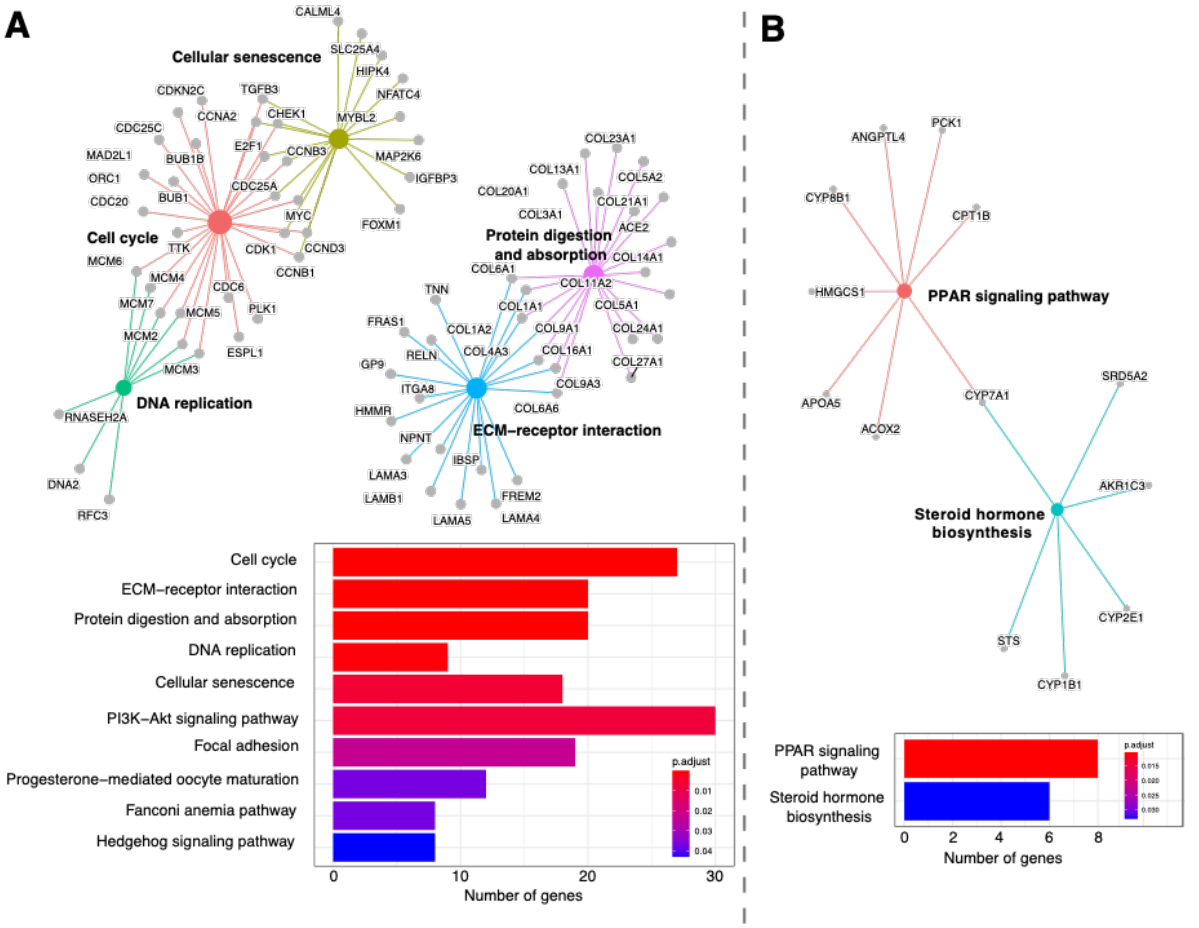


Figure 12: Enrichment by pathway terms is visualized using the cnetplot function from the "enrichplot" R package. Significantly enriched Kegg pathways (p-value ≤ 0.05) associated with the significant DEGs between juvenile and adult' brushtail possums (*Trichosurus vulpecula*) RNASeq liver samples, (A: upregulated, B: downregulated). Big and coloured nodes represent the pathways and grey nodes are the differentially expressed genes associated with those pathways. Nodes are sized using enrichment analysis p-values. The number of genes and FDR-adjusted p-values associated with each significantly enriched pathway are reported on the bar plot.

An alternative approach implemented with the R package Limma (Ritchie et al. 2015) was used to assess the stability of the findings from DESeq2. Limma uses a different method of normalisation that could lead to slightly different results. With a p-value cutoff of 0.05 and log-fold change cutoff of 2 the differential expression analysis using the Limma R package identified 929 genes, 665 upregulated and 264 downregulated. The software Cytoscape and the plug-in ClueGo helped visualize the enriched GO terms in the differentially expressed genes (Figure 13). Similar to what was found with DESeq2, the 665 significantly up-regulated genes in juvenile brushtail possums found using Limma are enriched gene ontologies associated with biological processes directly related to "regulation of cell cycles process" (GO:0010564) with 30 genes and "mitotic sister chromatid segregation" (GO:0000819) with 19 genes. Other functional groups appear in the enriched GOs among up-regulated genes such as tissue

development and organisation with “collagen fibril organization” (GO:003198) with 9 genes and “tissue development” (GO:0009888) with 72 genes.

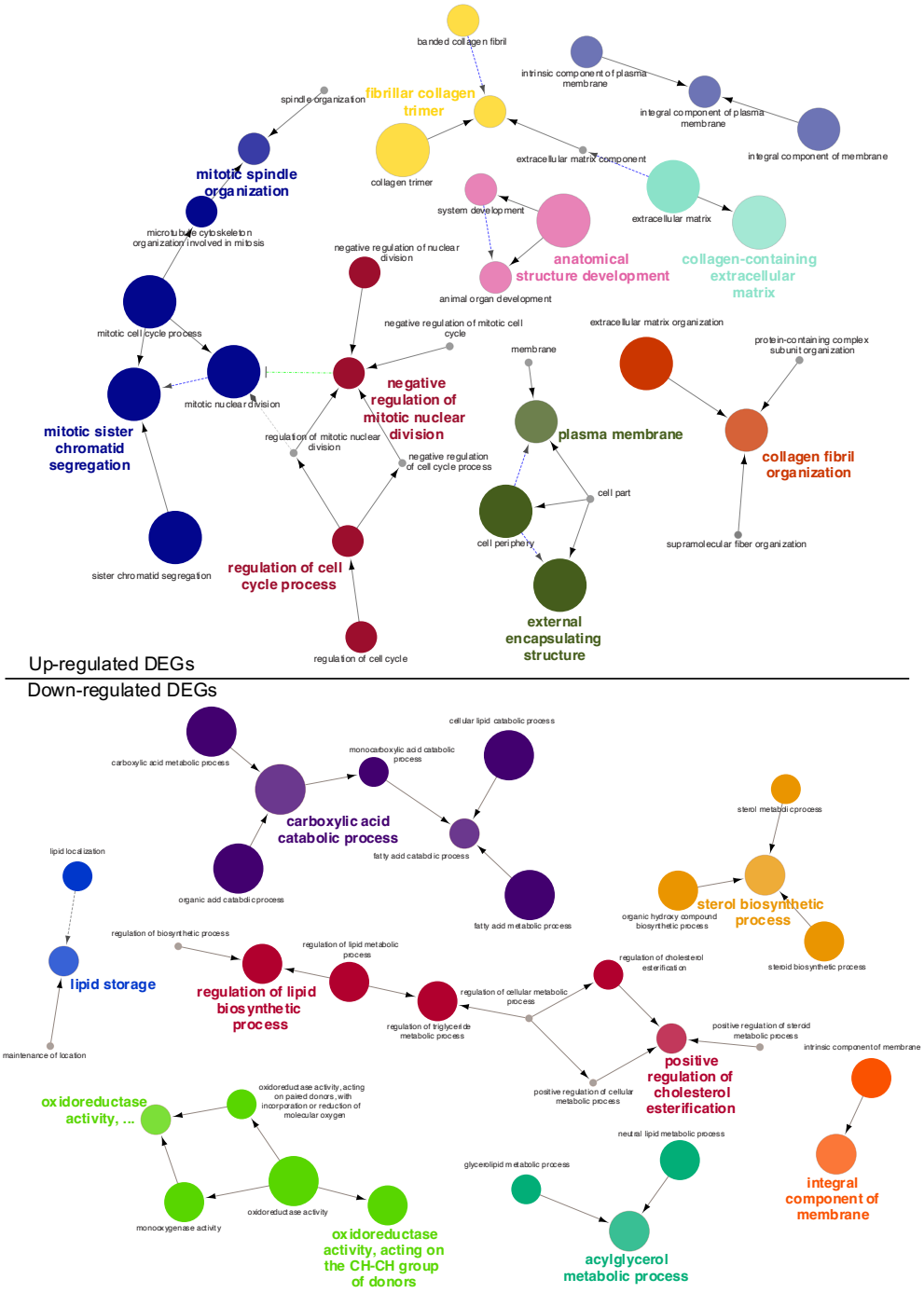


Figure 13: Enriched molecular functions, cellular component and biological processes associated with the DEGs (up-regulated and down-regulated) between juveniles and adults brushtail possums (*Trichosurus vulpecula*) RNASeq liver samples determined using Limma R-package. Related GO terms are connected. Coloured nodes indicate the significantly enriched GO terms (p-value ≤ 0.05), node size is constrained by p-value and node colour code indicates the functional class that they are involved. Small grey nodes are parents GO terms not significantly enriched.

Similar to functions of the genes identified with DESeq2, the 264 down-regulated genes identified with Limma are enriched in GOs associated with catabolic and metabolic processes, for example, “fatty acid catabolic process” (GO:0009062) with 7 genes, and “sterol biosynthetic process” (GO:0016126) with 6 genes. Associated regulation GO terms are also enriched such as "regulation of lipid biosynthetic process” (GO:0046890) with 9 genes and "regulation of triglyceride metabolic process" (GO:0090207) with 5 genes (Figure 13). The complete list of enriched GO terms from Limma DEGs can be found in Appendices (Appendix C3 Table 4 and 5).

WGCNA

The WGCNA analyses led to the identification of five clusters of genes that are significantly correlated with the maturity state of brushtail possums (p-value < 0.01). One cluster of genes (blue) is strongly negatively correlated with maturity (upregulated in juveniles) and the other four clusters (grey60, royalblue, cyan and magenta) are positively correlated with maturity (upregulated in adults, Figure 14).

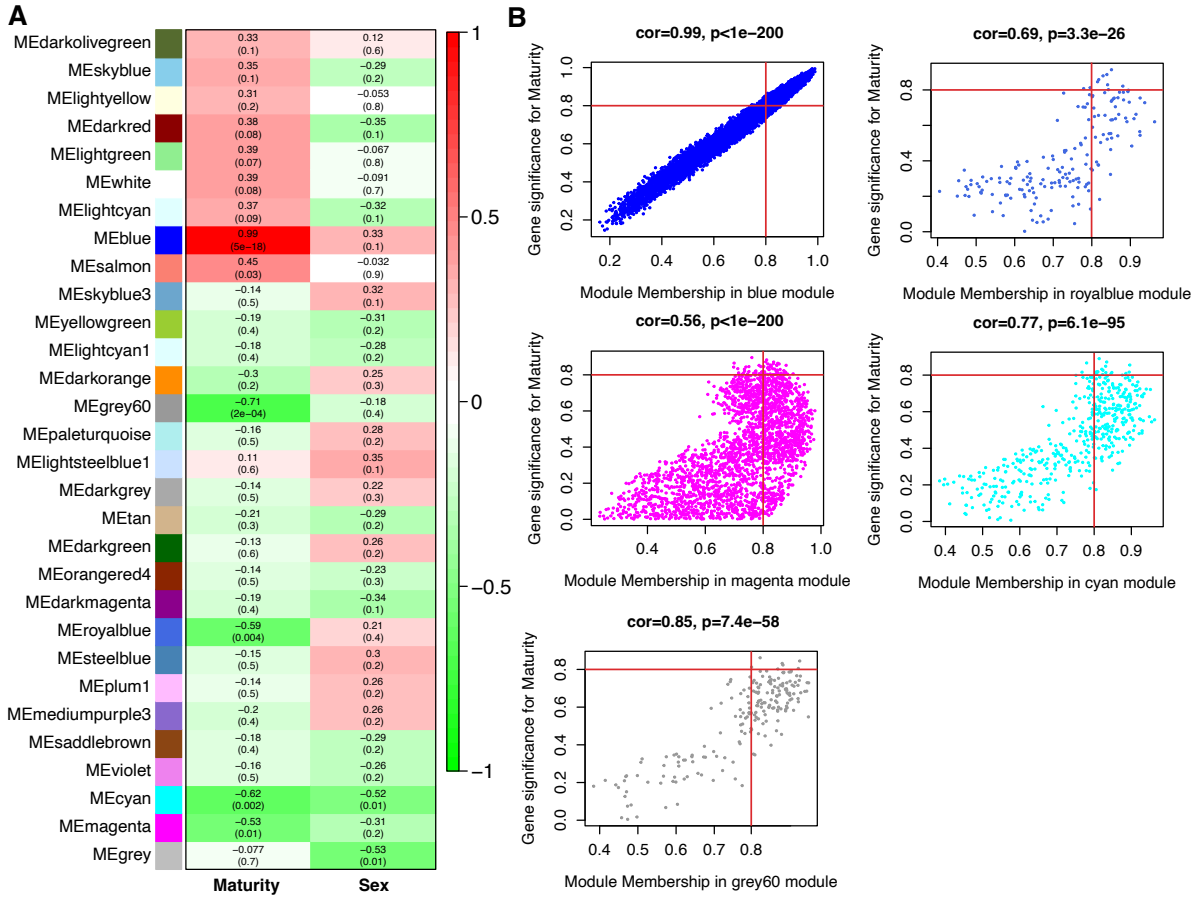


Figure 14: A) Heatmap of correlation of gene cluster with maturity (adult or juveniles) and sex of brushtail possums from WGCNA analysis. The colour and the first value of each cell correspond to the correlation score and the value in brackets corresponds to the p-value of the correlation. B) Plot of

Module Membership and Gene significance for maturity for each of the five significant clusters. Associated correlation score and the p-value is indicated on top of each plot, and MM and GS thresholds used to select significant genes are shown with red lines.

The module membership and gene significance cutoff led to the identification of 1761 differentially expressed genes associated with maturity (Figure 14) the majority of them from the blue cluster. Gene ontology enrichment analysis identified 64 enriched terms all from the blue cluster, upregulated in juveniles. Among the enriched terms, a lot are associated with cell cycle and replication (mitotic cell cycle process, cytokinesis, nuclear DNA replication). The rest of the biological processes upregulated in juveniles seem related to the inhibition of chromatids separation that can be associated with mitosis but also meiosis. Gene ontology enrichment on DEGs identified with WGCNA displays several genes associated with cellular components among which genes are associated with extracellular matrix or with chromosomes. (Figure 15) (Appendix C3 Table 6).

Figure 15: Enriched cellular component and biological processes associated with the DEGs (Down-regulated) between adult and juvenile brushtail possums (*Trichosurus vulpecula*) RNASeq liver samples using WGCNA analysis. Related GO terms are connected. Big nodes indicate the significantly

enriched GO terms (p-value ≤ 0.05), node sizes are constrained by p-value and node colour code indicates the functional class that they are involved. Small grey nodes are parents GO terms not significantly enriched.

Only three GO terms appear slightly enriched among genes downregulated in juveniles, the proteasome complex (GO:0000502), the endopeptidase complex (GO:1905369) and the peptidase complex (GO:1905368). The complete list of enriched GO terms from WGCNA DEGs can be found in Appendices (Appendix C3 Table 6).

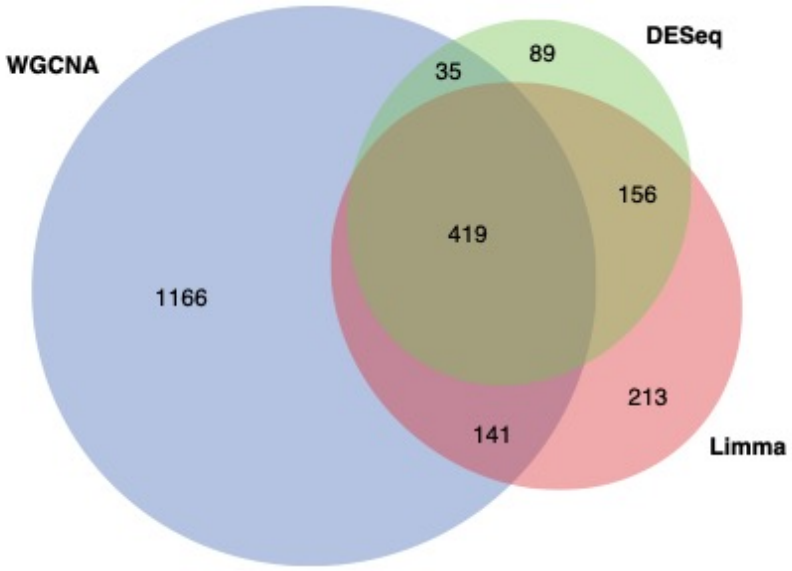


Figure 16: Proportional Venn diagram showing the number of common differentially expressed genes (DEGs) between juveniles and adults brushtail possums RNASeq liver samples for each analysis. Significance threshold for Limma and DEseq p-value < 0.05 and Log2FC >2 or < -2, and for WGCNA MM > 0.8 and GS > 0.8.

Comparing the discovered DEGs associated with maturity in the different analyses we can observe that the majority of genes discovered with DeSeq2 and Limma were also discovered using WGCNA (Figure 16). In contrast, WGCNA analysis resulted in a large number of differentially expressed genes unique to this analysis, while the results from DESeq2 and Limma results tend to overlap (Figure 16).

The large number of DEGs found only using WGCNA is not associated with different GO terms (Figure 17). Each of the three approaches identified many of the same GOs but also a significant proportion of unique GO terms (Figure 17).

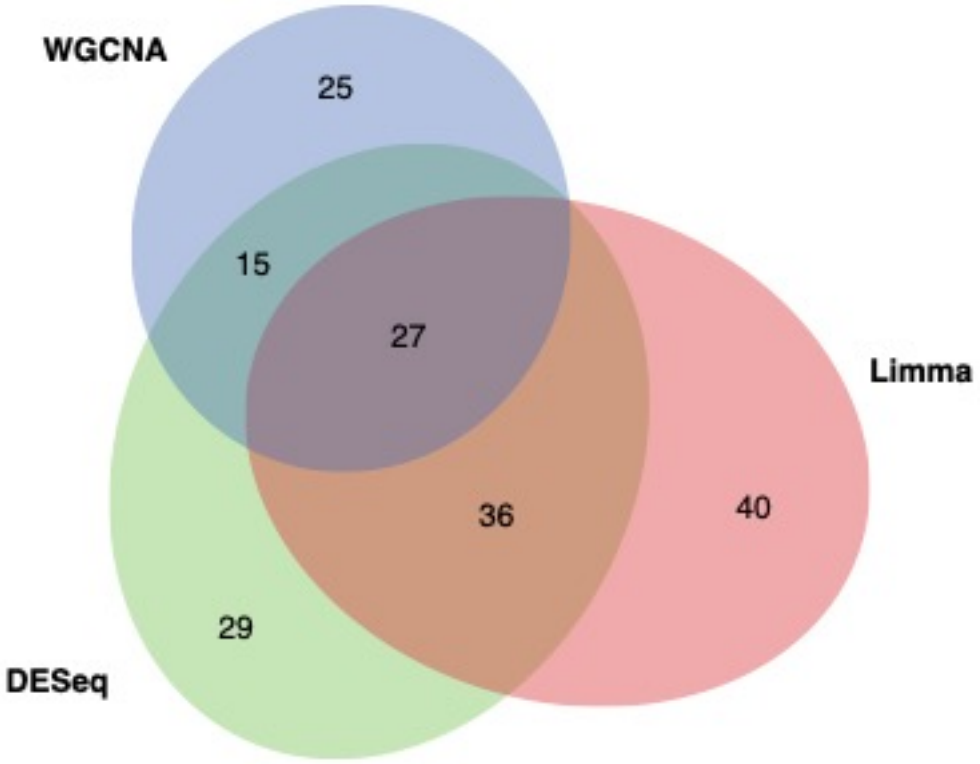


Figure 17: Proportional Venn diagram showing the number of common enriched gene ontology terms from the differentially expressed genes between juveniles and adults brushtail possums RNASeq liver samples for each analysis. Significance threshold for DEGs for Limma and DEseq are p-value < 0.05 and Log2FC >2 or < -2, and for WGCNA MM > 0.8 and GS > 0.8, the GOs significance threshold is p-value < 0.05.

Sexual differentiation

Adult possum samples were analysed to investigate differentially expressed genes (DGEs) between males and females. The juveniles and sample S 24 (suspected to be a furred juvenile) were removed from the analyses. The second principal component of variation (PC2) of individual expression partially distinguishes the males and females (Figure 18).

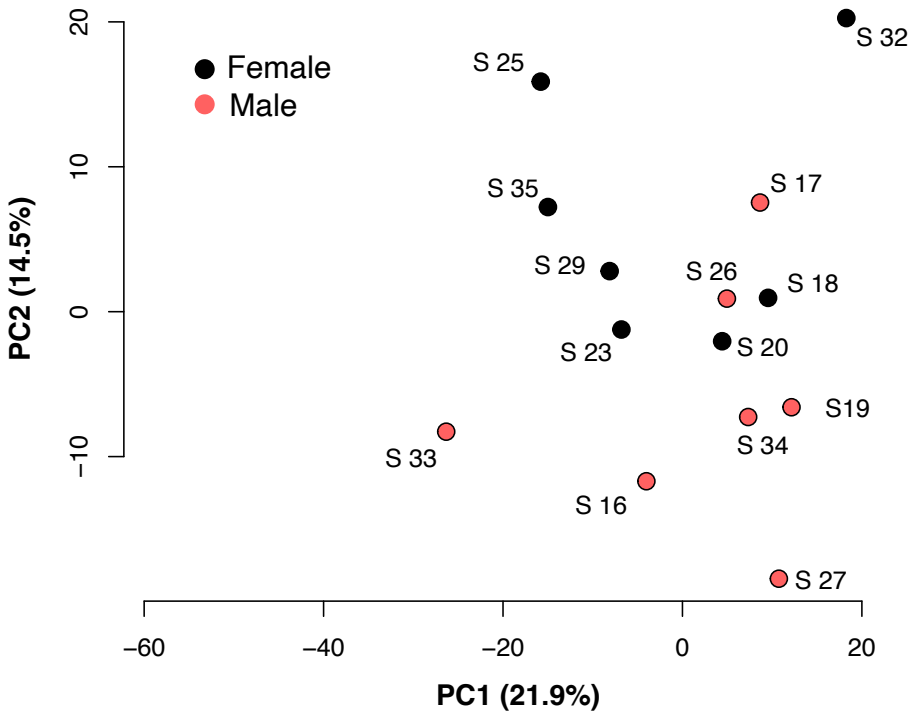


Figure 18: Plot of the first two principal components of the PCA applied to the normalized count data of the RNASeq data from the liver of adults brushtail possums. Samples are coloured by sex and the eigenvalue proportion of each axis is indicated below them.

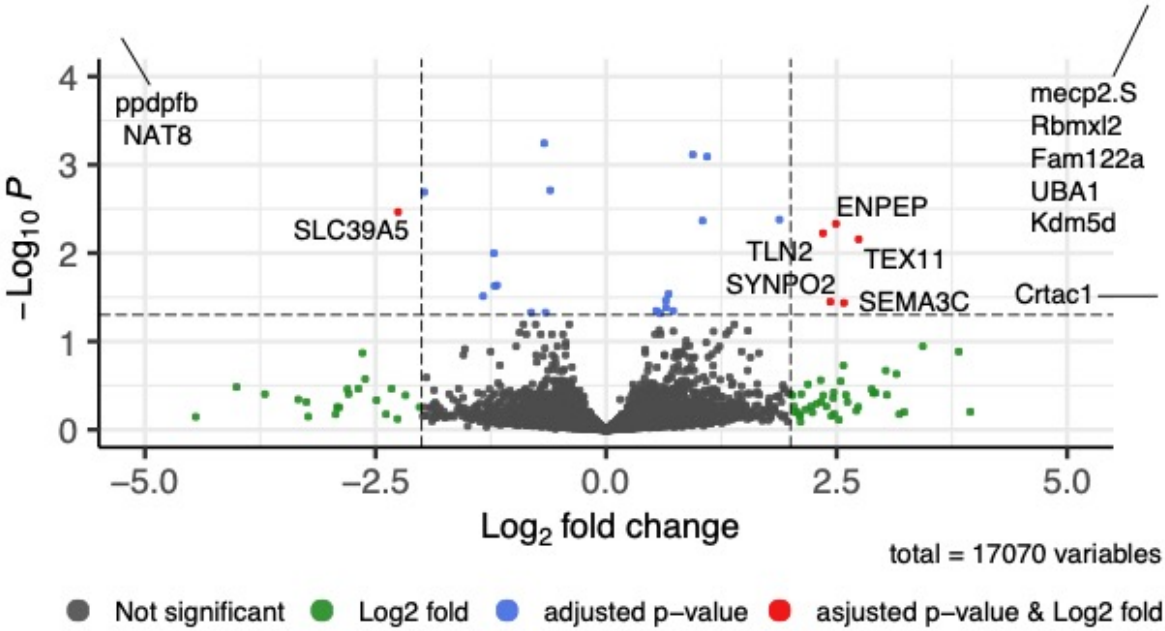


Figure 19: Visualization of DEGs volcano plots using R packages "DESeq2" from brushtail possums (*Trichosurus vulpecula)* RNASeq liver samples. The plot compared the DEGs between males and females. The representations are as follows: x-axis, log2 fold change; y-axis, -log10 of a p-value. The p-values < 0.05 are in blue dots, and logFC ≥ 2 and logFC ≤ –2 are in green dots; the significant DEGs with both satisfying values are in red dots and indicated with gene names. Black dots indicate the remaining genes present in the array that were not significantly changed. The genes that are upregulated in the males are on the right panel, and downregulated ones are on the left panel of the plot. Significant DEGs outside of the displayed intervals are indicated with the lines.

Using a cutoff of the FDR-adjusted p-value of 0.05 and the unshrank fold change lower than -2 or higher than 2 I identified 15 genes to be differentially expressed between females and males (Figure 19 & Table 2). 12 of those genes are upregulated in females and the others are downregulated. Despite the few significant DEGs, the produced heatmap shows a consistent pattern of expression between males of females in those genes (Figure 20) noticeable in normalized count data (Figure 21).

Table 2: 15 Significant DEGs (differently expressed genes) between females and males brushtail possums (*Trichosurus vulpecula*) liver samples. Genes are associated with their respective shrunk fold change and FDR-adjusted p-value.

GENES		FOLD CHANGE	P-VALUE
UBA1	Ubiquitin like modifier activating enzyme 1	13.400	2.53E-43
KDM5D	Lysine demethylase 5D	10.922	1.13E-38
HMGB3.S	High mobility group protein B3	10.359	1.73E-33
MECP2.S	Methyl-CpG-binding protein 2	7.838	9.82E-17
RBMXL2	RBMX like 2	6.855	7.13E-13
FAM122A	Protein FAM122A	5.980	1.96E-08
CRTAC1	Cartilage acidic protein 1	5.430	0.029
TEX11	Testis expressed 11	2.740	0.007
SEMA3C	Semaphorin 3C	2.578	0.037
ENPEP	Glutamyl aminopeptidase	2.492	0.005
SYNPO2	Synaptopodin 2	2.431	0.036
TLN2	Talin 2	2.351	0.006
SLC39A5	Solute carrier family 39 member 5	-2.256	0.003
PPDPFB	Pancreatic progenitor cell differentiation and proliferation factor B	-6.162	9.46E-05
NAT8	N-acetyltransferase 8 (putative	-22.027	3.14E-10

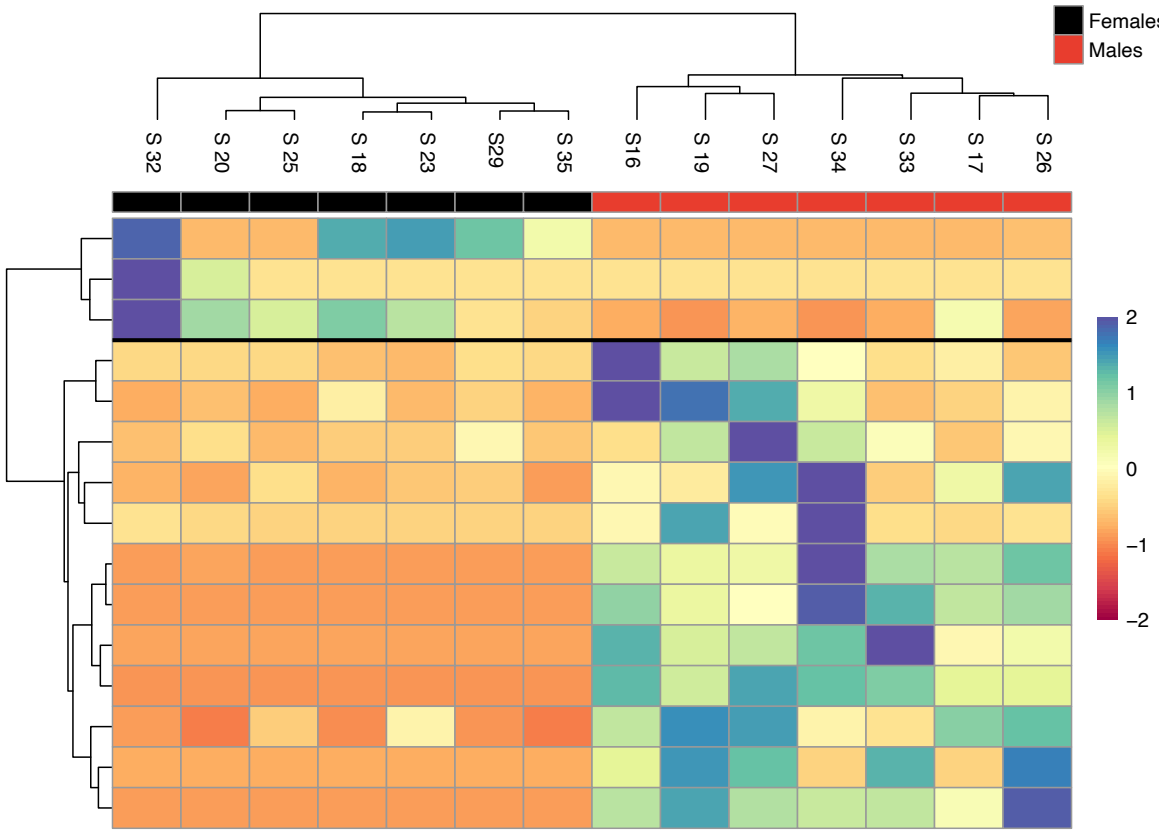


Figure 20: Heatmap and clustering based on a hierarchy of the differentially expressed genes (p.value < 0.00001) from the RNASeq data from liver samples of male and female brushtail possums, n = 15 DEGs. Upregulated genes are represented in red, and downregulated genes are represented in blue. Sex is indicated in the first row. (Males are in red and females are black).

Overall, the search for DEGs between adult males in females in liver samples led to few genes displaying significant expression differences. On the other hand, some of the genes with differential expression are associated with sex chromosomes providing an explanation of this pattern.

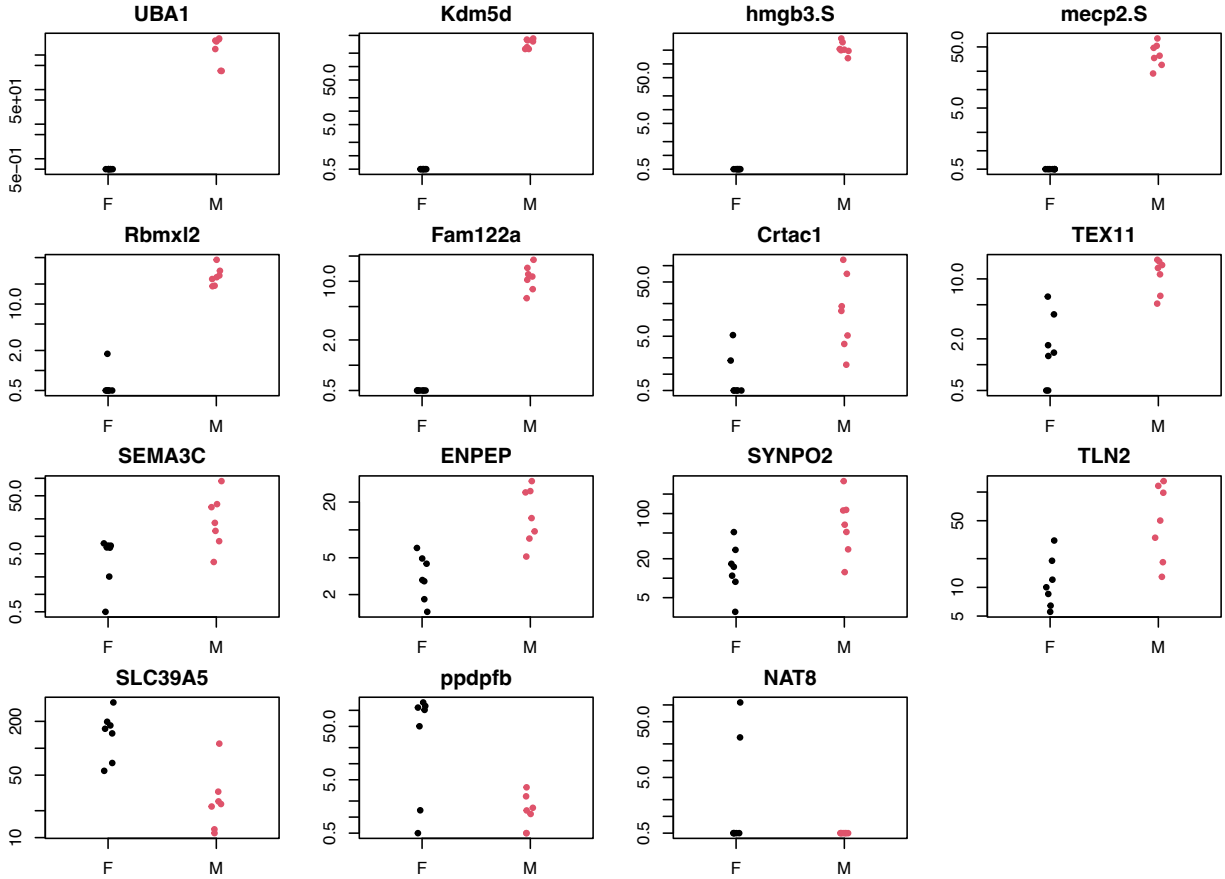


Figure 21: Normalized counts of each adult sample for the 15 significantly differently expressed genes between males and females (adjusted p-values < 0.05 and LFC < -2 or LFC > 2). Females are coloured in black and males in red.

In addition, the association with sex was also explored using WGCNA with two clusters (cyan and grey) positively correlated with sex (upregulated in females, Figure 14).

Fur colour phenotypes

Last, the same samples were put through similar analyses but looking at the three fur colour phenotypes' impact on the expression pattern in those adult brushtail possums (*Trichosurus vulpecula*) liver samples.

Principal components analysis on the normalized counts of each gene and samples do not display any apparent differentiation between fur colour phenotypes (Figure 22) and these results are confirmed with no significant DEGs between the three fur colour phenotypes. Looking more closely at the normalized counts of the 16 most significant DEGs shows no sign of different patterns of expression between the different fur colours among those genes

confirming that fur colour in *Trichosurus vulpecula* has no impact on the liver pattern of expression (Figure 23).

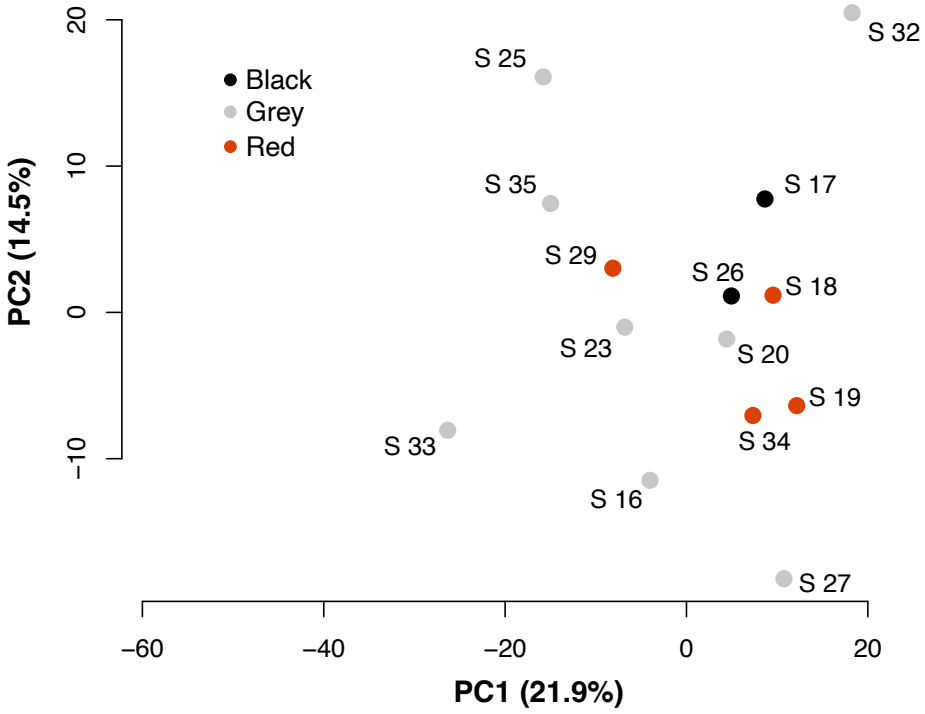


Figure 22: Plot of the first two principal components of the PCA applied to the normalized count data of the RNASeq data from the liver of adults brushtail possums. Samples are coloured by fur color and the eigenvalue proportion of each axis is indicated below them.

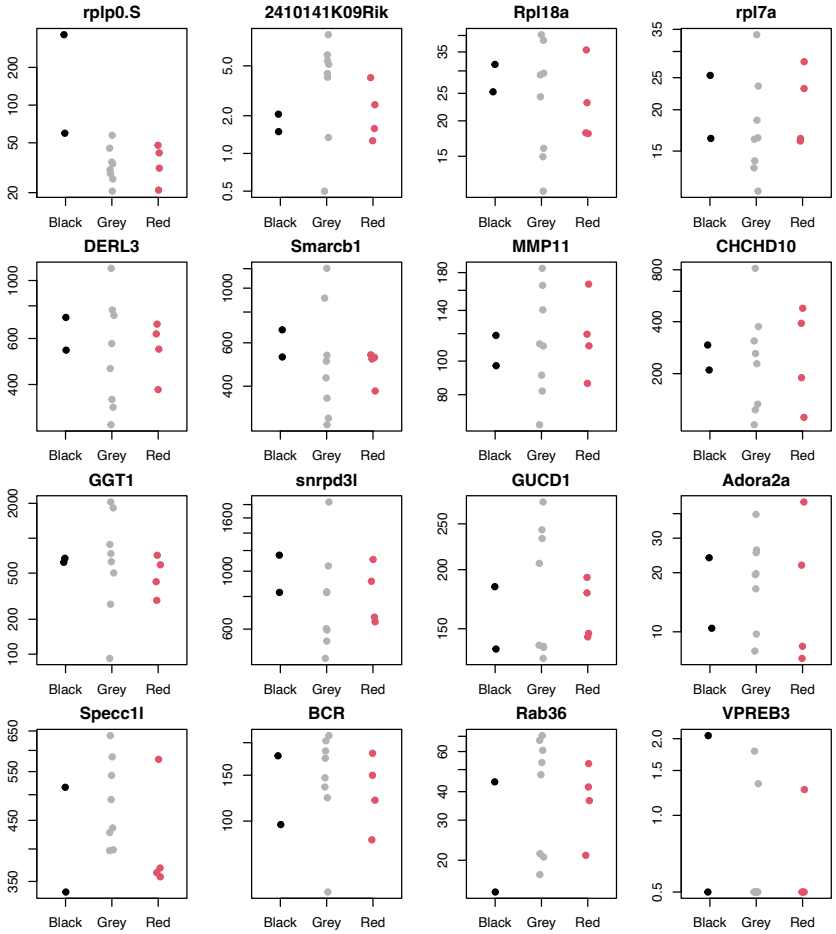


Figure 23: Normalized counts of each adult sample for the top 16 most significantly differentially expressed genes between the three fur colours phenotypes among brushtail possums (*Trichosurus vulpecula)* RNASeq liver samples.

Power algorithm

The first results are showing a significant influence of sample size on the proportion of common discovered genes with a median starting at less than 20% for 2 samples in each group and reaching around 80% for groups of 8 samples (Figure 24, A). The increasing number of common discovered genes seems to come with reduced variance and be consistent with the increase in sample size, the biggest gap seems to correspond with the group size from 2 to 3 samples (Figure 24, A). The reduced variance corresponds to the subset of the sample being more similar to each other with the increased sample size.

Another result to mention is the similar evolution of common discovered enriched GOs (Figure 15, B) but the variance is very high for group sizes of 3, 4 and 5 with the number of common GOs varying from 0% to 85%. Similarly, to the common DEGs, the variance is reducing with the number of samples per group and as the number of common GOs terms is increasing (Figure 24, B).

The threshold p-value seems also to be crucial when recovering DEGs, there is a clear increase of common DEGs with an increased maximum p-value. For 2 samples per group increasing the p-value from 0.00001 to 0.01 allows the proportion of common DEGs to reach almost 50%. This increase in common DEGs is less marked with a higher number of samples, from more than 25% to around 10% (Figure 24, C). A side effect of the increase of the maximum p-value for DEGs significance is an increase in the false discovery rate, this rate represents the proportion of significant DEGs that are not considered significant when looking at the full samples set and with a p-value of 0.00001. This false discovery rate also reduces with the number of samples but never reaches more than 26% of the number of discovered DEGs.

It is very interesting to notice how many common DEGs are recovered with increasing the number of samples in only one group (Figure 24, D), for example with 4 juveniles and 2 adults the mean of ten runs reaches 26% of common significant DEGs when 4 juveniles and 14 adults the value is slightly less than 63%.

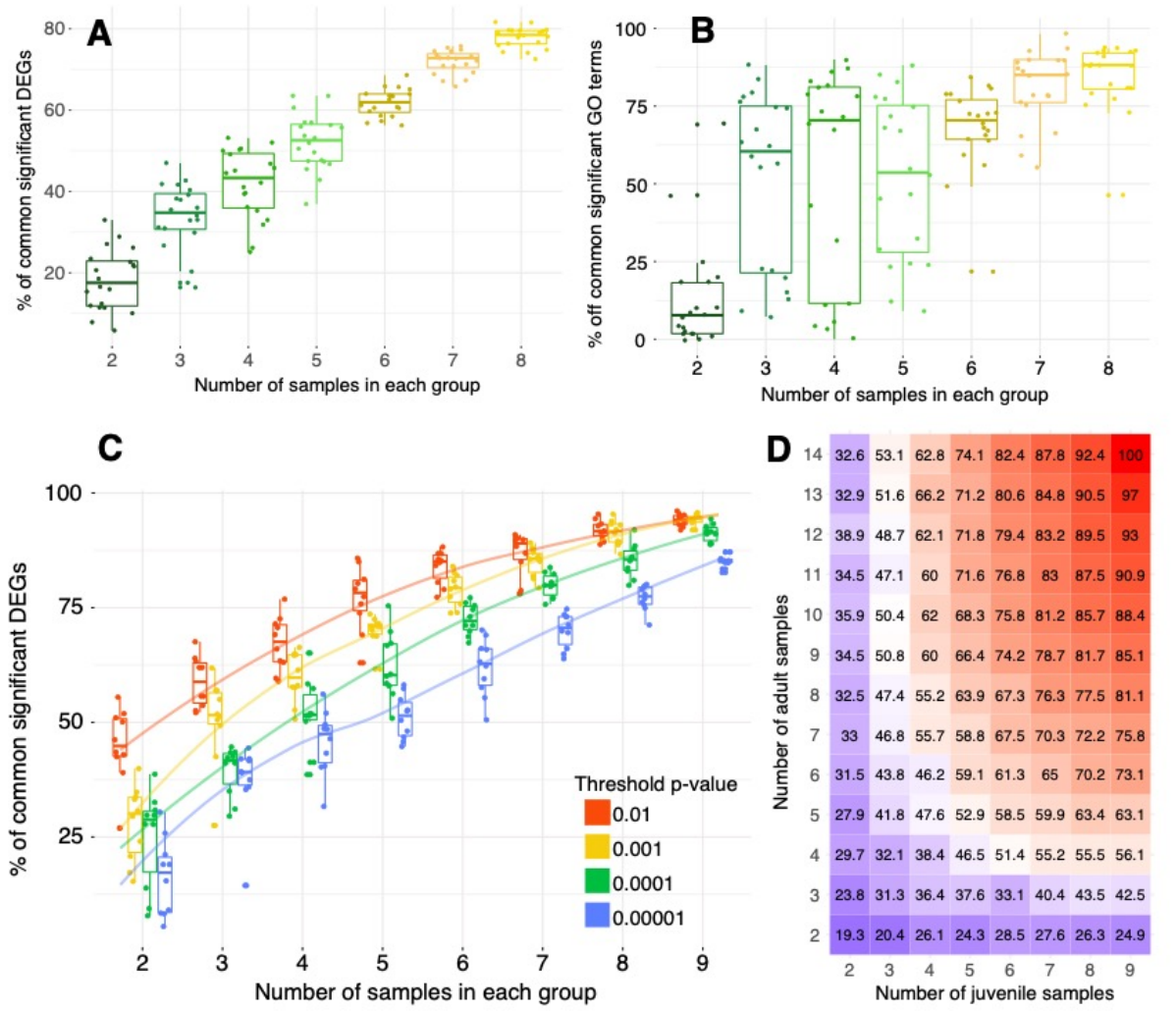


Figure 24: A) Proportion of common significant DEGs (differentially expressed genes) between the results from differential expression analysis on the whole set of samples (9 juvenile and 14 adult brushtail possum livers samples) and a subset of samples with a set number of samples. **B)** Proportion of common significant GO terms (Gene ontology) between the results of GO enrichment analysis from differential expression analysis and on the whole set of samples (9 juvenile and 14 adult brushtail possum livers samples) and a subset of samples with a set number of samples. **C)** Proportion of common significant DEGs (differentially expressed genes) between the results from differential expression analysis on the whole set of samples (9 juvenile and 14 adult brushtail possum livers samples) and a subset of samples with a set number of samples and a different set p-value. **D)** Mean proportion in % (from 10 runs) of common significant DEGs between the results from differential expression analysis on the whole set of samples (9 juvenile and 14 adult brushtail possum livers samples) and a subset of samples (randomly chosen).

Discussion

In this study I investigated differentially expressed genes (DEGs) in the liver of juvenile and adult brushtail possums (*Trichosurus vulpecula*). Despite using a reference genome, a significant proportion of the DNA sequences (reads) could not be assigned to any genes

because they did not map to annotated regions of the reference genome. This might result from the presence of DNA in the RNA samples that would lead to the amplification and sequencing of intronic and intergenic sequences, or the expression of genes that have not yet been annotated (Bond et al. 2023). It is essential to remember that this analysis was done using liver samples, so gene expression differences relate to the digestive and metabolic activities of this organ. Using the liver proved to be beneficial for this study but it does have limit inferences to a single tissue type. The first advantage of using liver is experimental as this organ tends to be easy to sample and after extraction consistently yields a good amount of high quality RNA. The liver also seems to be a good choice when looking at differential expression during development as it carries genes involved in organ and tissue formation and genes involved in food and/or toxin processing.

One interesting result from examination of gene expression levels is how one individual classified as an adult when tissue was sampled (S24), clustered with the juvenile possum samples in the principal component analysis of the normalized read count. The significant DEGs of sample S24 matched the expression pattern of the furred juvenile (sample S21) from which one can infer that sample S24 is also from a furred juvenile. The analysis based on age class (Figures 8-9) tends to confirm this inference, where the furred juvenile group variance is quite low for the genes investigated. This example of potential incorrect age identification suggests that maturity in the brushtail possum is not easily determined based on external phenotype but is more clear-cut in the expression phenotype. Juvenile possums go through a fast weight gain before reaching maturity and furred juveniles can have a comparable body mass and aspect of mature adults (Buaboocha and Gemmell 1997; Jackson et al. 2019). The pattern of differential gene expression across different juvenile age classes involve similar genes but I found evidence of the impact of age on the level of expression. Larger samples of juvenile possums covering the full range of days postpartum would allow these results to be further explored. Sampling possums from the wild could lead to including furred juveniles even if the targets are adults, therefore, it is valuable to identify the set genes involved in sexual maturity and development for further studies involving adult and near-adult individuals.

When comparing adult with juvenile possum samples I identified 475 significant DEGs and studied their potential function with gene ontology terms and Kegg pathways. As expected, the set of up-regulated genes in juveniles were enriched for terms and pathways associated with tissue development cell cycle and extracellular matrix. Compared to adults, juvenile possums

are investing more resources in their growth so this result resonates with a high rate of mitosis required for the development of tissues within the liver of the juvenile possum. Particularly Kegg pathways highlighted the collagens genes family, coding for the fibrous protein called collagen the main component of skin, bone, cartilage, blood vessels, and teeth in vertebrates but more importantly, collagen is involved in morphogenesis and tissue interaction during vertebrate development (Trelstad 1973). On the other hand, the genes downregulated in juveniles compared to adult possums are associated with catabolic processes that could be related to diet changes between juveniles and adults expressed in the liver.

In the comparison of gene expression in liver samples, seven genes of the cytochrome P450s (CYP) family were down-regulated in juvenile compared to adult possums. These genes code for membrane-bound proteins that catalyse the metabolism of a diverse array of xenobiotic compounds and endogenous substrates, they are regulated in a tissue-specific and temporal manner and are critically important in detoxifying and eliminating drugs, chemicals and environmental pollutants. In mammals, various CYPs also participate in the biosynthesis and metabolism of steroids (El-Merhibi et al. 2011; Seliskar and Rozman 2007). Increased expression in adults could be explained by the change of diet of juveniles during their development from maternal milk to leaves (P. E. Cowan 1989; How and Hillcox 2000; Jarman 2006). A similar increase of CYPS (P450s) expression during development has been documented in mice (Peng et al. 2012). These genes might be worth considering for further study that focuses on toxin resistance of the western Australian possum subspecies (Twigg et al. 2003) and knowing that their expression varies during development adds an extra evidence of their possible association with the diet of the possums. Extra care will be needed if cytochrome P450s are identified in differential expression analyses as they could be a sign of different maturity levels of the possums samples rather than adaptations to local diet.

Comparing approaches

Two tools were used to identified genes that differed significantly in expression levels within the liver of adult and juvenile possums. Even if the majority of the results from Limma are similar to the one found with DESeq2 some differences are noticeable. The first noticeable result is the sensitivity to FDR (false detection rate) adjusted p-value threshold. It seems that DESeq2 tends to produce lower p-values allowing the use of more stringent threshold when using Limma seems to be more sensitive to p-value threshold (Figure 25).

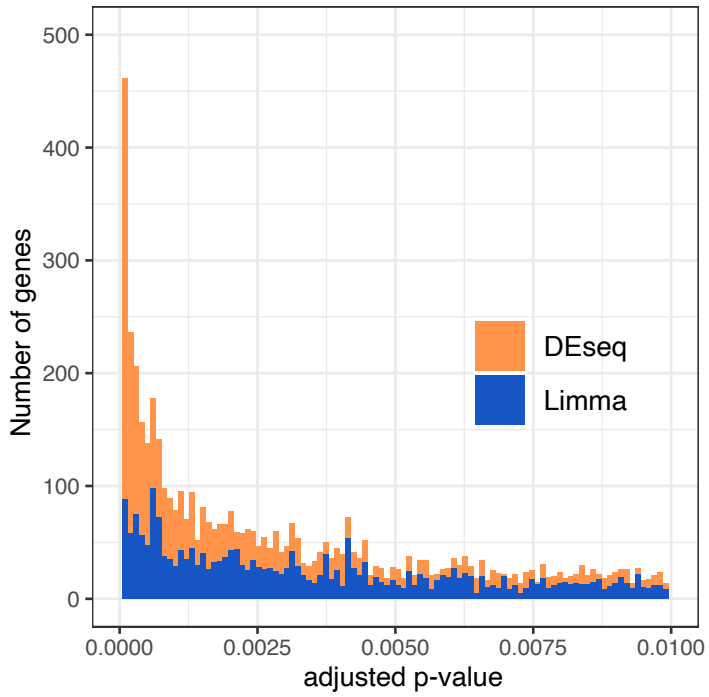


Figure 25: Histogram showing the number of genes for each associated p-value for two different differential expression analysis packages using brushtail possums RNASeq liver samples. In blue bars are the results of the analysis from Limma in orange is the results from DESeq2.

The differences between different differential expression software have been studied, usually using simulated data, and they tend to show that Limma and DESeq2 are both effective software with sensitivity differences (Seyednasrollah, Laiho, and Elo 2013). These differences led to the decision of using a threshold of 0.00001 for DESeq2 and a threshold of 0.05 for Limma.

Three tools were used to explore the functional community of genes identified as showing significant variation in expression levels within the liver of adult and juvenile possums. Looking at Gene ontology terms enrichment analyses, the results of the three analyses are consistent with each other. The enriched GOs among the upregulated genes in juveniles are associated with tissue organisation and cell cycle, corresponding to the development of the organ as the liver is developing, especially in pouch-young individuals. When the enriched GOs among the upregulated genes in adults (downregulated in juveniles) correspond to the metabolism of molecules such as triglyceride, lipid and/or carboxylic acid, which reflects the role of the liver in the digestion of adult individuals. Very similar results were found in the

mouse liver development (Kung et al. 2010; Li et al. 2009) confirming the relevance of the data and the performed analysis.

WGCNA analysis is less popular but also produced interesting results. The analysis identified more DEGs but used a threshold that is quite large. One downside observed is that most of the DEGs were upregulated in the juveniles and a very small number were upregulated in the adults (so few that the GO enrichment analysis did not produce any results). It is quite interesting to see that very similar GOs are showing enrichment between the DeSeq2 and WGCNA analyses despite the sample of genes being quite different. The analysis does not produce gene-specific p-value and FDR, making the comparison with the result of the other analyses difficult.

The choice of cutoff p-value in DGE analyses has a marked effect downstream and so deserves discussion. The normal admitted threshold for the p-value is 0.05 but lots of studies are adapting the cutoff value to suit their purpose. Using a very stringent value lowers the number of discovered genes and allows a focus on the most important gene functions (Dalman et al. 2012). Using a lower threshold provides a wider view with larger number of genes identified and a more general conclusion on the potential effect under investigation. For the next chapter, I will also use a stringent cutoff as it is appropriate to focus on strongly specific metabolism and reduce the False discovery rate (FDR).

As expected, when looking at the differences between adult females and adult males only 15 genes were detected as differing in their expression levels within the liver samples. This analysis used a less stringent p-value (> 0.5) and using unshrunk fold change. The absence of expression differentiation between the sexes is expected in the liver, and the 15 genes recovered can be explained as several of them are associated with sex chromosomes. Using gonads, brain or hormonal glands would be more suitable for isolating sex-related genes (Paranjpe et al. 2019; Tao et al. 2018; Yatsu et al. 2016). On the other hand, these results are quite positive because it means that sex is unlikely to interact with results of other studied characters in brushtail possums (such as toxin resistance) using liver samples.

Lastly, fur colour displays no significant DEGs between the three recorded phenotypes, those results could be explained in different ways, like the last analysis the choice of the liver must be the main reason for the lack of signal for fur colour-related genes. Secondly, those results resonate with the evidence of strong mixing in the New Zealand population, it appears that fur

colour does not carry any correlation with genetic distance in the New Zealand populations (Pattabiraman et al. 2021).

The goal of the power analysis on this particular sample set tries to establish the impact of sample size and p-value threshold on the number of discovered genes. The analysis is trying to help design and predict the efficiency of further differential expression analysis on brushtail possums' liver RNASeq data. Other subspecies of possums in Australia are of strong interest for differential expression study for their phenotypical differences but also are more challenging to sample. Given their conservation status within Australia, knowing the minimum number of samples that are needed to confidently acquire significant signals is crucial when designing experiments on wild and endangered animals. My analysis shows that sample size is crucial in differential expression analysis, first to discover as many gene as possible that are differentially expressed but also in order to reduce the number of false positives. Controlling the false discovery rate is a crucial step of every differential expression study (Korthauer et al. 2019), DESeq2 incorporate it in its pipeline using Benjamini and Hochberg's step-up procedure (Benjamini and Hochberg 1995; Love, Huber, and Anders 2014) that adjusts the p-value. Many studies have been trying to establish what the minimum number of samples per group is for a DE analysis that will provide reasonable results (Conesa et al. 2016; Hart et al. 2013). The current study is the perfect opportunity to quantify the effect of sample size for our particular case (possum liver) and assess the cost and benefit of more samples. Moreover, this study is showing us that GO terms enrichment analysis is quite sensitive to differences in the gene set and especially to potential false positives. Another very useful result is that increasing the size of one of the groups is beneficial for the discovery of significant DEGs, in the condition of the sampling of wild animals one of the characters can have limited sampling availability making this result very valuable. Possum trapping in New Zealand provides a sampling opportunity, but this situation is more difficult to translate in the western Australian context where *Trichosurus vulpecula* is conserved and is living in an arid low density environment.

This chapter has contributed significantly to our understanding of gene expression patterns in brushtail possums, shedding light on the genes associated with developmental processes within their liver tissue. Additionally, our investigation into the impact of sample size on data generation underscores the robustness of our findings, demonstrating that even with a smaller sample size in one group, a substantial amount of significant genes can be found. These insights not only deepen our understanding of the possums' biology but also bear significance in the

broader exploration of their toxin-resistance mechanisms. This distinction will allow for the distinction between genes that might be linked to toxin resistance and those that could be associated with different factors, such as developmental processes.

References

Abbott, Ian. 2012. "Original Distribution of Trichosurus Vulpecula (Marsupialia : Phalangeridae) in Western Australia, with Particular Reference to Occurrence Outside the Southwest." *Journal of the Royal Society of Western Australia* 95: 83–93. https://www.rswa.org.au/publications/Journal/95(2)/Abbottpp.83-93.pdf (May 10, 2022).

Alanagreh, Lo'ai, Caitlin Pegg, Amritha Harikumar, and Mark Buchheim. 2017. "Assessing Intragenomic Variation of the Internal Transcribed Spacer Two: Adapting the Illumina Metagenomics Protocol." *PLoS ONE* 12(7): e0181491. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0181491 (February 10, 2022).

Alexander, Helen K., Guillaume Martin, Oliver Y. Martin, and Sebastian Bonhoeffer. 2014. "Evolutionary Rescue: Linking Theory for Conservation and Medicine." *Evolutionary Applications* 7(10): 1161–79.

Anders, Simon, and Wolfgang Huber. 2010. "Differential Expression Analysis for Sequence Count Data." *Nature Precedings*: 1–1. https://www.nature.com/articles/npre.2010.4282.1 (July 5, 2022).

Andrews, Simon, and others. 2010. "FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010." *Https://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/* 1(1): http://www.bioinformatics.babraham.ac.uk/projects/. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/%0Ahttp://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/ (March 17, 2020).

Ansari, Mina Hojat et al. 2019. "Plio-Pleistocene Diversification and Biogeographic Barriers in Southern Australia Reflected in the Phylogeography of a Widespread and Common Lizard Species." *Molecular Phylogenetics and Evolution* 133: 107–19.

Archer, Michael et al. 2006. "Current Status of Species-Level Representation in Faunas from Selected Fossil Localities in the Riversleigh World Heritage Area, Northwestern Queensland." *Alcheringa* 30(4): 1–17. https://www.tandfonline.com/doi/abs/10.1080/03115510609506851 (October 1, 2021).

Bandelt, H J, P Forster, and A Rohl. 1999. "Median-Joining Networks for Inferring

Intraspecific Phylogenies." *Molecular Biology and Evolution* 16(1): 37–48. https://academic.oup.com/mbe/article-lookup/doi/10.1093/oxfordjournals.molbev.a026036 (March 19, 2019).

Bass, Chris, Ian Denholm, Martin S. Williamson, and Ralf Nauen. 2015. "The Global Status of Insect Resistance to Neonicotinoid Insecticides." *Pesticide Biochemistry and Physiology* 121(10): 78–87. https://www.sciencedirect.com/science/article/pii/S0048357515000826 (March 15, 2020).

Beck, Robin M.D. 2008. "A Dated Phylogeny of Marsupials Using a Molecular Supermatrix and Multiple Fossil Constraints." *Journal of Mammalogy* 89(1): 175–89. https://academic.oup.com/jmammal/article/89/1/175/1020874 (October 1, 2021).

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57(1): 289–300.

Bernt, Matthias et al. 2013. "MITOS: Improved de Novo Metazoan Mitochondrial Genome Annotation." *Molecular Phylogenetics and Evolution* 69(2): 313–19.

Besnier, Francois et al. 2014. "Human-Induced Evolution Caught in Action: SNP-Array Reveals Rapid Amphi-Atlantic Spread of Pesticide Resistance in the Salmon Ectoparasite Lepeophtheirus Salmonis." *BMC Genomics* 15(1): 1–18.

Bindea, Gabriela et al. 2009. "ClueGO: A Cytoscape Plug-in to Decipher Functionally Grouped Gene Ontology and Pathway Annotation Networks." *Bioinformatics* 25(8): 1091–93. https://academic.oup.com/bioinformatics/article/25/8/1091/324247 (July 4, 2022).

Bintanja, Richard, Roderik S.W. Van De Wal, and Johannes Oerlemans. 2005. "Modelled Atmospheric Temperatures and Global Sea Levels over the Past Million Years." *Nature* 437(7055): 125–28. https://pubmed.ncbi.nlm.nih.gov/16136140/ (May 30, 2022).

Black, Karen H., Michael Archer, Suzanne J. Hand, and Henk Godthelp. 2012. "The Rise of Australian Marsupials: A Synopsis of Biostratigraphic, Phylogenetic, Palaeoecologic and Palaeobiogeographic Understanding." In *Earth and Life: Global Biodiversity, Extinction Intervals and Biogeographic Perturbations Through Time*, Springer Netherlands, 983–1078. https://link.springer.com/chapter/10.1007/978-90-481-3428-1_35 (June 21, 2021).

Bond, Donna M. et al. 2023. "The Admixed Brushtail Possum Genome Reveals Invasion History in New Zealand and Novel Imprinted Genes." *Nature Communications 2023 14:1* 14(1): 1–17. https://www.nature.com/articles/s41467-023-41784-8 (October 18, 2023).

Bouckaert, Remco R. 2010. "DensiTree: Making Sense of Sets of Phylogenetic Trees." *Bioinformatics* 26(10): 1372–73. https://academic.oup.com/bioinformatics/article-abstract/26/10/1372/192963 (September 8, 2021).

Boyle, C. Mary. 1960. "Case of Apparent Resistance of Rattus Norvegicus Berkenhout to Anticoagulant Poisons." *Nature* 188(4749): 517.

Braithwaite, R. W. 1990. "Australia's Unique Biota: Implications for Ecological Processes." *Journal of Biogeography* 17(4–5): 347–54.

Brown, Jason L. et al. 2018. "Paleoclim, High Spatial Resolution Paleoclimate Surfaces for Global Land Areas." *Scientific Data* 5. https://www.nature.com/articles/sdata2018254 (March 18, 2022).

Brown, Thomas M., and Gregory T. Payne. 1988. "Experimental Selection for Insecticide Resistance." *Journal of economic entomology* 81(1): 49–56.

Bryant, Litticia M., and Susan J. Fuller. 2014. "Pleistocene Climate Fluctuations Influence Phylogeographical Patterns in Melomys Cervinipes across the Mesic Forests of Eastern Australia." *Journal of Biogeography* 41(10): 1923–35. https://onlinelibrary.wiley.com/doi/full/10.1111/jbi.12341 (October 5, 2022).

Buaboocha, W., and R. T. Gemmell. 1997. "Development of Lung, Kidney and Skin in the Brushtail Possum, Trichosurus Vulpecula." *Cells Tissues Organs* 159(1): 15–24. https://www.karger.com/Article/FullText/147960 (July 5, 2022).

Burridge, Christopher Paul. 2012. "Update: Divergence of Island Biotas When They Were Not Always Islands." *Frontiers of Biogeography* 3(4). https://escholarship.org/uc/item/74n463jr (October 17, 2022).

Byrne, M. 2008. "Evidence for Multiple Refugia at Different Time Scales during Pleistocene Climatic Oscillations in Southern Australia Inferred from Phylogeography." *Quaternary Science Reviews* 27(27–28): 2576–85.

de Castro, Érika C.P., Mika Zagrobelny, Márcio Z. Cardoso, and Søren Bak. 2018. "The Arms Race between Heliconiine Butterflies and Passiflora Plants – New Insights on an Ancient Subject." *Biological Reviews* 93(1): 555–73. https://onlinelibrary.wiley.com/doi/full/10.1111/brv.12357 (May 19, 2023).

Chapple, David G., J. Scott Keogh, and Mark N. Hutchinson. 2005. "Substantial Genetic Substructuring in Southeastern and Alpine Australia Revealed by Molecular Phylogeography of the Egernia Whitii (Lacertilia: Scincidae) Species Group." *Molecular Ecology* 14(5): 1279–92. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-294X.2005.02463.x (October 5, 2022).

Chaves-Campos, Johel, Steven G. Johnson, and C. Darrin Hulsey. 2011. "Spatial Geographic Mosaic in an Aquatic Predator-Prey Network." *PLoS ONE* 6(7): R75. http://genomebiology.biomedcentral.com/articles/10.1186/gb-2005-6-9-r75 (March 11, 2020).

Clark, Peter U. et al. 2009. "The Last Glacial Maximum." *Science* 325(5941): 710–14. https://pubmed.ncbi.nlm.nih.gov/19661421/ (May 30, 2022).

Clout, M, and K Ericksen. 2000. "Anatomy of a Disastrous Success: The Brushtail Possum as an Invasive Species." *The brushtail possum-biology, impact and management of an introduced marsupial*: 1–9. internal-pdf://0.0.3.94/10012031131.html.

Clout, M N, and Kris Ericksen. 2000. "Anatomy of a Disastrous Success: The Brushtail Possum as an Invasive Species." In *The Brushtail Possum: Biology, Impact and Management of an Introduced Marsupial*, , 1–9. https://ci.nii.ac.jp/naid/10012031131/ (March 15, 2020).

Coleman, Annette W. 2013. "Analysis of Mammalian RDNA Internal Transcribed Spacers." *PLoS ONE* 8(11). /pmc/articles/PMC3834078/ (February 10, 2022).

Conesa, Ana et al. 2016. "A Survey of Best Practices for RNA-Seq Data Analysis." *Genome Biology* 17(1): 1–19. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0881-8 (July 5, 2022).

Cooper, Christine E et al. 2018. "Geographical Variation in the Standard Physiology of Brushtail Possums ({Trichosurus}): Implications for Conservation Translocations." *Conserv Physiol* 6(1). https://academic.oup.com/conphys/article/6/1/coy042/5075449 (July 1, 2019).

Costa-Silva, Juliana, Douglas Domingues, and Fabricio Martins Lopes. 2017. "RNA-Seq Differential Expression Analysis: An Extended Review and a Software Tool." *PLoS ONE* 12(12): e0190152. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0190152 (July 5, 2022).

Cowan, P. 2016. "Characteristics and Behaviour of Brushtail Possums Initially Moving into a Depopulated Area." *New Zealand Journal of Zoology* 43(3): 223–33.

Cowan, P. E. 1989. "Changes in Milk Composition during Lactation in the Common Brushtail Possum, Trichosurus Vulpecula (Marsupialia: Phalangeridae)." *Reproduction, Fertility and Development* 1(4): 325–35. https://www.publish.csiro.au/rd/rd9890325 (July 5, 2022).

———. 1992. "The Eradication of Introduced Australian Brushtail Possums, Trichosurus Vulpecula, from Kapiti Island, a New Zealand Nature Reserve." *Biological Conservation*

61(3): 217–26. https://www.sciencedirect.com/science/article/pii/000632079291119D (March 15, 2020).

Cowan, P E. 1990. "Brushtail Possum." *The handbook of New Zealand mammals*: 68–89.

Cristiano, Luigi. 2021. "The Pseudogenes of Eukaryotic Translation Elongation Factors (EEFs): Role in Cancer and Other Human Diseases." *Genes & Diseases*.

Dalman, Mark R., Anthony Deeter, Gayathri Nimishakavi, and Zhong Hui Duan. 2012. "Fold Change and P-Value Cutoffs Significantly Alter Microarray Interpretations." *BMC bioinformatics* 13 Suppl 2(2): 1–4. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-S2-S11 (December 19, 2022).

Davis, M. B., and R. G. Shaw. 2001. "Range Shifts and Adaptive Responses to Quaternary Climate Change." *Science* 292(5517): 673–79. https://www.science.org/doi/full/10.1126/science.292.5517.673 (May 30, 2022).

Dawkins, R., and J. R. Krebs. 1979. "Arms Races between and within Species." *Proceedings of the Royal Society of London - Biological Sciences* 205(1161): 489–511.

Deakin, Janine E. et al. 2013. "Towards an Understanding of the Genetic Basis behind 1080 (Sodium Fluoroacetate) Tolerance and an Investigation of the Candidate Gene ACO2." *Australian Journal of Zoology* 61(1): 69–77. http://www.publish.csiro.au/?paper=ZO12108 (March 15, 2020).

Denholm, I, and M W Rowland. 1992. *TACTICS FOR MANAGING PESTICIDE RESISTANCE IN ARTHROPODS: Theory and Practice*. www.annualreviews.org (March 12, 2020).

Dermauw, Wannes et al. 2013. "A Link between Host Plant Adaptation and Pesticide Resistance in the Polyphagous Spider Mite Tetranychus Urticae." *Proceedings of the National Academy of Sciences of the United States of America* 110(2): E113–22.

———. 2018. "Does Host Plant Adaptation Lead to Pesticide Resistance in Generalist Herbivores?" *Current Opinion in Insect Science* 26: 25–33.

Drummond, Alexei J., Marc A. Suchard, Dong Xie, and Andrew Rambaut. 2012. "Bayesian Phylogenetics with BEAUti and the BEAST 1.7." *Molecular Biology and Evolution* 29(8): 1969–73. https://academic.oup.com/mbe/article-abstract/29/8/1969/1044583 (September 8, 2021).

Dunlop, Judy et al. 2021. "Industry Environmental Offset Funding Facilitates a Large Multi-Species Fauna Translocation Program." *Pacific Conservation Biology* 102: 98–133. https://www.publish.csiro.au/pc/PC20036 (May 10, 2022).

Eason, Charles. 2002. "Sodium Monofluoroacetate (1080) Risk Assessment and Risk Communication." *Toxicology* 181–182: 523–30.

Eason, Charles, Aroha Miller, Shaun Ogilvie, and Alastair Fairweather. 2011. "An Updated Review of the Toxicology and Ecotoxicology of Sodium Fluoroacetate (1080) in Relation to Its Use as a Pest Control Tool in New Zealand." *New Zealand Journal of Ecology* 35: 1–20. https://www.jstor.org/stable/24060627 (March 13, 2020).

Edger, Patrick P. et al. 2015. "The Butterfly Plant Arms-Race Escalated by Gene and Genome Duplications." *Proceedings of the National Academy of Sciences of the United States of America* 112(27): 8362–66. https://www.pnas.org/doi/abs/10.1073/pnas.1503926112 (May 19, 2023).

Efford, Murray, Bruce Warburton, and Nick Spencer. 2000. "Home-Range Changes by Brushtail Possums in Response to Control." *Wildlife Research* 27(2): 117–27. http://www.publish.csiro.au/WR/WR99005 (March 15, 2020).

Ehrlich, Paul R., and Peter H. Raven. 1964. "Butterflies and Plants: A Study in Coevolution." *Evolution* 18(4): 586. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1558-5646.1964.tb01674.x (March 15, 2020).

El-Merhibi, Adaweyah et al. 2011. "Cytochrome P450 CYP3A in Marsupials: Cloning and Characterisation of the Second Identified CYP3A Subfamily Member, Isoform 3A78 from Koala (Phascolarctos Cinereus)." *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* 154(4): 367–76.

Elena, Santiago F., and Richard E. Lenski. 2003. "Evolution Experiments with Microorganisms: The Dynamics and Genetic Bases of Adaptation." *Nature Reviews Genetics* 4(6): 457–69.

Endara, María José et al. 2017. "Coevolutionary Arms Race versus Host Defense Chase in a Tropical Herbivore–Plant System." *Proceedings of the National Academy of Sciences of the United States of America* 114(36): E7499–7505.

Endler, John A. 1977. "Geographic Variation, Speciation, and Clines." *Monographs in population biology* 10: 1–246.

Faith, J. Tyler et al. 2017. "Large Mammal Species Richness and Late Quaternary Precipitation Change in South-Western Australia." *Journal of Quaternary Science* 32(6): 760–69. https://onlinelibrary.wiley.com/doi/full/10.1002/jqs.2888 (October 1, 2021).

Felsenstein, J. 1985. "Phylogenies and the Comparative Method." *American Naturalist* 125(1): 1–15.

Felsenstein, Joseph, and Hirohisa Kishino. 1993. "Is There Something Wrong with the

Bootstrap on Phylogenies? A Reply to Hillis and Bull." *Systematic Biology* 42(2): 193.

Fick, SE, RJ Hijmans - International journal of climatology, and undefined 2017. 2017. "WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas." *Wiley Online Library* 37(12): 4302–15. https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5086?casa_token=bTtWGlmoSeAAAAAA:M50yU39CG-YMgytWWhb-H7bD5RMri3sszTwKqCTWDbLda8xgwuUYj1tBKsMj952jSGuzlAbdlOYuafBF (September 8, 2021).

Firman, Renée C. et al. 2020. "Extreme and Variable Climatic Conditions Drive the Evolution of Sociality in Australian Rodents." *Current Biology* 30(4): 691-697.e3.

Flannery, T F, W D Turnbull, T H V Rich, and E L Lundelius. 1987. "The Phalangerids (Marsupialia: Phalangeridae) of the Early Pliocene Hamilton Local Fauna, Southwestern Victoria." *Possums and opossums: studies in evolution. Surrey Beatty and Sons and the Royal Zoological Society of New South Wales, Sydney*: 537–46.

Georghiou, George P., and Roni B. Mellon. 1983. "Pesticide Resistance in Time and Space." In *Pest Resistance to Pesticides*, Springer US, 1–46.

Griffitts, Joel S., and Raffi V. Aroian. 2005. "Many Roads to Resistance: How Invertebrates Adapt to Bt Toxins." *BioEssays* 27(6): 614–24. https://onlinelibrary.wiley.com/doi/full/10.1002/bies.20239 (May 9, 2023).

Gupta, RC. 2015. "Handbook of Toxicology of Chemical Warfare Agents." https://books.google.com/books?hl=fr&lr=&id=MXKDBAAAQBAJ&oi=fnd&pg=PP1&dq=Goncharov+N,+et+al.+2015+Fluoracetate.+Chapter+16,+pages+193–214+in+Gupta+RC+(ed)+Handbook+of+Toxicology+of+Chemical+Warfare+Agents+(2nd+edition).+Academic+Press.+978-0-12-800159-2&ots=L0zowzfqi3&sig=zkRzHn12-_2tCGiRS4zNQ2OIWOM (March 13, 2020).

Harris, M. A. et al. 2004. "The Gene Oncology (GO) Database and Informatics Resource." *Nucleic Acids Research* 32(DATABASE ISS.): D258–61. https://academic.oup.com/nar/article/32/suppl_1/D258/2505186 (May 3, 2023).

Hart, Steven N. et al. 2013. "Calculating Sample Size Estimates for RNA Sequencing Data." *Journal of Computational Biology* 20(12): 970–78. https://pubmed.ncbi.nlm.nih.gov/23961961/ (July 5, 2022).

Hawkins, Nichola J., Chris Bass, Andrea Dixon, and Paul Neve. 2019. "The Evolutionary Origins of Pesticide Resistance." *Biological Reviews* 94(1): 135–55. http://doi.wiley.com/10.1111/brv.12440 (March 15, 2020).

He, Yuxin, and Huanye Wang. 2021. "Terrestrial Material Input to the Northwest Shelf of Australia Through the Pliocene-Pleistocene Period and Its Implications on Continental Climates." *Geophysical Research Letters* 48(17): e2021GL092745. https://onlinelibrary.wiley.com/doi/full/10.1029/2021GL092745 (May 30, 2022).

Helfrich, Philipp et al. 2019. "Treeannotator: Versatile Visual Annotation of Hierarchical Text Relations." In *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, , 1958–63. https://www.aclweb.org/anthology/L18-1308.pdf (September 8, 2021).

Hewitt, Godfrey M. 2004. "Genetic Consequences of Climatic Oscillations in the Quaternary." In *Philosophical Transactions of the Royal Society B: Biological Sciences*, The Royal Society, 183–95. /pmc/articles/PMC1693318/?report=abstract (May 30, 2022).

Hijmans, RJ et al. 2017. "Package 'Dismo.'" *gr.xemacs.org*. https://scholar.google.com/ftp://ftp.gr.xemacs.org/mirrors/CRAN/web/packages/dismo/dismo.pdf (March 18, 2022).

Hocknull, Scott A. et al. 2020. "Extinction of Eastern Sahul Megafauna Coincides with Sustained Environmental Deterioration." *Nature Communications* 11(1): 1–14. https://www.nature.com/articles/s41467-020-15785-w (March 9, 2022).

How, R. A., and S. J. Hillcox. 2000. "Brushtail Possum, Trichosurus Vulpecula, Populations in South-Western Australia: Demography, Diet and Conservation Status." *Wildlife Research* 27(1): 81–89. https://www.publish.csiro.au/wr/wr98064 (July 5, 2022).

Huelsenbeck, John P., and David M. Hillis. 1993. "Success of Phylogenetic Methods in the Four-Taxon Case." *Systematic Biology* 42(3): 247–64. https://academic.oup.com/sysbio/article/42/3/247/1629468 (September 30, 2021).

Jackson, R. et al. 2019. "Ecology of a Brushtail Possum (Trichosurus Vulpecula) Population at Castlepoint in the Wairarapa, New Zealand." *New Zealand Journal of Ecology* 43(2).

Jalili, Vahid et al. 2021. "The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2020 Update." *Nucleic Acids Research* 48(W1): W395–402. https://academic.oup.com/nar/article/48/W1/W395/5849904 (July 4, 2022).

James, C. H., and C. Moritz. 2000. "Intraspecific Phylogeography in the Sedge Frog Litoria Fallax (Hylidae) Indicates Pre-Pleistocene Vicariance of an Open Forest Species from Eastern Australia." *Molecular Ecology* 9(3): 349–58. https://onlinelibrary.wiley.com/doi/full/10.1046/j.1365-294x.2000.00885.x (January 10, 2022).

Jarman, Peter. 2006. "Life of Marsupials." *Austral Ecology* 31(5): 670–670.

https://books.google.com/books?hl=fr&lr=&id=gDFeNMhJlPUC&oi=fnd&pg=PP7&dq=life+of+marsupials&ots=e8g6ILivtt&sig=CnZRgadgx3IxwwmrYAZcxgptNlU (July 5, 2022).

Jombart, Thibaut. 2008. "Adegenet: A R Package for the Multivariate Analysis of Genetic Markers." *Bioinformatics* 24(11): 1403–5.

Jombart, Thibaut, Sébastien Devillard, and François Balloux. 2010. "Discriminant Analysis of Principal Components: A New Method for the Analysis of Genetically Structured Populations." *BMC Genetics* 11(1): 1–15. https://bmcgenomdata.biomedcentral.com/articles/10.1186/1471-2156-11-94 (February 16, 2022).

Kanehisa, Minoru, and Susumu Goto. 2000. "KEGG: Kyoto Encyclopedia of Genes and Genomes." *Nucleic Acids Research* 28(1): 27–30. https://academic.oup.com/nar/article/28/1/27/2384332 (May 3, 2023).

Karageorgi, Marianthi et al. 2019. "Genome Editing Retraces the Evolution of Toxin Resistance in the Monarch Butterfly." *Nature* 574(7778): 409–12. https://pubmed.ncbi.nlm.nih.gov/31578524/ (May 9, 2023).

Kearse, M et al. 2012. "Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data." *Bioinformatics* 28(12): 1647–49. https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bts199.

Kerle, A., G. M. McKay, and G. B. Sharman. 1991. "A Systematic Analysis of the Brushtail Possum, Trichosurus Vulpecula (Kerr, 1792) (Marsupialia: Phalangeridae)." *Australian Journal of Zoology* 39(3): 263–71. https://www.publish.csiro.au/zo/zo9910313 (November 23, 2021).

Kerle, A. 1984. "Variation in the Ecology of Trichosurus: Its Adaptive Significance." *Possums and gliders*: 115–28.

———. 2002. "Possums: The Brushtails, Ringtails and Greater Glider." *Choice Reviews Online* 39(10): 39-5816-39–5816. https://books.google.com/books?hl=fr&lr=&id=YDM0hjAwchUC&oi=fnd&pg=PR7&dq=opossum+Kerle&ots=4gYdItzt2O&sig=T5sZ2OXT-lr669669KDdfxjw7G4 (September 23, 2022).

Kerle, A, and RA How. 2008. "Common Brushtail Possum." In *The Mammals of Australia*, , 274–76.

Kim, Daehwan et al. 2019. "Graph-Based Genome Alignment and Genotyping with HISAT2

and HISAT-Genotype." *Nature Biotechnology 2019 37:8* 37(8): 907–15. https://www.nature.com/articles/s41587-019-0201-4 (July 4, 2022).

Kimura, Motoo. 1977. "Preponderance of Synonymous Changes as Evidence for the Neutral Theory of Molecular Evolution [33]." *Nature* 267(5608): 275–76. https://www.nature.com/articles/267275a0 (April 24, 2023).

King, D. R., A. J. Oliver, and R. J. Mead. 1978. "The Adaptation of Some Western Australian Mammals to Food Plants Containing Fluoroacetate." *Australian Journal of Zoology* 26(4): 699–712.

Koot, Emily M., Mary Morgan-Richards, and Steven A. Trewick. 2022. "Climate Change and Alpine-Adapted Insects: Modelling Environmental Envelopes of a Grasshopper Radiation." *Royal Society Open Science* 9(3). https://royalsocietypublishing.org/doi/abs/10.1098/rsos.211596 (March 23, 2022).

Korthauer, Keegan et al. 2019. "A Practical Guide to Methods Controlling False Discoveries in Computational Biology." *Genome Biology* 20(1): 1–21. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1716-1 (August 16, 2022).

Kozulin, Peter et al. 2022. "Divergent Evolution of Developmental Timing in the Neocortex Revealed by Marsupial and Eutherian Transcriptomes." *Development (Cambridge)* 149(3). https://journals.biologists.com/dev/article/149/3/dev200212/274319/Divergent-evolution-of-developmental-timing-in-the (July 5, 2022).

Kruskal, Joseph B. 1956. "On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem." *Proceedings of the American Mathematical Society* 7(1): 48–50. https://www.ams.org/proc/1956-007-01/S0002-9939-1956-0078686-7/ (February 16, 2022).

Kuch, Ulrich et al. 2005. "Phylogeography of Australia's King Brown Snake (Pseudechis Australis) Reveals Pliocene Divergence and Pleistocene Dispersal of a Top Predator." *Naturwissenschaften* 92(3): 121–27. https://link.springer.com/article/10.1007/s00114-004-0602-0 (October 5, 2022).

Kung, Janet W.C., Ian S. Currie, Stuart J. Forbes, and James A. Ross. 2010. "Liver Development, Regeneration, and Carcinogenesis." *Journal of Biomedicine and Biotechnology* 2010.

Lamarre, Sophie et al. 2018a. "Optimization of an RNA-Seq Differential Gene Expression Analysis Depending on Biological Replicate Number and Library Size." *Frontiers in Plant Science* 9: 108.

———. 2018b. “Optimization of an RNA-Seq Differential Gene Expression Analysis Depending on Biological Replicate Number and Library Size.” *Frontiers in Plant Science* 9: 108.

Lanfear, Robert et al. 2017. “Partitionfinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses.” *Molecular Biology and Evolution* 34(3): 772–73. https://academic.oup.com/mbe/article-abstract/34/3/772/2738784 (March 16, 2020).

Langfelder, Peter, and Steve Horvath. 2008. “WGCNA: An R Package for Weighted Correlation Network Analysis.” *BMC Bioinformatics* 9(1): 1–13. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559 (December 19, 2022).

Larsson, Johan, and Peter Gustafsson. 2018. “{eulerr}: Area-Proportional Euler and Venn Diagrams with Ellipses.” *Proceedings of {{International Workshop}} on {{Set Visualization}} and {{Reasoning}}* 2116: 84–91. https://cran.r-project.org/package=eulerr (December 19, 2022).

Leigh, Jessica W, and David Bryant. 2015. “Popart: Full-Feature Software for Haplotype Network Construction.” *Methods in Ecology and Evolution* 6(9): 1110–16. https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12410 (June 17, 2019).

Lenski, Richard E. 2017. “Convergence and Divergence in a Long-Term Experiment with Bacteria.” *The American naturalist* 190(S1): S57–68. http://www.ncbi.nlm.nih.gov/pubmed/28731830 (March 12, 2020).

Leong, Lex Ee Xiang et al. 2017. “Fluoroacetate in Plants - a Review of Its Distribution, Toxicity to Livestock and Microbial Detoxification.” *Journal of Animal Science and Biotechnology* 8(1): 55. http://jasbsci.biomedcentral.com/articles/10.1186/s40104-017-0180-6 (March 13, 2020).

Li, Tingting et al. 2009. “Multi-Stage Analysis of Gene Expression and Transcription Regulation in C57/B6 Mouse Liver Development.” *Genomics* 93(3): 235–42.

Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. “FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features.” *Bioinformatics* 30(7): 923–30. https://academic.oup.com/bioinformatics/article/30/7/923/232889 (July 4, 2022).

Livingstone, P. G., N. Hancox, G. Nugent, and G. W. de Lisle. 2015. “Toward Eradication: The Effect of Mycobacterium Bovis Infection in Wildlife on the Evolution and Future

Direction of Bovine Tuberculosis Management in New Zealand." *New Zealand Veterinary Journal* 63: 4–18.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15(12): 1–21. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8 (July 4, 2022).

Lynch, A. Jasmyn J., Robert J.S. Beeton, and Penelope Greenslade. 2019. "The Conservation Significance of the Biota of Barrow Island, Western Australia." *Journal of the Royal Society of Western Australia* 102: 98–133. https://www.researchgate.net/profile/A-Lynch/publication/336855839_The_conservation_significance_of_the_biota_of_Barrow_Island_Western_Australia/links/5dc261f94585151435ec6ff5/The-conservation-significance-of-the-biota-of-Barrow-Island-Western-Australia.p (May 10, 2022).

Maclean, Craig. 2010. "The Population Genetics of Antibiotic Resistance: Integrating Molecular Mechanisms and Treatment Contexts." *nature.com*. https://www.researchgate.net/publication/44608740 (March 15, 2020).

Macnair, M. R. 1991. "Why the Evolution of Resistance to Anthropogenic Toxins Normally Involves Major Gene Changes: The Limits to Natural Selection." *Genetica* 84(3): 213–19.

Macqueen, Peggy et al. 2010. "Phylogenetics of the Pademelons (Macropodidae: Thylogale) and Historical Biogeography of the Australo-Papuan Region." *Molecular Phylogenetics and Evolution* 57(3): 1134–48.

Mallet, James. 2008. "Hybridization, Ecological Races and the Nature of Species: Empirical Evidence for the Ease of Speciation." *Philosophical Transactions of the Royal Society B: Biological Sciences* 363(1506): 2971–86. https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2008.0081 (March 14, 2022).

Marioni, John C. et al. 2008. "RNA-Seq: An Assessment of Technical Reproducibility and Comparison with Gene Expression Arrays." *Genome Research* 18(9): 1509–17. https://genome.cshlp.org/content/18/9/1509.full (July 12, 2022).

Martin, H. A. 2006. "Cenozoic Climatic Change and the Development of the Arid Vegetation in Australia." *Journal of Arid Environments* 66(3 SPEC. ISS.): 533–63.

Martin, H A. 1998. "Tertiary Climatic Evolution and the Development of Aridity in Australia." In *Proceedings-Linnean Society Of New South Wales*, , 115–36.

Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." *EMBnet.journal* 17(1): 10.

Mcgowran, Brian, and Phil Bock. 2000. "Article in Memoirs of the Association of Australasian Palaeontologists." https://www.researchgate.net/publication/292144424 (October 1, 2021).

McIlroy, J. C. 1983. "The Sensitivity of the Brushtail Possum (Trichosurus Vulpecula) to 1080 Poison (Sodium Monfluoroacetate)." *New Zealand Journal of Ecology* 6: 125–31. https://www.jstor.org/stable/24052734 (March 15, 2020).

Mead, R. J., A. J. Oliver, D. R. King, and P. H. Hubach. 1985. "The Co-Evolutionary Role of Fluoroacetate in Plant-Animal Interactions in Australia." *Oikos* 44(1): 55.

Meredith, Robert W., Michael Westerman, and Mark S. Springer. 2009. "A Phylogeny of Diprotodontia (Marsupialia) Based on Sequences for Five Nuclear Genes." *Molecular Phylogenetics and Evolution* 51(3): 554–71. https://www.sciencedirect.com/science/article/pii/S1055790309000414 (June 24, 2021).

Meredith, Robert W, Michael Westerman, Judd A Case, and Mark S Springer. 2008. "A Phylogeny and Timescale for Marsupial Evolution Based on Sequences for Five Nuclear Genes." *Journal of Mammalian Evolution* 15(1): 1–36.

Miller, Gifford H. et al. 2005. "Anthropology: Ecosystem Collapse in Pleistocene Australia and a Human Role in Megafaunal Extinction." *Science* 309(5732): 287–90. https://www.science.org/doi/abs/10.1126/science.1111288 (March 9, 2022).

Mittermeier, Russell A. 2004. *Hotspots Revisited*. Cemex.

Modepalli, Vengamanaidu et al. 2018. "Gene Expression Profiling of Postnatal Lung Development in the Marsupial Gray Short-Tailed Opossum (Monodelphis Domestica) Highlights Conserved Developmental Pathways and Specific Characteristics during Lung Organogenesis." *BMC Genomics* 19(1): 732. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6173930/ (August 18, 2019).

Montague, T. L. (Thomas L.). 2000. *The Brushtail Possum : Biology, Impact and Management of an Introduced Marsupial*. https://trove.nla.gov.au/work/32729533?q&versionId=46367736 (March 15, 2020).

Müllner, Daniel. 2013. "Fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python." *Journal of Statistical Software* 53(9): 1–18. https://cran.r-project.org/package=fastcluster (December 19, 2022).

Myers, Norman et al. 2000. "Biodiversity Hotspots for Conservation Priorities." *Nature* 403(6772): 853–58. https://www.nature.com/articles/35002501 (November 8, 2021).

Naimi, Babak et al. 2014. "Where Is Positional Uncertainty a Problem for Species Distribution Modelling?" *Ecography* 37(2): 191–203.

https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0587.2013.00205.x (March 18, 2022).

Neaves, Linda E. et al. 2016. "Phylogeography of the Koala, (Phascolarctos Cinereus), and Harmonising Data to Inform Conservation." *PLoS ONE* 11(9): e0162207. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0162207 (February 11, 2022).

Neher, R. A., B. I. Shraiman, and D. S. Fisher. 2010. "Rate of Adaptation in Large Sexual Populations." *Genetics* 184(2): 467–81. https://academic.oup.com/genetics/article/184/2/467/6077864 (May 30, 2022).

Neve, Paul, Roberto Busi, Michael Renton, and Martin M. Vila-Aiub. 2014. "Expanding the Eco-Evolutionary Context of Herbicide Resistance Research." *Pest Management Science* 70(9): 1385–93.

New, Tim R. 2011. "'In Considerable Variety': Introducing the Diversity of Australia's Insects." *'In Considerable Variety': Introducing the Diversity of Australia's Insects*.

Nogués-Bravo, David et al. 2018. "Cracking the Code of Biodiversity Responses to Past Climate Change." *Trends in Ecology and Evolution* 33(10): 765–76. https://pubmed.ncbi.nlm.nih.gov/30173951/ (March 23, 2022).

Nugent, G., P. Sweetapple, J. Coleman, and P. Suisted. 2000. "Possum Feeding Patterns: Dietary Tactics of a Reluctant Folivore." In *The Brushtail Possum: Biology, Impact and Management of an Introduced Marsupial*, , 10–19.

O'Reilly-Wapstra, Julianne M., and Phil Cowan. 2010. "Native Plant/Herbivore Interactions as Determinants of the Ecological and Evolutionary Effects of Invasive Mammalian Herbivores: The Case of the Common Brushtail Possum." *Biological Invasions* 12(2): 373–87.

Ogilvie, S C et al. 2009. *Uptake of 1080 by Watercress and Puha-Culturally-+lmportant Plants Used for Food By*. http://dspace.lincoln.ac.nz/handle/10182/1389 (March 13, 2020).

Oliver, A. J., and D. R. King. 1979. "Fluoroacetate Tolerance, a Genetic Marker in Some Australian Mammals." *Australian Journal of Zoology* 27(3): 331–47. https://www.publish.csiro.au/zo/zo9790363 (April 22, 2022).

Oliver, Paul M. et al. 2017. "A Novel Hotspot of Vertebrate Endemism and an Evolutionary Refugium in Tropical Australia." *Diversity and Distributions* 23(1): 53–66. https://onlinelibrary.wiley.com/doi/full/10.1111/ddi.12506 (October 5, 2022).

Oliver, Paul M., Mark Adams, and Paul Doughty. 2010. "Molecular Evidence for Ten Species and Oligo-Miocene Vicariance within a Nominal Australian Gecko Species

(Crenadactylus Ocellatus, Diplodactylidae).” *BMC Evolutionary Biology* 10(1): 1–11. https://bmcecolevol.biomedcentral.com/articles/10.1186/1471-2148-10-386 (October 5, 2022).

Orzech, Kathryn M, and Mark Nichter. 2008. “From Resilience to Resistance: Political Ecological Lessons from Antibiotic and Pesticide Resistance.” www.annualreviews.org (March 12, 2020).

Palumbi, S. R. 2001. “Humans as the World’s Greatest Evolutionary Force.” *Science* 293(5536): 1786–90.

Paradis, Emmanuel. 2010. “Pegas: An R Package for Population Genetics with an Integrated–Modular Approach.” *Bioinformatics* 26(3): 419–20. https://academic.oup.com/bioinformatics/article/26/3/419/215731 (March 26, 2019).

Paranjpe, Monika et al. 2019. “Transcriptomic Analysis of MAP3K1 and MAP3K4 in the Developing Marsupial Gonad.” *Sexual Development* 13(4): 195–204. https://www.karger.com/Article/FullText/505799 (July 5, 2022).

Pattabiraman, Nimeshika, Mary Morgan-Richards, Ralph Powlesland, and Steven A. Trewick. 2021. “Unrestricted Gene Flow between Two Subspecies of Translocated Brushtail Possums (Trichosurus Vulpecula) in Aotearoa New Zealand.” *Biological Invasions* 24(1): 247–60. https://link.springer.com/article/10.1007/s10530-021-02635-z (May 31, 2022).

Peng, Lai et al. 2012. “RNA Sequencing Reveals Dynamic Changes of MRNA Abundance of Cytochromes P450 and Their Alternative Transcripts during Mouse Liver Development.” *Drug Metabolism and Disposition* 40(6): 1198–1209. https://pubmed.ncbi.nlm.nih.gov/22434873/ (December 21, 2022).

Perez, John C., Sathit Pichyangkul, and Vivian E. Garcia. 1979. “The Resistance of Three Species of Warm-Blooded Animals to Western Diamondback Rattlesnake (Crotalus Atrox) Venom.” *Toxicon* 17(6): 601–7. https://pubmed.ncbi.nlm.nih.gov/524385/ (May 2, 2023).

Peterson, B., C. C. Bezuidenhout, and J. Van den Berg. 2017. “An Overview of Mechanisms of Cry Toxin Resistance in Lepidopteran Insects.” *Journal of Economic Entomology* 110(2): 362–77. https://pubmed.ncbi.nlm.nih.gov/28334065/ (May 10, 2023).

Phillips, M. J., Y. H. Lin, G. L. Harrison, and D. Penny. 2001. “Mitochondrial Genomes of a Bandicoot and a Brushtail Possum Confirm the Monophyly of Australidelphian Marsupials.” *Proceedings of the Royal Society B: Biological Sciences* 268(1475): 1533–38.

Poran, Naomie S., Richard G. Coss, and Eli Benjamini. 1987. “Resistance of California Ground

Squirrels (Spermophilus Beecheyi) to the Venom of the Northern Pacific Rattlesnake (Crotalus Viridis Oreganus): A Study of Adaptive Variation." *Toxicon* 25(7): 767–77. https://pubmed.ncbi.nlm.nih.gov/3672545/ (May 2, 2023).

Porder, Stephen. 2014. "Coevolution of Life and Landscapes." *Proceedings of the National Academy of Sciences of the United States of America* 111(9): 3207–8. www.pnas.org/cgi/doi/10.1073/pnas.1400954111 (March 9, 2022).

Potter, Sally et al. 2012. "Phylogenetic Relationships of Rock-Wallabies, Petrogale (Marsupialia: Macropodidae) and Their Biogeographic History within Australia." *Molecular Phylogenetics and Evolution* 62(2): 640–52.

Powles, Stephen B., and Qin Yu. 2010. "Evolution in Action: Plants Resistant to Herbicides." *Annual Review of Plant Biology* 61(1): 317–47.

Pracy, L. 1974. "Opposums." *New Zealand Nature Heritage* 3(32): 873–82.

Prideaux, Gavin J. et al. 2007. "Mammalian Responses to Pleistocene Climate Change in Southeastern Australia." *Geology* 35(1): 33–36. http://pubs.geoscienceworld.org/gsa/geology/article-pdf/35/1/33/3532085/i0091-7613-35-1-33.pdf (February 16, 2022).

R Development Core Team. 2021. "R: A Language and Environment for Statistical Computing (Version 4.1. 2)[Computer Software]." *R Foundation for Statistical Computing*.

Rambaut, Suchard, Xie, and A Drummond. 2015. "Tracer v1.7." http://tree.bio.ed.ac.uk/software/tracer/ (March 16, 2020).

Reid, Noah M., and Bryan C. Carstens. 2012. "Phylogenetic Estimation Error Can Decrease the Accuracy of Species Delimitation: A Bayesian Implementation of the General Mixed Yule-Coalescent Model." *BMC Evolutionary Biology* 12(1): 1–11. https://bmcecolevol.biomedcentral.com/articles/10.1186/1471-2148-12-196 (February 11, 2022).

Reznick, David N., and Cameron K. Ghalambor. 2001. "The Population Ecology of Contemporary Adaptations: What Empirical Studies Reveal about the Conditions That Promote Adaptive Evolution." *Genetica* 112–113(1): 183–98. https://link.springer.com/article/10.1023/A:1013352109042 (May 30, 2022).

Ritchie, Matthew E. et al. 2015. "Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies." *Nucleic Acids Research* 43(7): e47. /pmc/articles/PMC4402510/ (December 19, 2022).

Rix, Michael G et al. 2015. "Biogeography and Speciation of Terrestrial Fauna in the South-Western Australian Biodiversity Hotspot." *Biological Reviews* 90(3): 762–93.

Robbins, Robert K. et al. 2021. "A Switch to Feeding on Cycads Generates Parallel Accelerated Evolution of Toxin Tolerance in Two Clades of Eumaeus Caterpillars (Lepidoptera: Lycaenidae)." *Proceedings of the National Academy of Sciences of the United States of America* 118(7): e2018965118. https://www.pnas.org/doi/abs/10.1073/pnas.2018965118 (May 2, 2023).

Roberts, Karen Kristine, Mina Bassarova, and Michael Archer. 2008. "Oligo-Miocene Ringtail Possums of the Genus Paljara (Pseudocheiridae: Marsupialia) from Queensland, Australia." *Geobios* 41(6): 833–44. https://www.sciencedirect.com/science/article/pii/S0016699508000806 (September 30, 2021).

Roberts, R. G. et al. 2001. "New Ages for the Last Australian Megafauna: Continent-Wide Extinction about 46,000 Years Ago." *Science* 292(5523): 1888–92. https://www.science.org/doi/abs/10.1126/science.1060264 (March 9, 2022).

Robinson, Weldon B. 1953. "Coyote Control with Compound 1080 Stations in National Forests." *Journal of Forestry* 51(12): 880–85.

Robles, José A. et al. 2012. "Efficient Experimental Design and Analysis Strategies for the Detection of Differential Expression Using RNA-Sequencing." *BMC Genomics* 13(1): 1–14. https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-13-484 (July 4, 2022).

Ronquist, Fredrik, and John P Huelsenbeck. 2003. "MrBayes 3: Bayesian Phylogenetic Inference under Mixed Models." *Bioinformatics* 19(12): 1572–74. https://academic.oup.com/bioinformatics/article-abstract/19/12/1572/257621 (March 16, 2020).

Rosenheim, Jay A. et al. 1996. "Biochemical Peradaptations, Founder Events, and the Evolution of Resistance in Arthropods." *Journal of Economic Entomology* 89(2): 263–73.

Ross, James G., Kathryn Bicknell, and G.J. Hickling. 1999. "COST-EFFECTIVE CONTROL OF 1080 BAIT-SHY POSSUMS."

Roush, R, and BE Tabashnik. 2012. "Pesticide Resistance in Arthropods." https://books.google.com/books?hl=fr&lr=&id=ooHjBwAAQBAJ&oi=fnd&pg=PA7&dq=pesticide+resistance+review+&ots=b4HTRVe3DT&sig=gatc_jXjdzJN2icrjAZpQkbtNgo (March 12, 2020).

Rowe, Ashlee H., and Matthew P. Rowe. 2008. "Physiological Resistance of Grasshopper Mice (Onychomys Spp.) to Arizona Bark Scorpion (Centruroides Exilicauda) Venom." *Toxicon* 52(5): 597–605. https://pubmed.ncbi.nlm.nih.gov/18687353/ (May 2, 2023).

Rowe, Kevin C. et al. 2008. “Pliocene Colonization and Adaptive Radiations in Australia and New Guinea (Sahul): Multilocus Systematics of the Old Endemic Rodents (Muroidea: Murinae).” *Molecular Phylogenetics and Evolution* 47(1): 84–101.

Saltré, Frédérik et al. 2016. “Climate Change Not to Blame for Late Quaternary Megafauna Extinctions in Australia.” *Nature Communications* 7(1): 1–7. https://www.nature.com/articles/ncomms10511 (March 9, 2022).

Scheuerl, Thomas, Johannes Cairns, Lutz Becks, and Teppo Hiltunen. 2019. “Predator Coevolution and Prey Trait Variability Determine Species Coexistence.” *Proceedings of the Royal Society B: Biological Sciences* 286(1902).

Seliskar, Matej, and Damjana Rozman. 2007. “Mammalian Cytochromes P450-Importance of Tissue Specificity.” *Biochimica et Biophysica Acta - General Subjects* 1770(3): 458–66. https://pubmed.ncbi.nlm.nih.gov/17097232/ (July 5, 2022).

Seyednasrollah, Fatemeh, Asta Laiho, and Laura L. Elo. 2013. “Comparison of Software Packages for Detecting Differential Expression in RNA-Seq Studies.” *Briefings in Bioinformatics* 16(1): 59–70. https://academic.oup.com/bib/article/16/1/59/240754 (December 19, 2022).

Shiao, Yih-Horng. 2021. *Discoveries and Biological Implications of Mammalian 45S RDNA Variants and Non-Structural RRNAs*. Preprints. https://www.preprints.org/manuscript/202102.0613/v1 (February 11, 2022).

Smith, Kathleen K., and Anna L. Keyte. 2020. “Adaptations of the Marsupial Newborn: Birth as an Extreme Environment.” *Anatomical Record* 303(2): 235–49. https://onlinelibrary.wiley.com/doi/full/10.1002/ar.24049 (March 6, 2023).

Smoot, Michael E. et al. 2011. “Cytoscape 2.8: New Features for Data Integration and Network Visualization.” *Bioinformatics* 27(3): 431–32. https://academic.oup.com/bioinformatics/article/27/3/431/321742 (July 4, 2022).

Sniderman, J. M. K. et al. 2016. “Pliocene Reversal of Late Neogene Aridification.” *Proceedings of the National Academy of Sciences of the United States of America* 113(8): 1999–2004. www.pnas.org/cgi/doi/10.1073/pnas.1520188113 (May 30, 2022).

Sniderman, J. M.K. et al. 2019. “Vegetation and Climate Change in Southwestern Australia During the Last Glacial Maximum.” *Geophysical Research Letters* 46(3): 1709–20. https://onlinelibrary.wiley.com/doi/full/10.1029/2018GL080832 (October 1, 2021).

Sochorová, Jana et al. 2021. “Analyses of the Updated ‘Animal Rdna Loci Database’ with an Emphasis on Its New Features.” *International Journal of Molecular Sciences* 22(21): 11403. https://www.mdpi.com/1422-0067/22/21/11403/htm (February 11, 2022).

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30(9): 1312–13. https://academic.oup.com/bioinformatics/article-abstract/30/9/1312/238053 (March 16, 2020).

Steel, Mike, and Andy McKenzie. 2001. "Properties of Phylogenetic Trees Generated by Yule-Type Speciation Models." *Mathematical Biosciences* 170(1): 91–112.

Strasburg, Jared L., Michael Kearney, Craig Moritz, and Alan R. Templeton. 2007. "Combining Phylogeography with Distribution Modeling: Multiple Pleistocene Range Expansions in a Parthenogenetic Gecko from the Australian Arid Zone." *PLoS ONE* 2(8): e760. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000760 (October 5, 2022).

Szücs, M. et al. 2017. "Rapid Adaptive Evolution in Novel Environments Acts as an Architect of Population Range Expansion." *Proceedings of the National Academy of Sciences of the United States of America* 114(51): 13501–6. www.pnas.org/cgi/doi/10.1073/pnas.1712934114 (May 30, 2022).

Tao, Wenjing et al. 2018. "Transcriptome Display during Tilapia Sex Determination and Differentiation as Revealed by RNA-Seq Analysis." *BMC Genomics* 19(1): 1–12. https://link.springer.com/articles/10.1186/s12864-018-4756-0 (July 5, 2022).

Taylor, Andrea C. et al. 2004. "High Microsatellite Diversity and Differential Structuring among Populations of the Introduced Common Brushtail Possum, Trichosurus Vulpecula, in New Zealand." *Genetical Research* 83(2): 101–11.

van Thiel, Jory et al. 2022. "Convergent Evolution of Toxin Resistance in Animals." *Biological Reviews* 97(5): 1823–43. https://onlinelibrary.wiley.com/doi/full/10.1111/brv.12865 (May 1, 2023).

Thompson, Jim A., and Peter J.S. Fleming. 1994. "Evaluation of the Efficacy of 1080 Poisoning of Red Foxes Using Visitation to Non-Toxic Baits as an Index of Fox Abundance." *Wildlife Research* 21(1): 91–104.

Thuiller, Wilfried et al. 2016. "Package 'Biomod2.'" *Species distribution modeling within an ensemble forecasting framework.*

Thuiller, Wilfried, Damien Georges, and Robin Engler. 2015. "Package 'Biomod2': Ensemble Platform for Species Distribution Modeling."

Todd, Erica V., Michael A. Black, and Neil J. Gemmell. 2016. "The Power and Promise of RNA-Seq in Ecology and Evolution." *Molecular Ecology* 25(6): 1224–41. https://onlinelibrary.wiley.com/doi/full/10.1111/mec.13526 (July 12, 2022).

Trelstad, R. L. 1973. "The Developmental Biology of Vertebrate Collagens." *Journal of Histochemistry and Cytochemistry* 21(6): 521–28.

Triggs, S. J., and W. Q. Green. 1989. "Geographic Patterns of Genetic Variation in Brushtail Possums Trichosurus Vulpecula and Implications for Pest Control." *New Zealand Journal of Ecology* 12: 1–10. https://science.sciencemag.org/content/293/5536/1786.short?casa_token=_2rXnNJtJmwAAAAA:4ONmMBnglQOi7ZQqbP3A0z22dV7pzPhvDSFT_1h9wVi12Vjh7yP-rGi4ieBRnRrR7QzNRQWqjQBPp1g (March 13, 2020).

Turney, Chris S.M. et al. 2008. "Late-Surviving Megafauna in Tasmania, Australia, Implicate Human Involvement in Their Extinction." *Proceedings of the National Academy of Sciences of the United States of America* 105(34): 12150–53. www.pnas.org/cgi/content/full/ (March 9, 2022).

Twigg, L. E. et al. 1996. "Fluoroacetate Content of Some Species of the Toxic Australian Plant Genus, Gastrolobium, and Its Environmental Persistence." *Natural Toxins* 4(3): 122–27. http://doi.wiley.com/10.1002/19960403NT4 (March 15, 2020).

———. 2003. "Sensitivity of Some Australian Animals to Sodium Fluoroacetate (1080): Additional Species and Populations, and Some Ecological Considerations." *Australian Journal of Zoology* 51(5): 515–31. https://www.publish.csiro.au/zo/zo03040 (March 23, 2022).

Twigg, L. E., and D. R. King. 1991. "The Impact of Fluoroacetate-Bearing Vegetation on Native Australian Fauna: A Review." *Oikos* 61(3): 412. https://www.jstor.org/stable/3545249?casa_token=wxSjPb9PE-QAAAAA:AqA216Efy0uOSzZJwBTgcg7Eo64RP0a9Hmm8Ua_rf4Faa-Qp-yot16PIkohRXPufbktoxRP5zU0RpZ0KIa9YOsEXHe5Co014rR-2rRyJe0hgP43FRiJi (March 15, 2020).

Ujvari, Beata et al. 2015. "Widespread Convergence in Toxin Resistance by Predictable Molecular Evolution." *Proceedings of the National Academy of Sciences of the United States of America* 112(38): 11911–16. www.pnas.org/cgi/doi/10.1073/pnas.1511706112 (March 15, 2020).

Umbrello, Linette S., Raphael K. Didham, Ric A. How, and Joel A. Huey. 2020. "Multi-Species Phylogeography of Arid-Zone Sminthopsinae (Marsupialia: Dasyuridae) Reveals Evidence of Refugia and Population Expansion in Response to Quaternary Change." *Genes 2020, Vol. 11, Page 963* 11(9): 963. https://www.mdpi.com/2073-4425/11/9/963/htm (February 11, 2022).

Valavi, Roozbeh, Jane Elith, José J. Lahoz-Monfort, and Gurutzeta Guillera-Arroita. 2019. "BlockCV: An r Package for Generating Spatially or Environmentally Separated Folds for k-Fold Cross-Validation of Species Distribution Models." *Methods in Ecology and Evolution* 10(2): 225–32. https://www.biorxiv.org/content/10.1101/357798.full (September 8, 2021).

VanDerWal, Jeremy et al. 2014. "Package 'SDMTools.'" *R package*.

Vignali, S, AG Barras, … R Arlettaz - Ecology and, and undefined 2020. 2020. "SDMtune: An R Package to Tune and Evaluate Species Distribution Models." *Wiley Online Library* 10(20): 11488–506. https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.6786 (September 8, 2021).

Williams, Kristen J et al. 2011. "Forests of East Australia: The 35th Biodiversity Hotspot." In *Biodiversity Hotspots*, , 295–310.

de Wit, Cynthia A., and Björn R. Weström. 1987. "Venom Resistance in the Hedgehog, Erinaceus Europaeus: Purification and Identification of Macroglobulin Inhibitors as Plasma Antihemorrhagic Factors." *Toxicon* 25(3): 315–23. https://pubmed.ncbi.nlm.nih.gov/3590212/ (May 2, 2023).

Wood, R.J. 1981. "Insecticide Resistance: Genes and Mechanisms." *The Genetic Consequences of Man Made Change.*: Pp 53-96. http://agris.fao.org/agris-search/search.do?recordID=US201302634817 (March 12, 2020).

Worth, J. R.P. et al. 2017. "Habitat Type and Dispersal Mode Underlie the Capacity for Plant Migration across an Intermittent Seaway." *Annals of Botany* 120(4): 539–49. https://academic.oup.com/aob/article/120/4/539/4049543 (April 25, 2022).

Yang, Ziheng. 2007. "PAML 4: Phylogenetic Analysis by Maximum Likelihood." *Molecular Biology and Evolution* 24(8): 1586–91. https://academic.oup.com/mbe/article/24/8/1586/1103731 (March 10, 2023).

Yang, Ziheng, and Joseph R. Bielawski. 2000. "Statistical Methods for Detecting Molecular Adaptation." *Trends in Ecology and Evolution* 15(12): 496–503. https://pubmed.ncbi.nlm.nih.gov/11114436/ (May 3, 2023).

Yatsu, Ryohei et al. 2016. "RNA-Seq Analysis of the Gonadal Transcriptome during Alligator Mississippiensis Temperature-Dependent Sex Determination and Differentiation." *BMC Genomics* 17(1): 1–13. https://link.springer.com/articles/10.1186/s12864-016-2396-9 (July 5, 2022).

Young, Matthew, Maintainer Nadia Davidson, Anthony Hawkins, and G O biocViews Sequencing. 2013. "Package 'Goseq.'"

Yu, Guangchuang. 2021. "Enrichplot: Visualization of Functional Enrichment Result." https://yulab-smu.top/biomedical-knowledge-mining-book/.

Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. "ClusterProfiler: An R Package for Comparing Biological Themes among Gene Clusters." *Omics: a journal of integrative biology* 16(5): 284–87.

Zachos, J et al. 2001. "Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present." *Science* 292(5517): 686–93. https://science.sciencemag.org/content/292/5517/686.abstract (October 1, 2021).

Zhang, Zong Hong et al. 2014. "A Comparative Study of Techniques for Differential Expression Analysis on RNA-Seq Data." *PLoS ONE* 9(8): e103207. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0103207 (July 5, 2022).

Zuur, Alain F., Elena N. Ieno, and Chris S. Elphick. 2010. "A Protocol for Data Exploration to Avoid Common Statistical Problems." *Methods in Ecology and Evolution* 1(1): 3–14. https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/J.2041-210x.2009.00001.X (March 18, 2022).

Chapter 4: Differential gene expression among brushtail possums reveals metabolism discrepancies between subspecies.

Abstract

This chapter documents differential gene expression in brushtail possums (*Trichosurus vulpecula*) from Western Australian compared to those from New Zealand. The study uses RNA-Seq data to investigate gene expression differences and provides valuable insights into variation introduced by an experimental design that used more than one DNA sequencing platform. The main goal was to identify genetic pathways that might be involved in the tolerance shown by southwest Australian possums to the plant toxin sodium fluoroacetate. I used three approaches to identify differentially expressed genes associated with cell signalling, encapsulating structure, cell mobility, and tricarboxylic acid cycle. As expected, I found numerous significant differences in gene expression. The emergence of physiological adaptations specific to regional diet demonstrates the evolutionary response to local selection resulting in genomic divergence of spatial populations. The gene expression differences detected allowed me to hypothesise which metabolic pathways are most likely to be associated with sodium fluoroacetate resistance in these marsupials.

Introduction

Gene expression analysis is an essential tool for understanding how organisms are adapted to their environment. Advances in high-throughput sequencing technologies have facilitated the investigation of gene expression differences among populations and species of non-model organisms. Recently, the tools and platforms available for RNA-seq research have rapidly expanded, reducing costs but adding potential confounding factors into analyses. Differential gene expression studies have provided valuable insights into the genetic mechanisms underlying phenotypic diversity, as well as the evolutionary processes that shaped it (Costa-Silva, Domingues, and Lopes 2017; Todd, Black, and Gemmell 2016; Zhang et al. 2014). In this chapter, I explore differential gene expression of two populations of brushtail possum; the natural western Australian population and the translocated New Zealand population, representing two subspecies of *Trichosurus vulpecula* (How and Kerle 1995; Kerle, McKay, and Sharman 1991). Specifically, we focus on genes that show evidence of being involved in the tolerance of plant toxins, which has important implications for the management and conservation of this species.

In Australia, brushtail possums (*Trichosurus vulpecula*) have a very wide geographic range, from the southwest to the northeast of the country. The different lineages within this species

are considered subspecies despite their morphological similarities (Kerle, McKay, and Sharman 1991). The main delimiting factor is the geographical distance separating the different populations but genetic data (Campbell et al. 2021; Pattabiraman et al. 2021; Taylor et al. 2004) and a calibrated phylogeny (Carmelet-Rescan et al. 2022) provide support for the subspecies' delimitations with lineage differentiation dating to the Pliocene. Brushtail possums are arboreal herbivores who feed on their local plant species, thus regional differences in the vegetation across Australia cover have a strong impact on possum diet. In southwest Australia they live in Jarrah and Kerri eucalyptus forests; and in rainforests in Queensland. These differences in their habitat have led to an observed difference in their exposure and tolerance of plant chemicals. In southwest Australia brushtail possums forage on plants known to contain toxic compounds including *Erythophleum, Acacia, Eucalyptus* and *Gastrolabium*. *Gastrolabium* species contain sodium fluoroacetate (also known as 1080), and brushtail possums in southwest Australia have an LD50 160 times higher than their cousins in East Australia when exposed to potent toxin fluoroacetate (Leong et al. 2017; Twigg et al. 1996). This emergence of physiological adaptations specific to regional flora demonstrates local selection pressures has resulted in genomic divergence of spatial populations (Mead et al. 1985; Oliver and King 1979). It is likely that geographic isolation facilitated the development and maintenance of toxin resistance in *T. v. hypoleucus* in Western Australia in contrast with populations elsewhere (Twigg and King 1991). From the translocation records, and confirmed by the phylogenetic analysis, we know that New Zealand brushtail possums came from individuals brought from southeast Australian and Tasmanian lineages (*T. v. fuliginosus* and *T. v. vulpecula* subspecies respectively) in the midst of the nineteenth century (Pattabiraman et al. 2021; Pracy 1974, see Chapter 1 and 2).

As possum density is dependent on their foraging opportunities (Efford, Warburton, and Spencer 2000), New Zealand provides them an extensive and attractive habitat with dense palatable vegetation. Brushtail possums are a threat to New Zealand biodiversity as they feed on native endangered plant species (Nugent et al. 2000) but also show signs of opportunistic feeding on native bird eggs and insects. Possums are also a vector for bovine tuberculosis (TB) which is costly for the agricultural industry. In order to reduce livestock infection, possum populations are managed, using traps and the toxin sodium fluoroacetate in aerial baiting (McIlroy 1983; Ross, Bicknell, and Hickling 1999). Killing mammalian pest species has the added benefit of reducing the threat of brushtail possums on New Zealand's biodiversity (Byrom et al. 2015; Nugent, Buddle, and Knowles 2015). The application in New Zealand of

synthetic 1080 (McIlroy 1983; Ross, Bicknell, and Hickling 1999) to a million hectares each year results in repeated high local kill, as aerial drop can reach more than 80% efficiency (Nugent et al. 2012). However, this is followed by rapid population recovery (P. Cowan 2016; P. E. Cowan 1992; Gupta 2015; Ross, Bicknell, and Hickling 1999). This management practice may exert selection pressure favouring the evolution of tolerance to toxins in New Zealand populations, just as they have in wild southwest Australian possums and within experimental insect lines (Brown and Payne 1988). It is worth noting that not all regions of New Zealand have been exposed to 1080 for pest management, notably the possum population on Stewart Island at the southern end of the country is managed with traps so they might be under different selection pressure. The evolution of tolerance to 1080 in New Zealand possums would have major implications for agriculture and conservation, which is why determining the underlying basis of toxin tolerance is crucial for the future of possum management in New Zealand.

Differential expression analysis using RNASeq data is the perfect tool to identify gene expression differences of different molecular pathways between populations and species. In chapter 3 I showed that with the aid of the fully sequenced genome, differential expression analysis can be applied to brushtail possums following standard pipelines (Chapter 3). It is likely that tolerance of plant toxins is associated with differential gene expression (Adamczyk et al. 2001; Jenkins et al. 2015), and as metabolism of toxins is a primary function of the liver in mammals, my sampling and analysis was constrained to gene expression in the liver of possums. Building upon the valuable insights gleaned from the previous chapter, the findings related to gene expression patterns, particularly those associated with development and the impact of sample size, provide a solid foundation for the investigations and analyses presented in this chapter.

This chapter aims at identifying novel gene expression associated with the western Australian brushtail possum phenotype and seek evidence for candidate genes involved in toxin resistance of candidate genes. I do this by comparing expression data from different possum lineages and expect to find many differentially expressed genes associated with a wide range of cellular-pathways due to the phylogenetic distance between the subspecies sampled. However, I expect that genes that are associated with metabolic pathways, and especially acetate or those associated with other toxin resistance mechanisms will be differentially expressed. I generated RNASeq data from fresh liver samples of brushtail possum from New Zealand and Western Australia and performed differential gene expression analysis. I then investigated the pathways

associated with the genes showing evidence of differential expression to explore their potential association with tolerance of toxins and/or sodium fluoroacetate metabolism.

Material and Method

Fresh liver samples were taken from three euthanised adults possums in Perth, Western Australia, by WA Wildlife. Nine adults possums livers were sampled across New Zealand (Turitea, Manawatū, North Island n =3; Marlborough, Manaroa South Island, n=2; Stewart Island n=4) from individuals killed as part of pest management. The tissues were immediately emersed in ample RNALater (Invitrogen) preservation liquid and stored in -20 °C storage as soon as possible until extraction. The metadata for each sample is given in Table 1.

Messenger RNA is unstable making it challenging to transport samples between New Zealand and Western Australia. Therefore a number of variations were introduced in the protocol applied to each sample (from sampling to sequencing). Using a local laboratory for western Australian samples facilitated sample transfer and limited potential degradation prior to and after extraction. Samples from Otago were extracted and sequenced and mapped by the Hore laboratory and Otago Genomics Facility and correspond to the adult samples analysed in Chapter 2. The fresh New Zealand samples were extracted using the Nucleopsin RNA (Macherey-Nagel) after grinding the samples using liquid nitrogen, on-site at Massey University. The procedures were carried out according to the manufacturer's protocol. Western Australian samples were extracted by the Australia Genome Research Facility (AGRF Ltd, Melbourne, Victoria) using the Qiagen RNeasy Mini Kit automated on a QiaCube according to the manufacturer's instructions. Extractions were followed by a DNA removal step using DNase before library preparation. The concentration of DNase-treated RNA was determined using the Qubit RNA BR Assay Kit with the Qubit Fluorometer (Invitrogen). Quantity and quality were assessed using PerkinElmer LabChip® GX Touch HT. Library preparations were performed according to the sequencing platform and the sample quality. Poly(A) enriched libraries and rRNA removal were preferred for higher quality samples. Two of the Western Australian samples were not suitable for poly-A capture and were sequenced on a NovaSeq RNA Sequencing platform at AGRF Melbourne generating 150bp PE reads, and samples from Stewart Island were also sequenced using a NovaSeq platform via Custom Science Ltd NGS services (Auckland, New Zealand) generating 150bp PE reads. Other samples were sequenced

on the DNBSeq platform from BGI Tech Solutions Ltd (Tai Po, Hong Kong) producing 100bp PE reads.

Prior to mapping, the quality of the sequencing reads was assessed using FastQC (Andrews et al. 2010) and adapter sequence and quality-based trimming was performed using Trimmomatic (Bolger, Lohse, and Usadel 2014) removing reads ≤ 20 bp in length and reads with a quality score of ≤ 20. Sequences were then aligned to the brushtail possums (*Trichosurus vulpecula*) genome (Genebank : mTriVul1.pri - GCA_011100635.1) using HISAT2 (Kim et al. 2019), a spliced read aligner, with default parameters. Samtools (Li et al. 2009) was used to sort, format and output .bam alignment files, and the Picards tools (Institute 2019) implemented to collect alignment quality metrics. The number of DNA reads mapped to each gene was reported using featureCounts software (Liao, Smyth, and Shi 2014), providing a table of the number of mapped reads for each gene that will form the base data for differential expression analysis. Details of these steps are presented in a flow chart (Figure 1).

Feature-specific quantile normalisation was applied to the reads count variance within each category to eliminate platform-based bias using the FSQN package (Franks, Cai, and Whitfield 2018). This eliminates distribution-based differences resulting from the use of different gene expression profiling platforms.

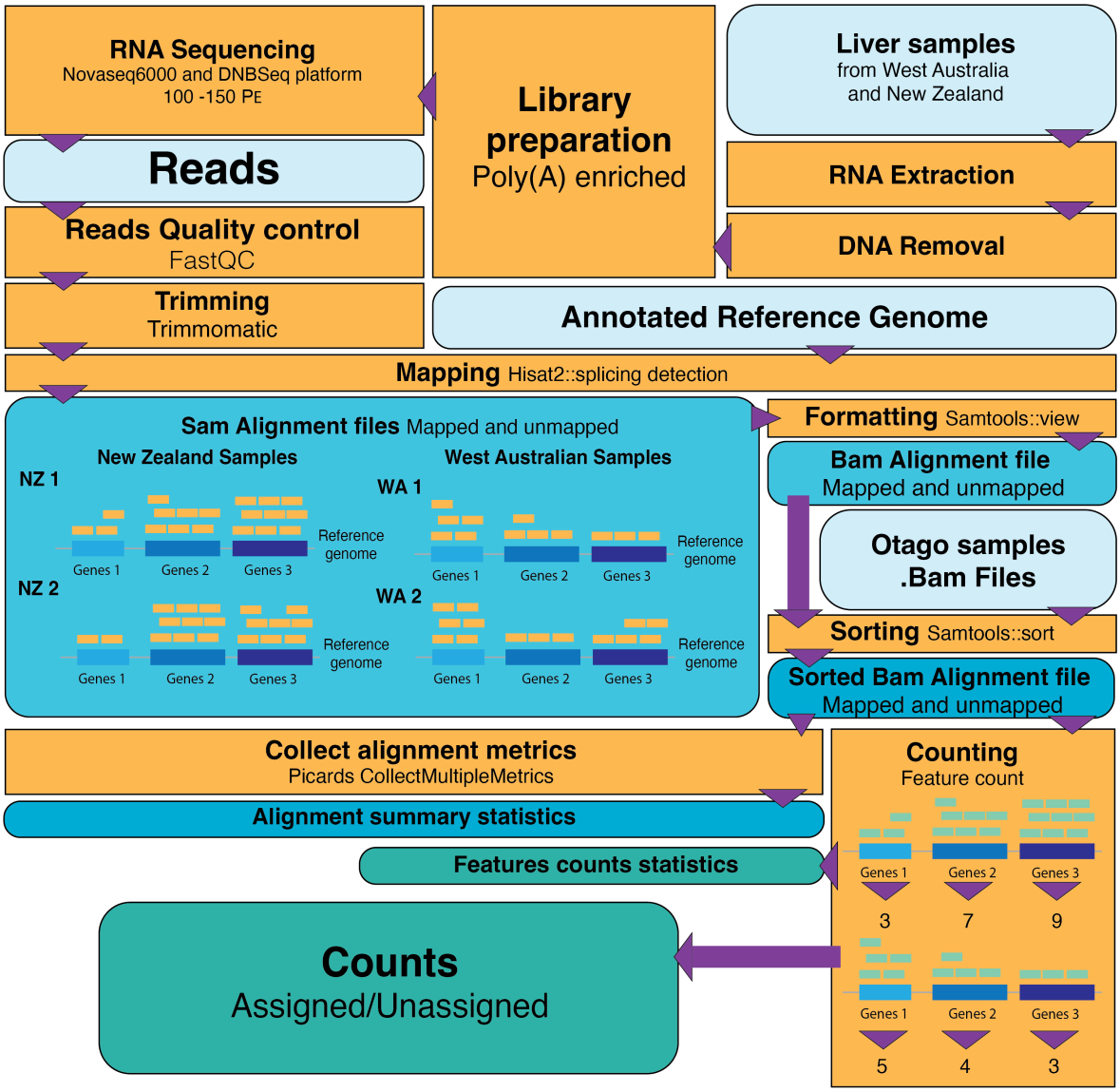


Figure 1: Flowchart of the method used to prepare data and compute read counts from RNASeq data obtained from brushtail possums (*Trichosurus vulpecula*).

Differential gene expression analysis based on the negative binomial distribution (DeSeq2)

To compute the differential expression analysis the R/Bioconductor package DeSeq2 was used (Love, Huber, and Anders 2014). The DeSeq function estimates the size factor (length of the gene) and the dispersion before fitting the data to a negative binomial generalized model and computing the Wald statistic for significance testing. Previous studies showed that parametric models are an appropriate approach when replicates are few (Kim et al. 2019) and that DeSeq2 responds to higher read depth by assigning smaller p-values to transcripts with small fold-change (Robles et al. 2012). Significantly differentially expressed genes (DEGs) were selected using FDR-adjusted p-values lower than 0.00001, using the insights of chapter 3 results, and a

log fold change of expression level above 2 or below -2 (corrected using shrinkage estimation) to improve stability and interpretability (Love, Huber, and Anders 2014). Prior analyses were performed using principal component analysis, clustered heatmapping of the significant DEGs and volcano plots to visualize those differences using DeSeq2 R package (Love, Huber, and Anders 2014). Genes identified in Chapter 3 as being associated with developmental process have been systematically filtered and excluded from further analysis in order to focus on the specific genetic factors relevant to the current research objectives.

To explore how western Australian samples group in relation to New Zealand adults and juveniles based on genes associated with development, we conducted a differential expression analysis utilizing DeSeq2. This analysis focused on the 475 differentially expressed genes (DEGs) related to development, as identified in the previous chapter. I used both juvenile and adult samples from Otago, as used in the previous chapter, along with the Western Australian samples collected for this chapter.

Since the *Trichosurus vulpecula* genome is not included in the database of GOs (Gene Ontologies) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways I used the genome (*Pan troglodites*) that returned the most identical genes (> 83% of found genes) to identify putative GO terms and Kegg pathways associated to those genes. Gene Ontology enrichment analysis was also computed on the significantly differentially expressed genes using goseq and clusterProfiler packages (Young et al. 2013; Yu et al. 2012) to determine the over-represented GO terms present in the list. The significance of enriched GO terms is determined using Wallenius non-central hypergeometric distribution (Wallenius 1963) of significantly differentially expressed genes p-values, including Benjamini & Hochberg FDR control (Benjamini and Hochberg 1995). Those GO terms and their parent relations were explored using REVIGO (Supek et al. 2011) and visualized as circle charts using the program CirGO (Kuznetsova et al. 2019).

KEGG enrichment analysis used the enrichKEGG function of the “clusterProfiler” R package (Yu et al. 2012). This analysis includes an FDR control. The significantly enriched pathways alongside the associated under-expressed and over-expressed genes were then plotted using the “enrichplot “ R package (Yu 2021).

Weighted correlation network analysis (WGCNA)

The search for differentially expressed genes was also carried out utilizing weighted correlation network analysis and the WGCNA R package (Langfelder and Horvath 2008). Genes with more than 50% of missing data were filtered out reducing the number of studied genes to 15,674 and expression values were normalized using DESeq2 variance stabilisation. The function "hclust" from the package fastcluster (Müllner 2013) is used on the expression results to cluster samples and exclude samples that deviated based on cluster height value (> 150). The construction of the co-expression network was realized using Pearson's correlation and then the adjacency matrix using the function $a_{mn} = |c_{mn}|^b$ (a_{mn}: adjacency between gene m and gene n, c_{mn}: Pearson's correlation, b: soft-power threshold). Soft-power threshold selection uses the lowest power for which the scale-free topology fit index reaches 0.80. Noise and spurious association effects were minimized by transforming the adjacency matrix into a Topological Overlap Matrix and then calculating the associated dissimilarity. Hierarchical clustering on the TOM-based dissimilarity is then performed to produce a hierarchical clustering dendrogram of genes. The dendrogram modules were identified using Dynamic Tree Cut then similar modules were merged based on co-expression similarity of entire modules. Module association with the traits of interest (maturity and sex) were quantified by performing principal components analysis of each module, the first components of each module are called the module eigengenes (MEs). Pearson correlation between the traits and the MEs is then calculated with their associated P-value. Modules with a significant p-value ($P< 0.01$) were selected to look for significantly differentially expressed genes. Gene significance (GS) was computed as the correlation between each individual gene of a module with the biological trait (subspecies), and module membership (MM) as the correlation between the gene and the module expression profile. Significant genes were then determined using a threshold of MM > 0.8 and GS > 0.8 in the significant modules. Similarly, gene ontology enrichment analysis was performed on significant DEGs discovered by WGCNA using the same method as above.

Differential expression analysis based on linear model (Limma)

The data were also analysed using an approach with a different normalization process implemented in the Limma R package (Ritchie et al. 2015). The first step calculates the scaling factor for each gene from the count data according to the library size and then transforms the counts to normalized log2-counts per million. The next step fits multiple linear models and then computes associated statistics needed to select the significantly differentially expressed

genes. The p-value cutoff for significance was set to 0.00001 and log2-fold over 2 and under -2. GO enrichment analysis was performed the same way on Limma DEGs as on DESeq2 ones.

To compare the results of the three different analyses proportional Venn diagrams were computed using the R package eulerr (Larsson and Gustafsson 2018). A flowchart detailing the method associated with differential expression analysis using the three different packages is represented in Figure 2.

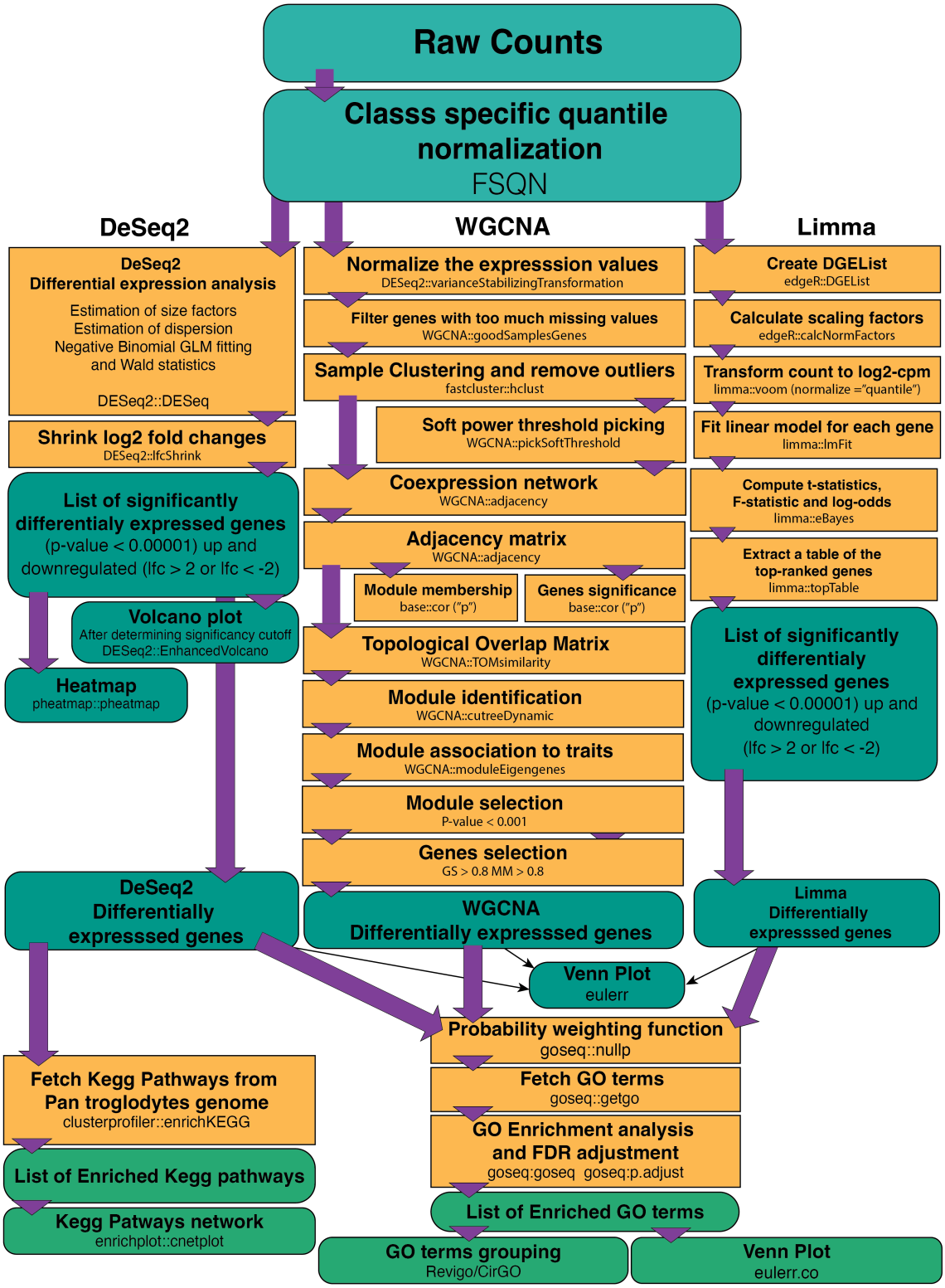


Figure 2: Flowchart of the method to compute differential expression analysis from count data from RNASeq sequences from liver of brushtail possums (*Trichosurus vulpecula*) using DESeq2, Limma and WGCNA.

Results

RNA-Seq analysis was performed on liver samples of brushtail possums (*Trichosurus vulpecula*) to explore gene expression differences among brushtail possums sampled from Western Australia and New Zealand, representing two major lineages within this species. A strict quality control analysis for each sample was conducted to confirm the quality of the short DNA sequences (reads) yielding more than 13 million reads of retained clean data per sample, representing a total of > 1.3 Gbps (Table 1).

Table 1: Samples of brushtail possums (*Trichosurus vulpecula*) used to study differential expression of genes in the liver. Data output quality and mapping quality for the examined samples, number of reads in million, number of base pairs in Gb and the proportion of aligned base pairs with at least a Q20 quality. In addition, the sample origin sequencing provider and sequencing platform are also displayed.

SAMPLE	ORIGIN	SEQUENCING PROVIDER	SEQUENCING PLATFORM	NUMBER OF READS	BASE PAIRS (GB)	Q20 BASES (%)
WA_1	Western Australia	AGRF	NovaSeq	40.56	5.89	89.86
WA_2	Western Australia	AGRF	NovaSeq	59.22	8.59	89.84
WA_3	Western Australia	BGI	DNBSeq platform (low quantity)	14.78	1.47	92.59
MA_1	Manaroa	BGI	DNBSeq platform	46.04	4.59	87.94
MA_2	Manaroa	BGI	DNBSeq platform	40.90	4.01	82.71
OT_1	Otago	Otago Genomics	HiSeq 2000 (Illumina)	26.08	2.63	92.43
OT_2	Otago	Otago Genomics	HiSeq 2000 (Illumina)	26.98	2.72	92.00
OT_3	Otago	Otago Genomics	HiSeq 2000 (Illumina)	34.98	3.53	90.91
OT_4	Otago	Otago Genomics	HiSeq 2000 (Illumina)	25.61	2.58	92.00
OT_5	Otago	Otago Genomics	HiSeq 2000 (Illumina)	17.68	1.78	92.41
OT_6	Otago	Otago Genomics	HiSeq 2000 (Illumina)	24.10	2.43	92.38
OT_7	Otago	Otago Genomics	HiSeq 2000 (Illumina)	27.89	2.81	92.50
OT_8	Otago	Otago Genomics	HiSeq 2000 (Illumina)	17.65	1.78	92.30
OT_9	Otago	Otago Genomics	HiSeq 2000 (Illumina)	39.96	4.03	91.89
OT_10	Otago	Otago Genomics	HiSeq 2000 (Illumina)	27.26	2.75	92.12
OT_11	Otago	Otago Genomics	HiSeq 2000 (Illumina)	28.54	2.88	91.33
OT_12	Otago	Otago Genomics	HiSeq 2000 (Illumina)	13.02	1.31	91.51
OT_13	Otago	Otago Genomics	HiSeq 2000 (Illumina)	25.09	2.53	92.74
OT_14	Otago	Otago Genomics	HiSeq 2000 (Illumina)	24.76	2.50	93.32
ST_1	Stewart Island	Custom Science	NovaSeq	58.45	8.73	91.68
ST_2	Stewart Island	Custom Science	NovaSeq	34.62	5.15	91.58
ST_3	Stewart Island	Custom Science	NovaSeq	36.64	5.46	91.54
ST_4	Stewart Island	Custom Science	NovaSeq	45.79	6.84	91.29
MN_1	Manawatu	BGI	DNBSeq platform	26.49	2.64	91.56
MN_2	Manawatu	BGI	DNBSeq platform	25.29	2.47	94.35
MN_3	Manawatu	BGI	DNBSeq platform	24.90	2.48	90.16

The time and place of sampling possum populations coupled with constraints on tissue preservation for RNA meant than the optimal procedure required RNA extraction and sequencing to be done in several different laboratories. There is an excess of New Zealand samples compared to western Australian samples and there is a correlation between sequencing provider provenance of samples. Different sequencing technologies generated varying numbers of reads and reads of differing length. Initial analysis revealed a relatively high number of unmapped reads from the AGRF platform, and the Otago samples yielded less

sequence depth with more than 20% of reads mapping to unannotated segments of the genome (Figure 1). Nevertheless, the overall quality of the reads was good among all samples.

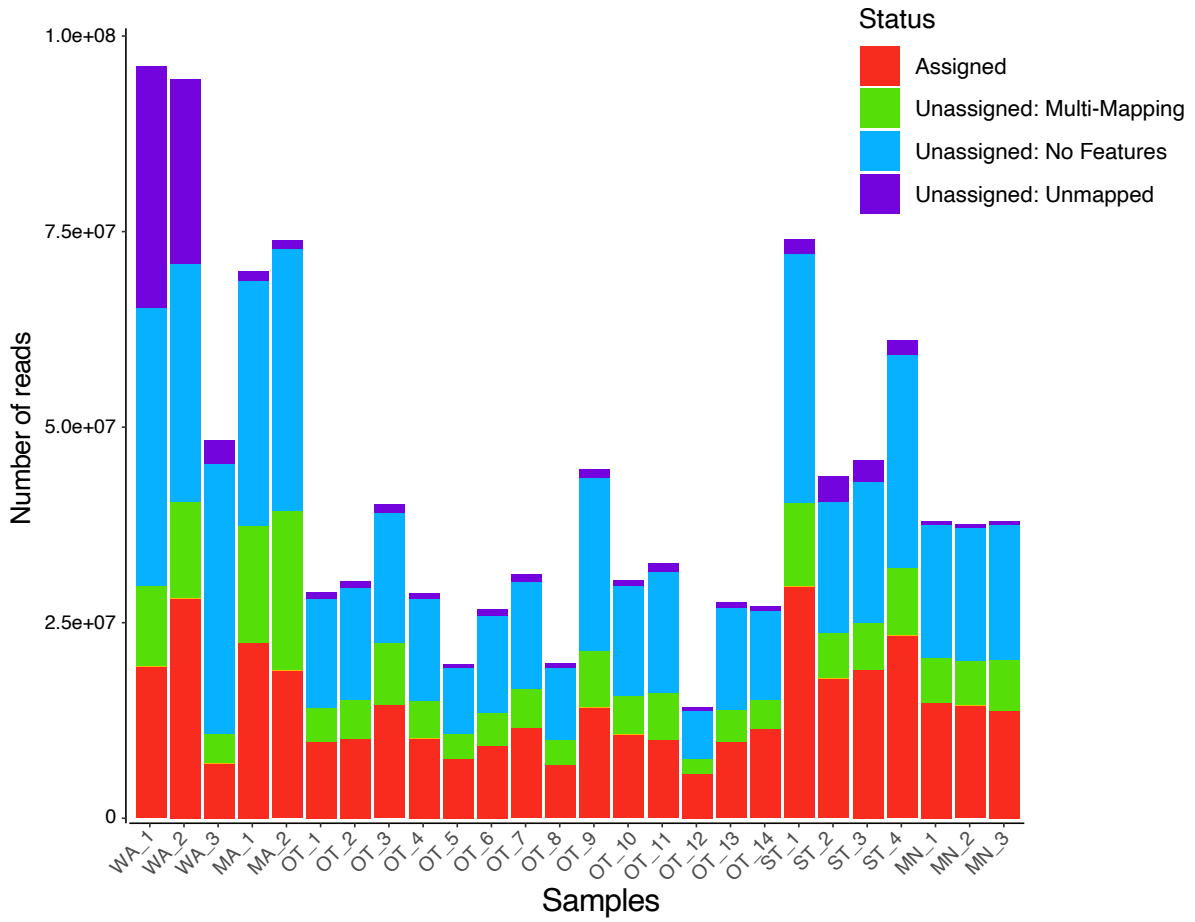


Figure 1: Number and assignment of reads from RNASeq for *Trichosurus vulpecula* samples used in differential expression analysis. Assigned corresponds to read that mapped one time on an annotated genetic region. Multi-mapping corresponds to reads mapping several times on the genome. No feature to reads mapping on the genome but to regions without any feature (gene annotation). And Unmapped to read that unsuccessfully mapped

The principal component analysis of the normalized counts from the differential gene expression analysis using DESeq2 R package (Love, Huber, and Anders 2014) clustered samples according to their sequencing platform (Figure 2A). This clustering disappeared after using feature-specific quantile normalisation (FSQN) with the principal difference in gene expression separating the New Zealand and Western Australian samples (PC1 = 39.9% of variation; Figure 2B).

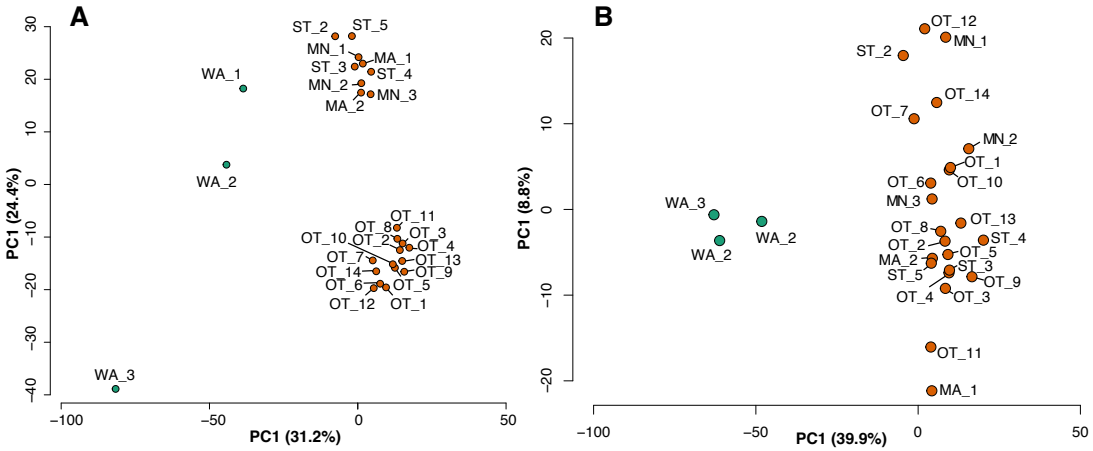


Figure 2: A) Principal Component Analysis (PCA) plot showing the results of differential gene expression analysis applied to the normalized count data of each brushtail possum (*Trichosurus vulpecula*) liver sample. Each dot represents a sample, with the samples coloured based on origin (Green: Western Australia and Orange: New Zealand) and the eigenvalue proportion of each axis is indicated below them. The plot illustrates the distribution and separation of samples based on their gene expression patterns, highlighting any differences between groups. B) Principal Component Analysis (PCA) applied to the normalized count data of each brushtail possum (*Trichosurus vulpecula*) liver sample after feature-specific quantile normalisation (FSQN).

Significantly differentially expressed genes were selected with a p-value threshold of 1e-05 and a log fold change of 2 (Figure 3). I identified 1041 differentially expressed genes (DEGs), 198 upregulated in New Zealand possums compared to Western Australia and 843 downregulated in New Zealand populations compared to Western Australia (Figure 4).

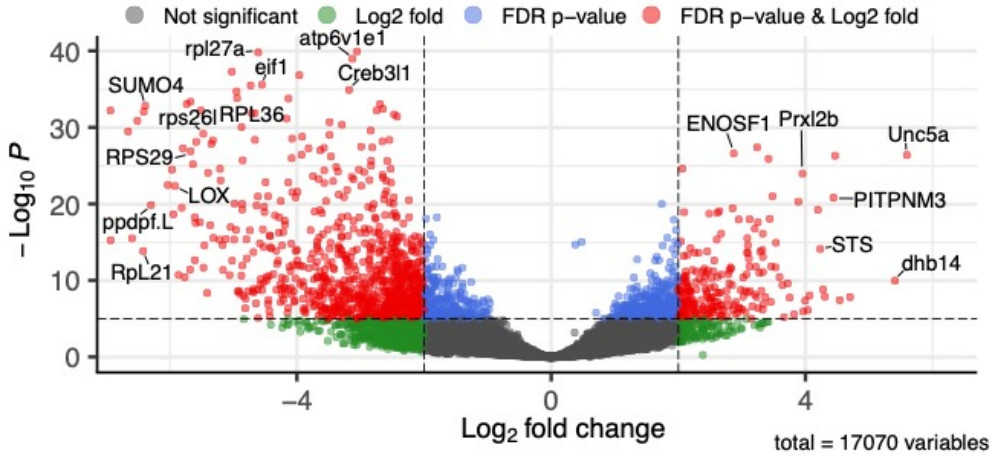


Figure 3: Volcano plots displaying differentially expressed genes (DEGs) samples between New Zealand and Western Australia brushtail possums (*Trichosurus vulpecula*) liver samples using R packages "DESeq2". The representations are as follows: x-axis, log2 fold change; y-axis, -log10 of a p-value. The genes with p-values < 0.00001 are the blue dots, and the genes with logFC ≥ 2 and logFC ≤ –2 are the green dots; the significant DEGs satisfying both values thresholds are in red and some are indicated with gene names. Black dots indicate the remaining genes present in the array that were not significantly differently expressed. The genes that are upregulated in the New Zealand populations are on the right panel, and those upregulated in Western Australia are on the left panel of the plot.

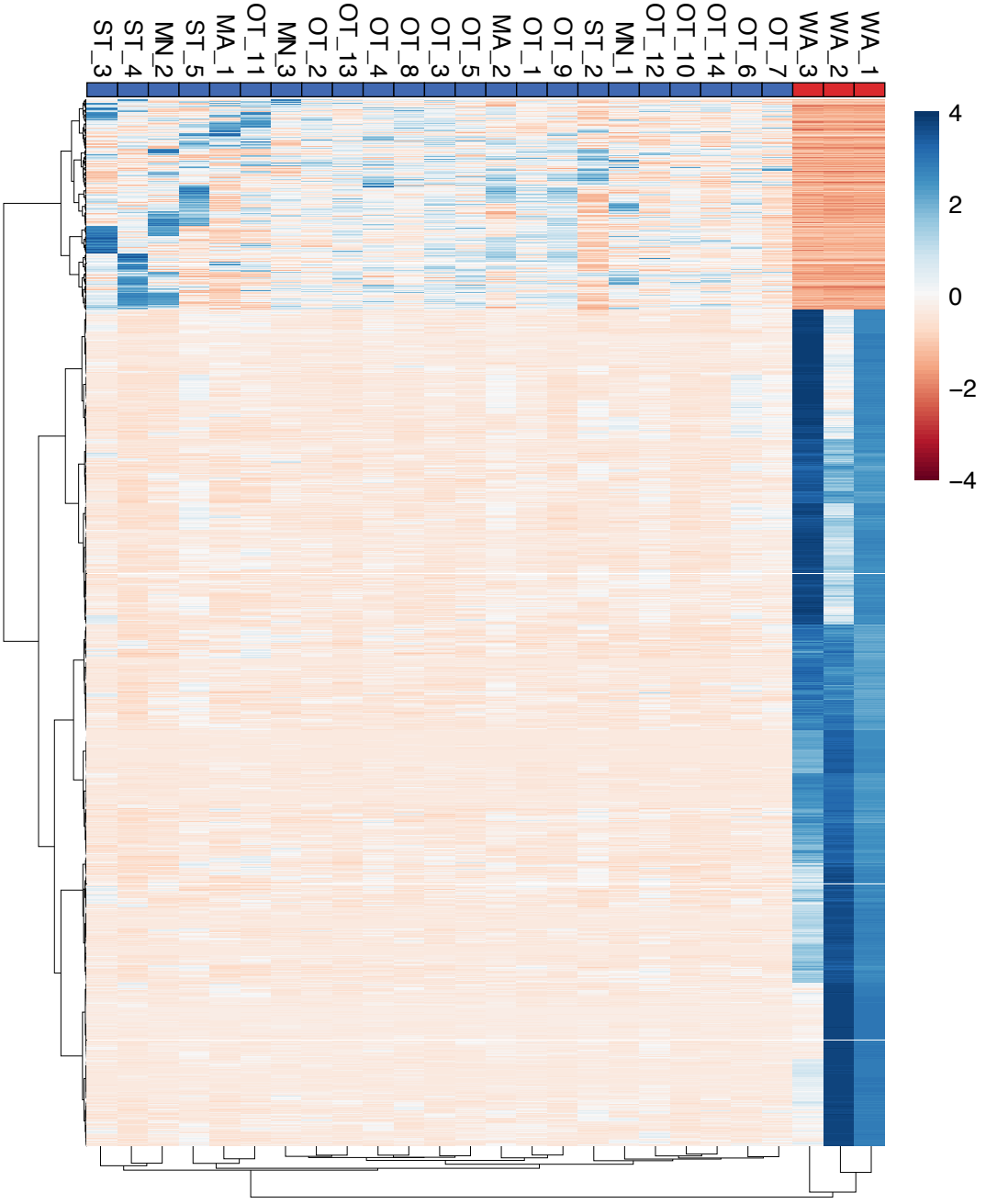


Figure 4: Heatmap representation of differentially expressed genes (DEGs) between samples from New Zealand and Western Australia brushtail possum (*Trichosurus vulpecula*) liver samples using R packages "DESeq2". n = 1147 DEGs. Rows represent individual genes, columns represent samples, and the colour intensity represents the expression levels of each gene in each sample. Sample origin is indicated with the colour band (Red: Western Australian, Blue: New Zealand). The heatmap is clustered to group together similar expression patterns and highlight any differences between samples based on gene expression

Two groups (Western Australia and New Zealand possum samples) were clearly distinguished using hierarchical clustering of gene expression (Figure 4). This heatmap representation also reveals within-group differences in gene expression among upregulated genes.

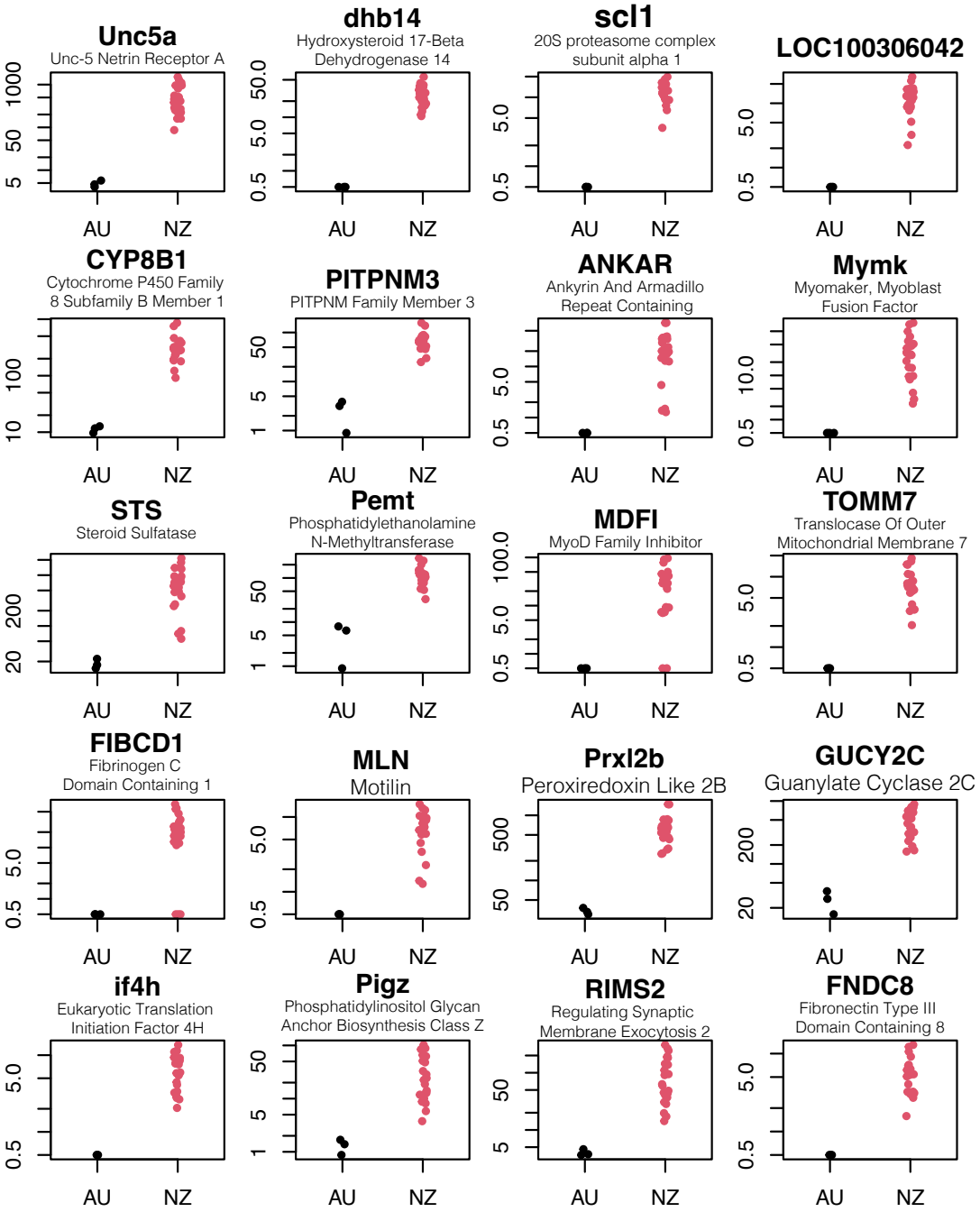


Figure 5: Normalized counts of the 20 most significantly differentially expressed downregulated genes in Western Australian compared to New Zealand brushtail possums (*Trichosurus vulpecula*) liver RNASeq samples. Western Australia black; New Zealand red.

Examination of the top twenty upregulated and downregulated genes (Figures 5 and 6) illustrates the expression differences between populations, but also reveals high variation among the New Zealand samples in terms of expression levels of some genes such as MDFI

(MyoD Family Inhibitor), FIBCD1 (Fibrinogen C Domain Containing 1) and ANKAR (Ankyrin and Armadillo Repeat Containing).

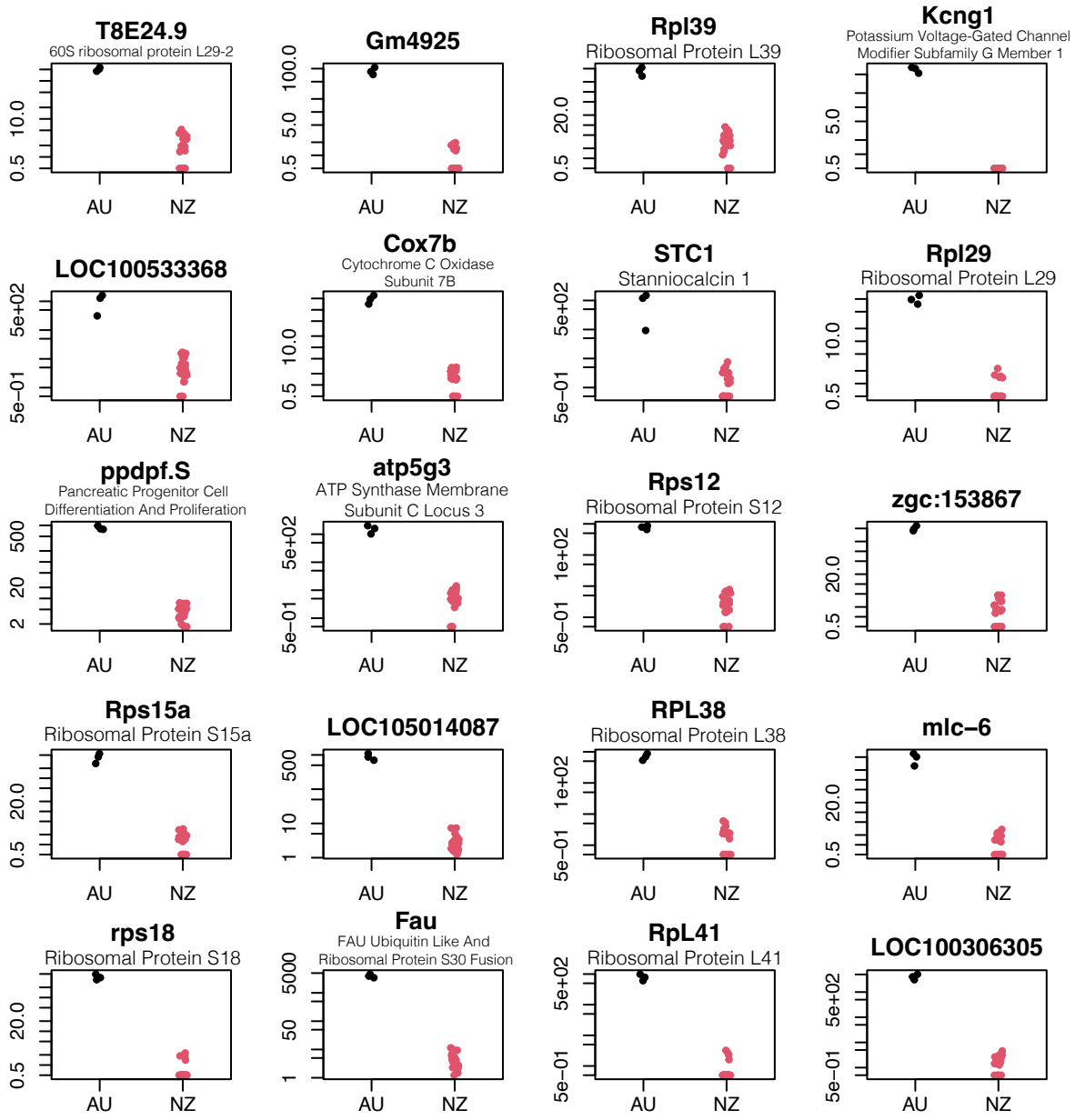


Figure 6: Normalized counts of the 20 most significantly differentially expressed upregulated genes between western Australian and New Zealand brushtail possums (*Trichosurus vulpecula*) liver RNASeq samples. Western Australia in black and New Zealand red.

In the analysis, out of the initial set of 475 genes associated with development 133 genes exhibited differential expression in the WA samples. However, due to this substantial overlap, it is challenging to definitively attribute this pattern solely to development, as other factors may also be influencing gene expression. Furthermore, principal component analysis of the

normalised read counts of the 475 genes linked to development revealed distinct clustering of the Western Australian samples, setting them apart from both New Zealand adults and juveniles (Appendix C4 Figure 1). These results imply that the gene expression patterns associated with development in WA possums exhibit notable differences, thereby warranting their utilisation in the investigation of toxin resistance mechanisms.

A greater proportion of genes were upregulated in the Western Australian possum samples than were upregulated within the New Zealand samples. Because many ribosomal genes appeared as upregulated in the western Australian population it is possible that the rRNA removal during sample preparation was less efficient for these samples. This could lead to unequal representation of rRNA sequences among samples so rRNA genes were excluding from subsequent analyses as well as gene associated with development found in previous chapter (Chapter 3) reducing the list to 959 differentialy expressed genes. Gene ontology enrichment analysis revealed 287 enriched terms, 268 significantly enriched terms among upregulated genes in Western Australia and 19 significantly enriched terms among downregulated genes. The Revigo (Supek et al. 2011) grouping of the enriched terms is visualized using CirGO (Kuznetsova et al. 2019) (Figure 7) and shows that a majority of the upregulated genes are associated with the biological processes of regulation of cell migration and anatomical structure development and the cellular components of the cytosolic ribosomes and external encapsulating structures. On the other hand, the downregulated genes are associated with small molecule catabolic process and flavin adenine dinucleotide binding (Appendix C4 Table 1 and 2).

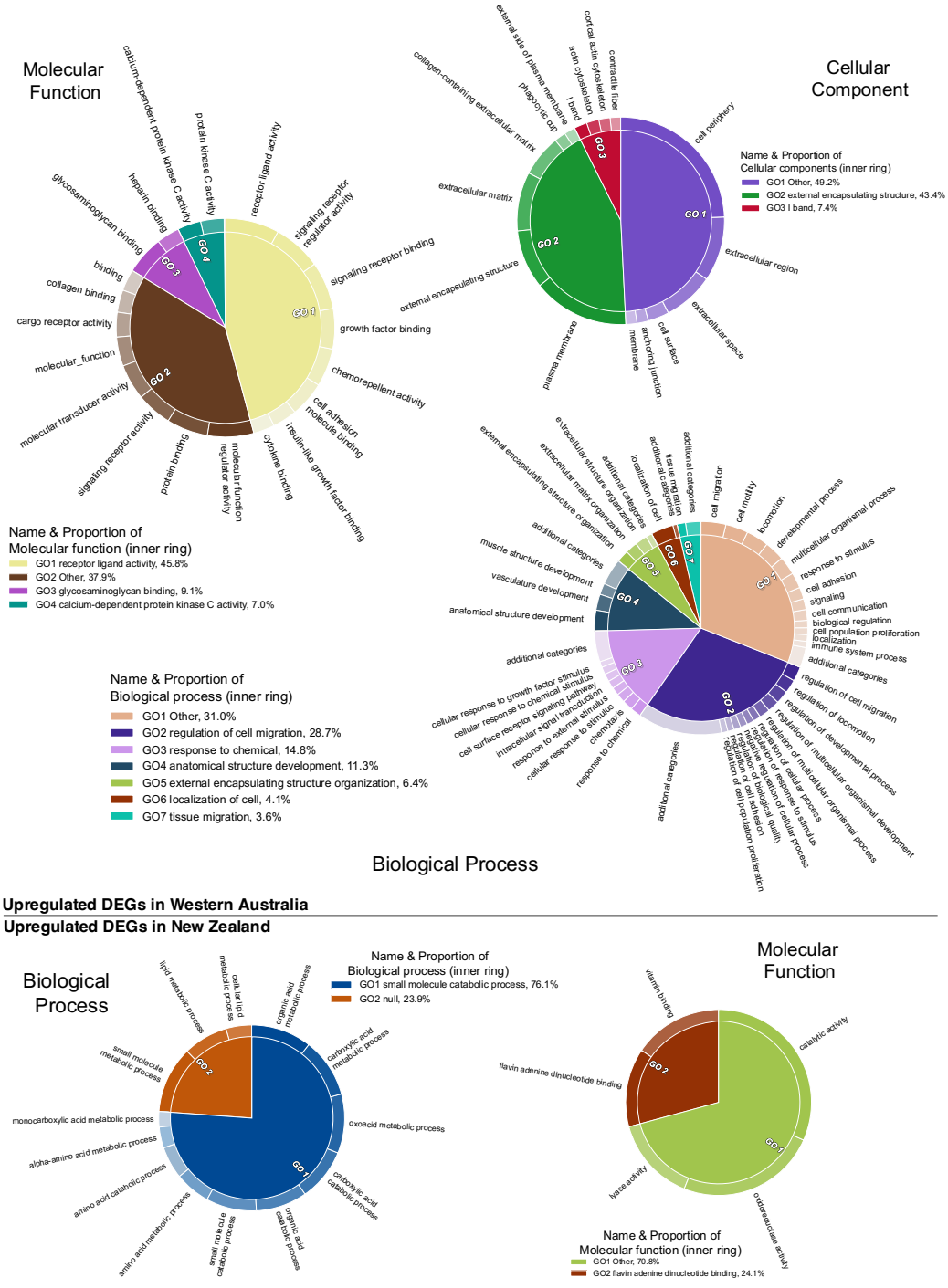


Figure 7: Gene Ontology (GOs) enrichment maps analysis among the significantly differentially regulated genes isolated using DESeq2 between Western Australian and New Zealand brushtail possum (*Trichosurus vulpecula*) liver. Terms are grouped by hierarchical clustering. Parent terms are identified in the legend and their respective proportions, directly proportional to statistical significance. GO terms were first summarized based on a semantic similarity of 0.4 using REVIGO and visualized in CirGO. Circles correspond to one ontology group (BP: Biological process, MF: Molecular function and CC: Cellular component) and are separated between the GOs associated with the upregulated DEGs in Western Australia and in New Zealand.

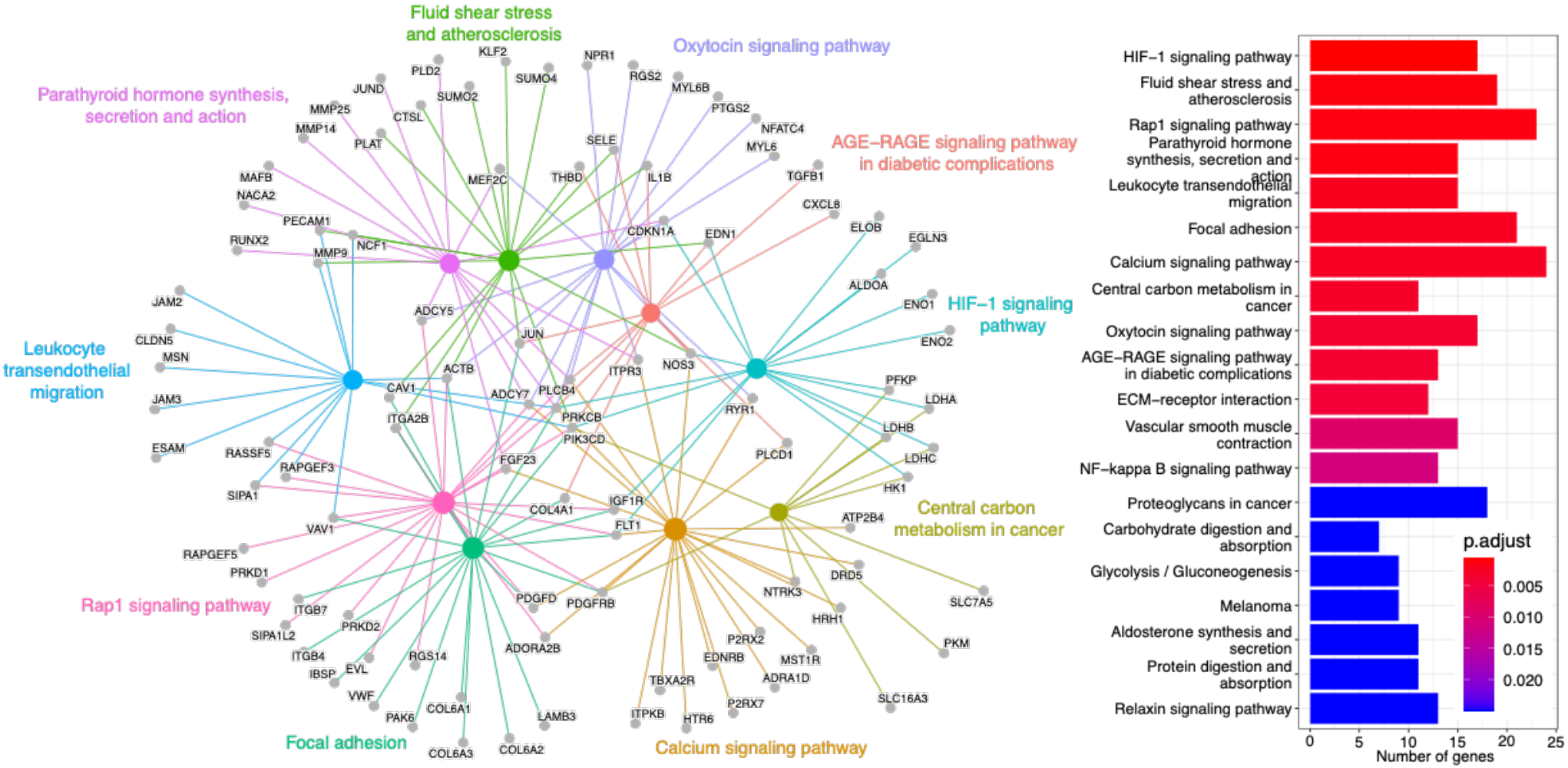


Figure 8: Enrichment by pathway terms is visualized using the cnetplot function from the "enrichplot" R package. Significantly enriched KeGG pathways (FDR ≤ 0.05) associated with the significant upregulated DEGs between Western Australian and New Zealand brushtail possums (Trichosurus vulpecula). Big and coloured nodes represent the pathways and grey nodes are the differentially expressed genes associated with those pathways. Nodes are sized using enrichment analysis p-value. The number of genes and FDR-adjusted p-values associated with the top twenty enriched pathways are reported on the bar plot.

significantly enriched KeGG pathways associated with the significant upregulated DEGs also display genes associated with ribosomes. Several other pathways are significantly enriched such as some signalling pathways (HIF-1 and Rap1) (Figure 8). These results correlate with the GO enrichment analysis but also add details on the possible functions that could be associated with the significant upregulated DEGs in Western Australian possums.

KeGG pathways found among the significantly downregulated DEGs also correlate with the GOs analysis by displaying small molecule metabolism pathways, but it also displays seven genes associated with the peroxisome (Figure 9).

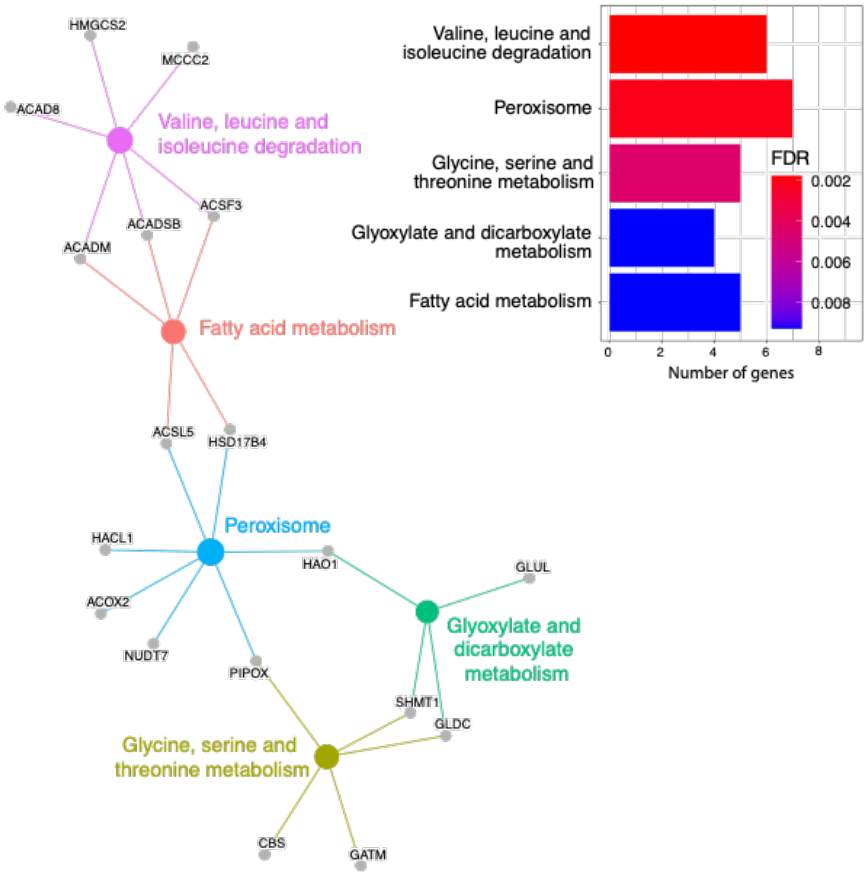


Figure 9: Enrichment by pathway terms is visualized using the cnetplot function from the "enrichplot" R package. Significantly enriched KeGG pathways (FDR ≤ 0.05) associated with the significant downregulated DEGs between Western Australian and New Zealand brushtail possums (*Trichosurus vulpecula*). Big and coloured nodes represent the pathways and grey nodes are the differentially expressed genes associated with those pathways. The number of genes and FDR-adjusted p-values associated with each significantly enriched pathway are reported on the bar plot.

Two other approaches (linear models and correlation networks) were used to compute significantly differentially expressed genes. The Limma package generated 1221 differentially expressed genes (1073 upregulated in Western Australian possums 148 downregulated). Those genes allowed me to identify 325 enriched GO terms (Figure 10) that closely match the terms identified from the DEGs identified with DESeq2. The common enriched terms among upregulated DEGs are associated with ribosome,

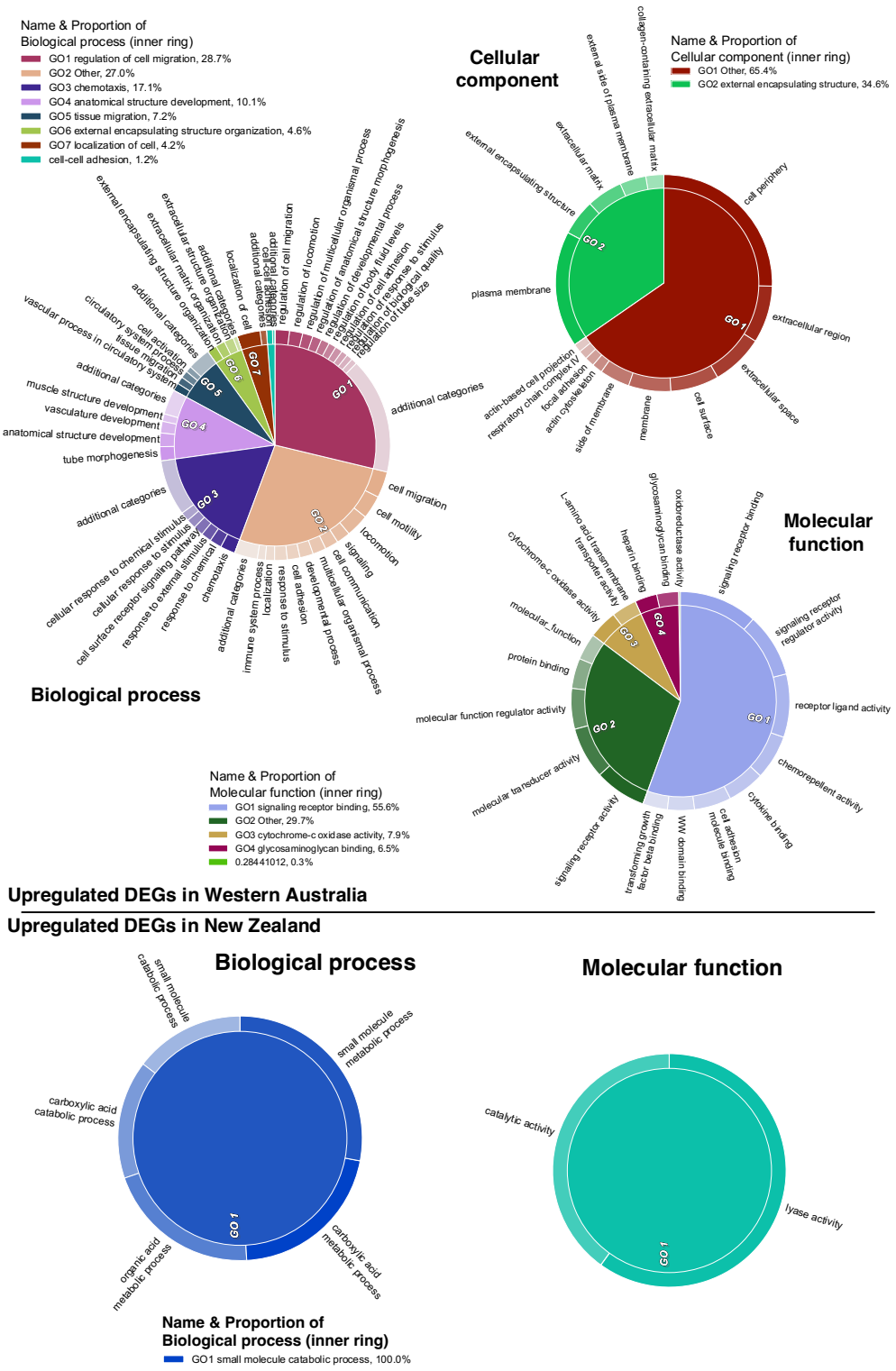


Figure 10: Gene Ontology (GOs) enrichment maps analysis among the significantly differentially regulated genes isolated using Limma between Western Australian and New Zealand brushtail possum (*Trichosurus vulpecula*) liver. Terms are grouped by hierarchical clustering. Parent terms are identified in the legend along with their respective proportions, which are directly proportional to statistical significance. GO terms were first summarized based on a semantic similarity of 0.4 using REVIGO and visualized in CirGO. Circles correspond to one ontology group (BP: Biological process, MF: Molecular function and CC: Cellular component) and are separated between the GOs associated with the upregulated DEGs in Western Australia and in New Zealand.

regulation of cell migration, chemotaxis and external encapsulating structure. Among downregulated DEGs are small molecule catabolic process (Appendix C4 Table 3 and 4).

Differential expression analysis using the WGCNA package (Langfelder and Horvath 2008) identified three modules significantly associated with the Western Australian samples compared to the New Zealand ones (Figure 11A); two of them are positively correlated (brown and lightyellow) and the other one is negatively correlated (grey60). Within those modules, the genes having a gene significance and module membership both over 0.8 were defined as differentially expressed (Figure 11B). Overall, the WGNA package identified 1006 differentially expressed genes only two of them were downregulated in Western Australian samples the 1004 others were upregulated. This led to the finding of only enriched gene ontology (GO) terms associated with upregulated genes and none associated with downregulated genes.

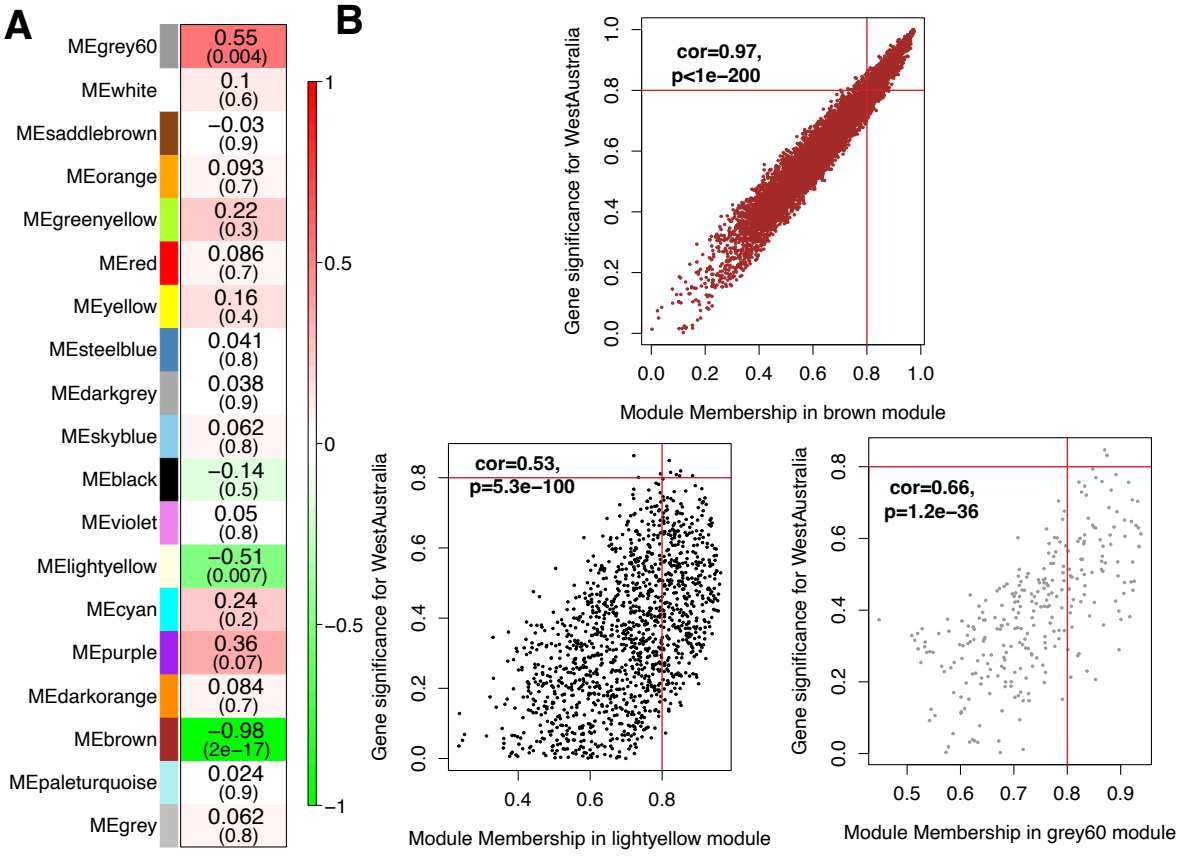


Figure 11: A) Heatmap of correlation of gene modules and the association of the samples to the western Australian population. The colour and the first value of each cell correspond to the correlation score and the value in brackets corresponds to the p-value of the correlation. B) Plot of Module Membership and Gene significance for maturity for each of the five significant clusters. The associated correlation score and the p-value are indicated on top of each plot, and MM and GS thresholds used to select significant genes are shown with red lines.

As with results from Limma, many common terms were found in the enriched GOs analysis of the DEGs using the WGNA and the DESeq2 packages, but their ranking of importance differs (Figure 12). Among the new GO terms discovered among the WGCNA DEGs, many are associated with "G protein-coupled receptor signalling pathway" being the first category of GO term within Biological processes (Appendix C4 Table 5).

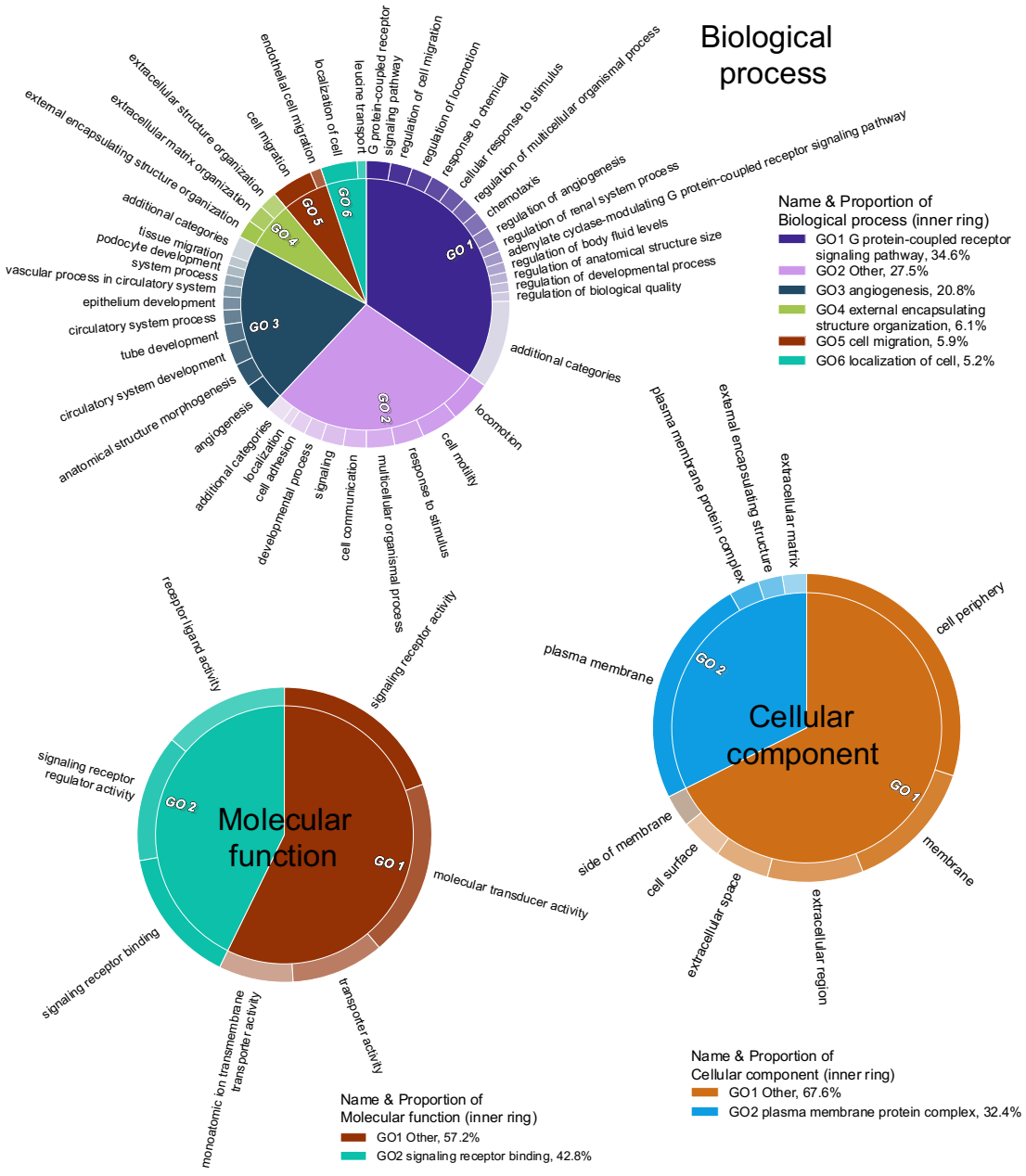


Figure 12: Gene Ontology (GOs) enrichment maps analysis among the significantly differentially regulated genes isolated using WGCNA between western Australian and New Zealand brushtail possum (*Trichosurus vulpecula*) liver. Terms are grouped by hierarchical clustering. Parent terms are identified in the legend and their respective proportions, directly proportional to statistical significance. GO terms were first summarized based on a semantic similarity of 0.4 using REVIGO and visualized in CirGO. Circles correspond to one ontology group (BP: Biological process, MF: Molecular function and CC: Cellular component) and are separated between the GOs associated with the upregulated DEGs in western Australian and in New Zealand.

The majority of differentially expressed genes were discovered by all three approaches used in this chapter (Figure 13) but each package identified a similar proportion of unique genes. Comparing the GO terms, the correlation network approach (package WGCNA) had fewer terms than the other two approaches but each analysis shares more than half of their discovered enriched GO terms with another analysis.

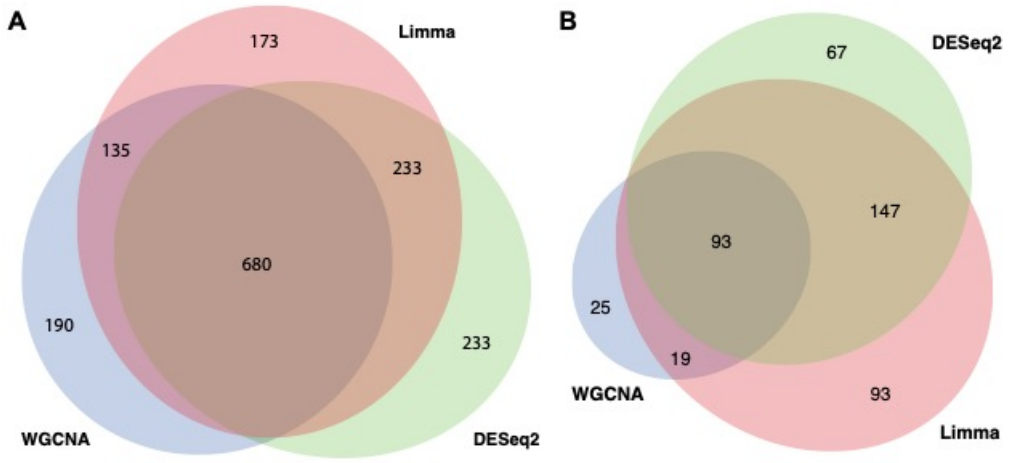


Figure 13: A: Visualization of the number of common differentially expressed genes (DEGs) using proportional Venn plots comparing the different packages used. B: Visualization of the number of common enriched GO terms using proportional Venn plots comparing the different packages used.

In order to explore the possible pathways associated with the toxin resistance, KeGG pathway enrichment analysis was performed on all 1622 genes with FDR-adjusted p-values below 0.00001 discarding the previous log-fold change (LFC) threshold. This analysis indicated several significantly enriched pathways there were different from the previous analyses (Figure 14). The “Peroxisome”,“Rap1 signalling” and “HIF-1 signalling” pathways were found in both analysis. Apart from “Complement and coagulation cascades” the other new pathways show a high degree of overlap in the gene sets implicated in the enriched pathways. Those overlapping pathways are centred around carbon metabolism, which is known to be relevant in sodium fluoroacetate metabolism.

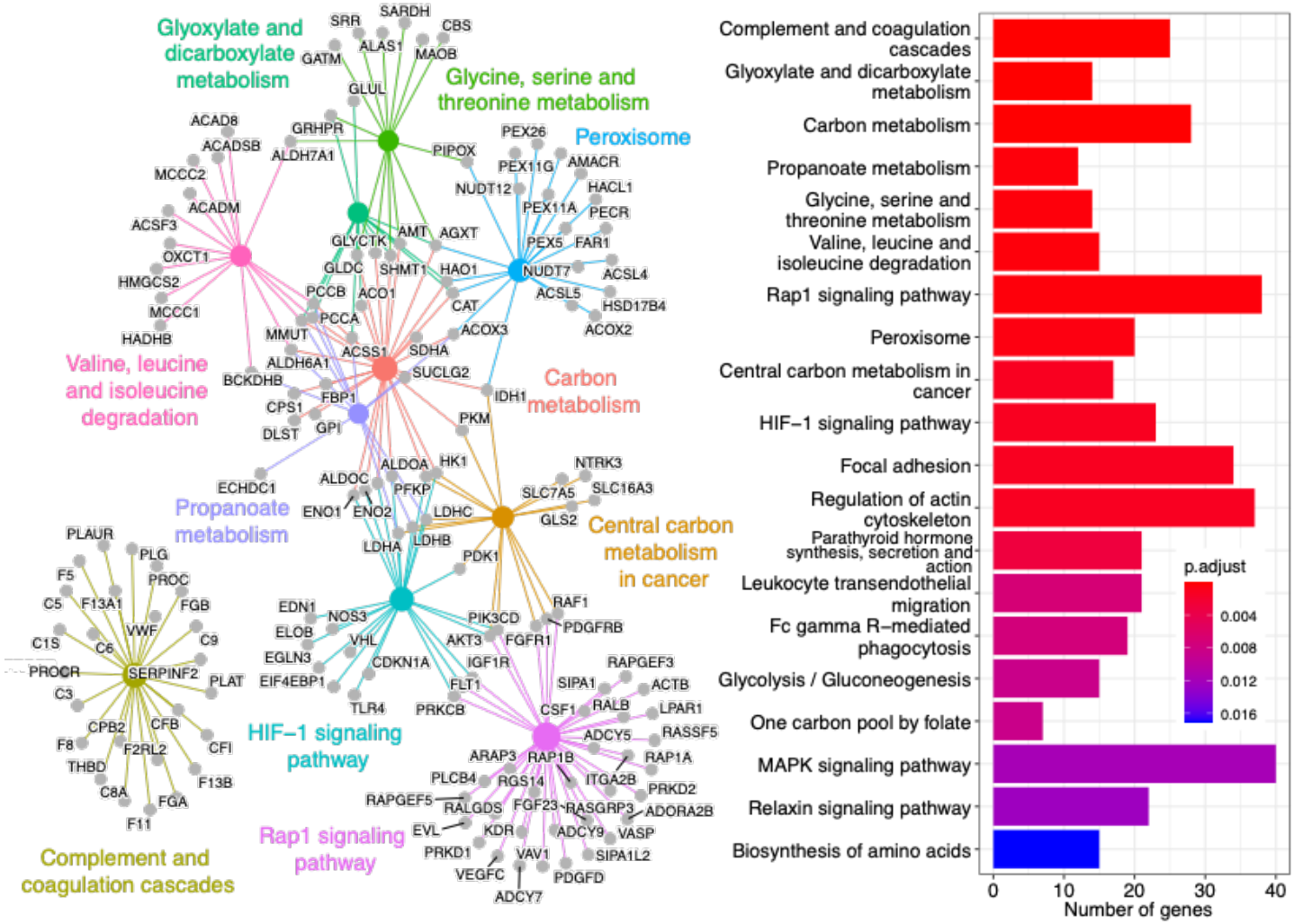


Figure 14: Enrichment by pathway terms is visualized using the cnetplot function from the "enrichplot" R package. Significantly enriched KeGG pathways (FDR ≤ 0.05) associated with the significant differentially regulated DEGs between western Australian and New Zealand brushtail possums (*Trichosurus vulpecula*) without using LFC cutoff. Big and coloured nodes represent the pathways and grey nodes are the differentially expressed genes associated with those pathways. The number of genes and FDR-adjusted p-values associated with each significantly enriched pathway are reported on the bar plot.

Looking at the genes associated with carbon metabolism shows that those genes show obvious differential expression, some genes are upregulated in western Australian samples (9) and others are downregulated (18) (Figure 15). Among the genes being part of carbon metabolism-related pathways, several are significantly overexpressed among western Australian samples (FDR corrected p-value < 0.00001 & LFC < -2) and expression level are shown on Figure 16.

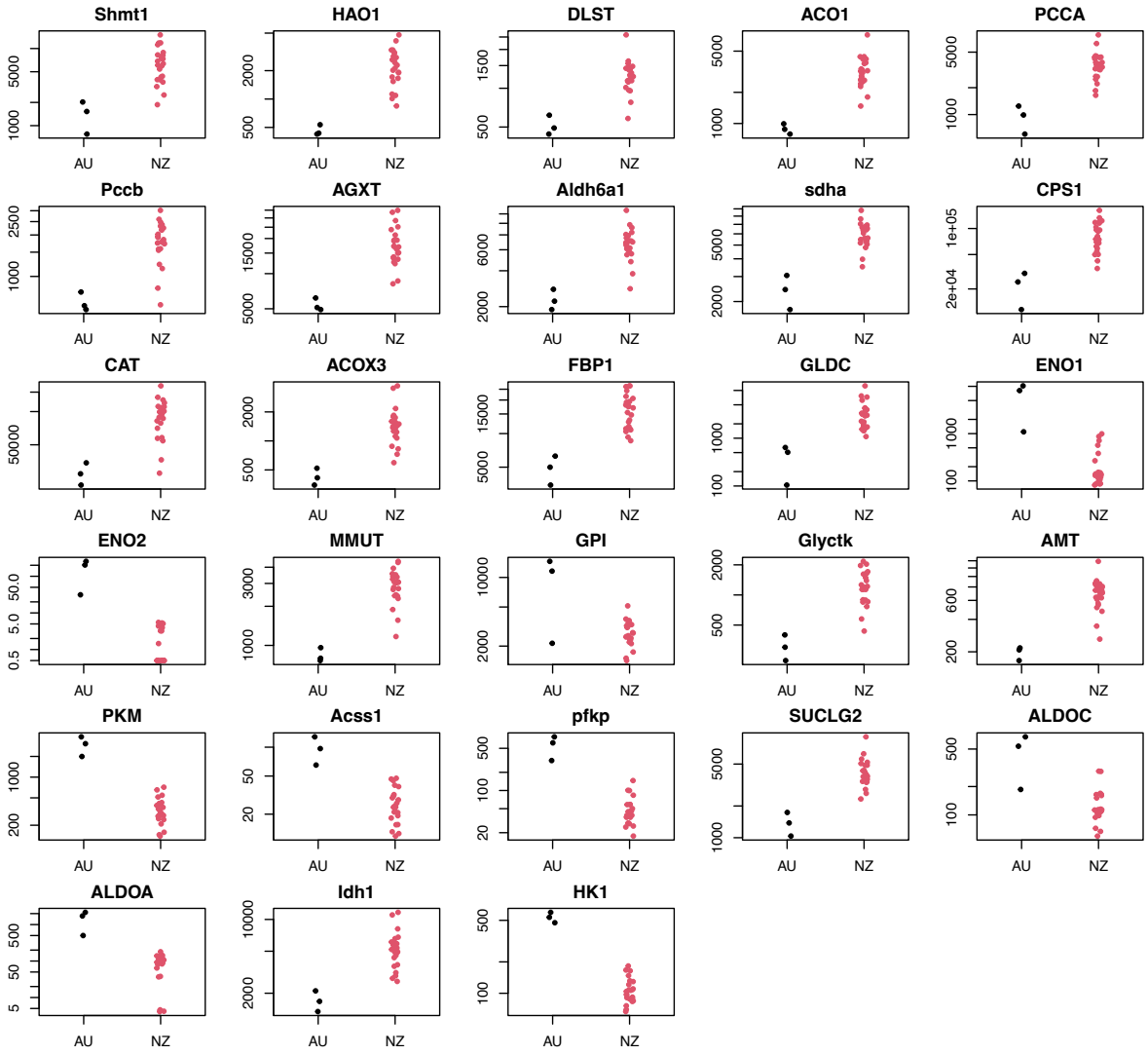


Figure 15: Genes associated with the carbon metabolism pathway differentially expressed in brushtail possums from western Australian (black) compared to New Zealand (red). Each dot corresponds to the normalized counts of the upregulated genes for each liver RNASeq sample (LFC <-2 & FDR-adjusted p-value < 0.00001).

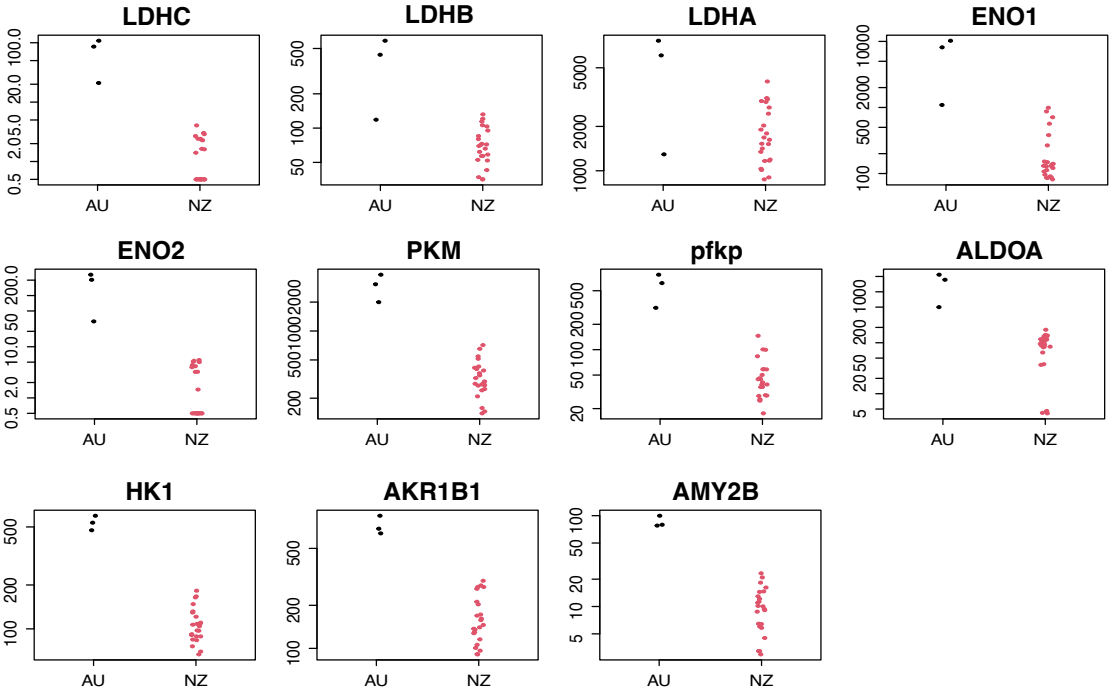


Figure 16: Genes associated with tricarboxylic cycle (TCA) differentially expressed in brushtail possums from western Australian (black) compared to New Zealand (red). Each dot corresponds to the normalized counts of the upregulated genes for each liver RNASeq sample (LFC <-2 & FDR-adjusted p-value < 0.00001).

Tables of read counts and gene associated p-values for DESeq2 and limma analysis can be found on the data repository: https://figshare.com/s/4e0f2dc7c9a05e441e03

Discussion

My objective was to identify gene expression associated with the western Australian phenotype and investigate their possible involvement in toxin resistance. Using three different approaches I identified differentially expressed genes associated with very different pathways in relation to cellular signalling, encapsulating structure, cell mobility and tricarboxylic (TCA) cycle.

Despite good mapping results (prior to differential expression analysis) many reads mapped to non-annotated parts of the possum genome. The presence of a significant quantity of unmapped reads in two samples from Western Australia can be attributed to incomplete execution of the poly-A capture step during the library preparation process. The notable increase in ribosomal gene counts identified in the differential expression analysis can also be accounted for by variations in library preparation, particularly during the rRNA depletion stage (Tellier and Murphy 2020). The RNA library preparation performed by AGRF did not include poly-A

capture, a rRNA removal technique, in contrast to AZENTA (Custom Science), Otago Genomic, and BGI Genomics. These libraries were consequently excluded from subsequent analyses.

The use of three approaches that used different pipelines to compute differential analysis was useful in confirming the overall picture, using two well-known and proven software Limma and DESeq2 especially with low rate of false positive (Seyednasrollah, Laiho, and Elo 2013) and WGCNA using drastically different approach (Langfelder and Horvath 2008). More than half the DEGs found with each approach was shared with the other and similar to the GO terms, displaying a strong level of similarity and confirmation. Some unique genes and pathways identified were specific to the approach employed which broaden the result to some extent.

Brushtail possums are arboreal herbivores and across their geographic range, they interact with and eat regional floras in ecologically distinct regions of Australia. The species lives in hot, dry regions, wet tropical and cool temperate forests (Kerle 1984; Carmelet-Rescan et al. 2022). New Zealand possums originated from Eastern Australia and Tasmania, so adapted to temperate coastal climates and sub-tropical forests which have a more consistent rainfall throughout the year than southwest Australia. These different environments likely result in differences in food availability, temperature fluctuations, and exposure to pathogens and parasites. As expected, numerous genes were identified as showing significant expression level differences between the western Australia and New Zealand population samples, although their adaptive significance is unclear. When considering gene properties, four major differences were detected from enrichment of gene ontology terms and are briefly highlight here:

1. Gene ontology (GO) terms related to glycosaminoglycan binding, ECM-receptor interaction, and proteoglycans in cancer were enriched which suggests divergence related to extracellular matrix interactions.
2. GO terms related to semaphorin receptor binding (Alto and Terman 2017), rap1 signalling pathway, leucocyte trans-endothelial migration, chemotaxis, and regulation of cell migration were enriched suggesting differences among possum populations related to cell signalling and migration.
3. The enrichment of GO terms related to DNA-binding transcription activator activity and NF-kappa B signalling pathway (Baltimore 2009) is as expected with divergent gene expression and response to environmental stressors.

4. The enrichment of GO terms related to external encapsulating structure organization, anatomical structure development, and external encapsulating structure in liver suggest physiological/morphological adaptation.

Overall, these significant differences in gene expression are as expected of separate brushtail possum populations (subspecies) that have undergone adaptations to their unique environmental conditions, including differences in vegetation cover, heat drought and pathogens and parasite pressure.

Candidate genes for tolerance of toxins

Cellular and metabolic functions form pathways that have been mapped to produce networks of molecular interactions and reactions and these are linked to genes functions and products within the KEGG database. KEGG pathway enrichment provided an opportunity to identify candidate genes for tolerance of plant toxins. The toxin sodium fluoroacetate interferes with the tricarboxylic acid (TCA) cycle by inhibiting the enzyme aconitase. This inhibition leads to a build-up of citrate, which is supposed to be converted to oxaloacetate by the aconitase, depleting cellular energy stores and ultimately causing death (Goncharov, Jenkins, and Radilov 2006). Examination of the KEGG pathway enrichment analysis on all genes with adjusted p-value under 0.00001 shows a strong association with carbon metabolism with pathways such as “Glyoxylate and dicarboxylate metabolism”, “Carbon metabolism”, “Propanoate metabolism” and more. This might be relevant to exposure to sodium fluoroacetate which disrupts the main energy source (tricarboxylic acid cycle) and thus regulation of genes associated with carbon metabolism is of importance. The differential expression of these pathways suggests it might include the genes responsible for tolerance of sodium fluoroacetate. Examining the pathways and associated genes in more detail provide some support for this hypothesis.

- Propanoate metabolism: One of the explored potential therapy/treatments of fluoroacetate (FA) poisoning is using competing substances to bind with CoA one of them being the propionate (Goncharov, Jenkins, and Radilov 2006). The propanoate metabolism pathway consists of all the genes associated with incorporating propanoate into the TCA and the first step consists of propionate to propionyl-CoA binding the molecule to CoA.
- Glyoxylate and dicarboxylate metabolism: The glyoxylate cycle bypasses the CO2-generating steps of the TCA cycle and allows organisms to synthesize malate from glyoxylate using malate synthase (Figure 17). This pathway could potentially be

upregulated in organisms that are resistant to sodium fluoroacetate, allowing them to use fluoroacetyl-CoA instead of acetyl-CoA to synthesise malate (fluoromalate) (Marletta, Srere, and Walsh 1981; Powell and Beevers 1968).

- The genes ENO1, ENO2, PKM, PFKP, ALDOA, and HK1 are all strongly upregulated in western Australian samples and encode for enzymes involved in the glycolytic pathway. Enolase (ENO) catalyses the conversion of 2-phosphoglycerate to phosphoenolpyruvate, while pyruvate kinase (PKM) catalyses the conversion of phosphoenolpyruvate to pyruvate, generating ATP in the process. Phosphofructokinase (PFKP) and aldolase A (ALDOA) are also involved in the early stages of glycolysis, while hexokinase 1 (HK1) catalyses the first step of glycolysis, the conversion of glucose to glucose-6-phosphate. The upregulation of these genes in response to sodium fluoroacetate exposure could be an attempt to shift energy production towards glycolysis, bypassing the TCA cycle and its inhibition by the toxin.
- AKR1B1 encodes for aldose reductase, an enzyme that catalyses the conversion of glucose to sorbitol and AMY2B encodes for pancreatic alpha-amylase, an enzyme involved in the breakdown of starch and glycogen. Their upregulation in response to sodium fluoroacetate exposure is not immediately clear, but it is possible that it could play a role in detoxification, energy production or the mobilization of energy reserves in response to the toxin.
- The genes LDHC, LDHB, and LDHA encode for the isoenzymes of lactate dehydrogenase (LDH), which catalyse the reversible conversion of pyruvate to lactate, or lactate to pyruvate, depending on the energy needs of the cell. LDH is an important enzyme in the metabolism of glucose, which is converted to pyruvate through glycolysis, and then further metabolized to acetyl-CoA via the TCA cycle. It is possible that the upregulation of LDH isoforms in response to sodium fluoroacetate exposure could be an attempt to compensate for the inhibition of the TCA cycle.
- ACSS1 gene is upregulated in the western Australian sample, ACSS1 code for the acetyl-CoA synthetase, an enzyme that is involved in the production of acetyl-CoA and is also involved in the synthesis of fluoroacetyl-CoA the first step of sodium fluoroacetate intoxication. This upregulation could mean that either fluoroacetyl-CoA is dealt with as we hypothesize through the glyoxylate shunt or that this version of the gene is different in the western Australian possums and synthesises acetyl-CoA faster than fluoroacetyl-CoA avoiding the inhibition of aconitase and the build-up of citrate.

Overall, the combination of genes differentially regulated and associated with the TCA cycle strongly suggests that this regulation is associated with sodium fluoroacetate resistance from the western Australian brushtail possums.

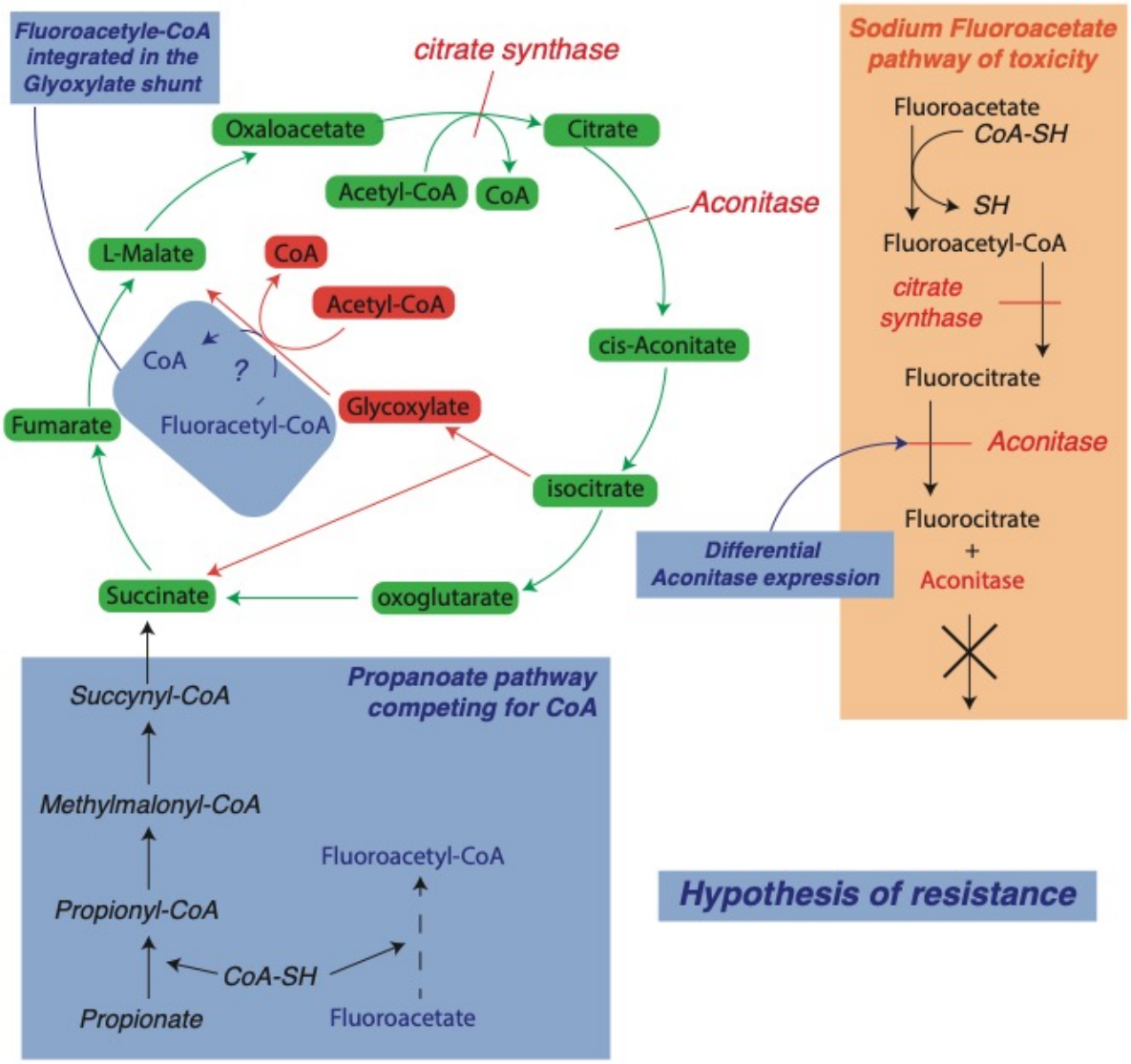


Figure 17: Molecular pathway associated with sodium fluoroacetate poisoning and hypothesis of resistance associated with the differential expression results from RNA-Seq data from western Australian brushtail possum liver.

In order to validate the results from those differential expression analyses and especially targeting particular overexpressed genes it would be interesting to use qRT-PCR (Bustin et al. n.d.) and look at the correlation it displays with RNA-Seq fold changes and assess the accuracy and reliability of the differential expression analysis.

It has been suggested that toxin resistance results from a difference in the aconitase gene ACO2, this hypothesis was tested on tammar wallaby trying to identify sequence differences without success (Deakin et al. 2013). In our data, the gene ACO1 is present and displays some evidence of overexpression in New Zealand samples (Figure 15) suggesting that aconitase expression and regulation could be at work in western Australian possums. It is probable that sodium fluoroacetate tolerance in possums, like many traits that are adaptive, is not controlled by a single gene, but rather is the result of the combined effects of multiple genes as the TCA cycle regulation is a highly polygenic trait (Barghi, Hermisson, and Schlötterer 2020; Margres et al. 2017).

The current results inform us that, like the phylogeny suggesting (Carmelet-Rescan et al. 2022), there are strong differences in gene expression between the two populations. This high amount of difference and the polygenic aspect of toxin resistance tend to indicate that toxin resistance would take a while to evolve in New Zealand despite the high selective pressure the brushtail possums are subjected to. A study designed to investigate New Zealand resistance could use Steward Island possums (not subject to sodium fluoroacetate poisoning), as it was intended here but keeping the same platform for sequencing, to look for differentially expressed genes common to the one found in this chapter between western Australian and New Zealand possums.

Future studies could analyse differential expression between mainland New Zealand exposed and Stewart Island unexposed brushtail possums, focusing on the candidate pathway identified here. Sampling brushtail possums from Stewart Island was realized here with the intent of comparing them to mainland New Zealand ones trying to see if 1080 exposure happening on the mainland and not on Stewart Island was already noticeable in the expression pattern of the possums. However, the use of a different sequencing platform because of simplicity and better results has led to a perfect correlation between the sequencing platform and the origin. This correlation means the FSQN step (Feature Specific Quantile Normalization) cannot be applied and therefore it is impossible to separate the variance associated with different sequencing platforms and the variance associated with the differences within the populations. Additional mRNA sequences generated for possum samples from the mainland using the sequencing platform used for the Stewart Island samples would allow the incorporation of all data to investigate differential gene expression between populations exposed and unexposed to 1080 in New Zealand.

An experimental approach, which involves LD50 testing and examining the expression patterns related to differential exposure in a controlled setting, is a highly potent method. However, its feasibility and ethical implications raise significant questions.

In conclusion, this study shows western Australian brushtail possums (*Trichosurus vulpecula hypoleucus*) have numerous genes that are differentially expressed compared to the New Zealand population. Many have the potential to be associated with local adaptation to their local environments. I have identified candidate genes that are upregulated and have the potential to be associated with sodium fluoroacetate tolerance (Twigg et al. 2003). These results can be further investigated with other analyses such qRT-PCR but are an important first step to study sodium fluoroacetate tolerance in other species and could be useful to look at genes and pathways that could be associated to the adaptation of mammal species to dry conditions. In addition to expression level data, the RNA-Seq data also provide sequence data that can be explored to look for evidence of selection in gene sequences.

References

Alto, Laura Taylor, and Jonathan R. Terman. 2017. "Semaphorins and Their Signaling Mechanisms." *Methods in Molecular Biology* 1493: 1–25. /pmc/articles/PMC5538787/ (February 24, 2023).

Adamczyk, J. J., D. D. Hardee, L. C. Adams, and D. V. Sumerford. 2001. "Correlating Differences in Larval Survival and Development of Bollworm (Lepidoptera: Noctuidae) and Fall Armyworm (Lepidoptera: Noctuidae) to Differential Expression of Cry1A (c) δ-Endotoxin in Various Plant Parts among Commercial Cultivars of Transgenic ." Journal of Economic Entomology 94(1): 284–90. https://dx.doi.org/10.1603/0022-0493-94.1.284 (September 22, 2023).

Andrews, Simon, and others. 2010. "FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010." *Https://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/* 1(1): http://www.bioinformatics.babraham.ac.uk/projects/. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/%0Ahttp://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/ (March 17, 2020).

Baltimore, David. 2009. "Discovering NF-KappaB." *Cold Spring Harbor perspectives in biology* 1(1). https://cshperspectives.cshlp.org/content/1/1/a000026.short (March 1, 2023).

Barghi, Neda, Joachim Hermisson, and Christian Schlötterer. 2020. "Polygenic Adaptation: A Unifying Framework to Understand Positive Selection." *Nature Reviews Genetics* 21(12): 769–81. https://www.nature.com/articles/s41576-020-0250-z (March 2, 2023).

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57(1): 289–300.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30(15): 2114–20. https://academic.oup.com/bioinformatics/article-abstract/30/15/2114/2390096 (March 17, 2020).

Brown, Thomas M., and Gregory T. Payne. 1988. "Experimental Selection for Insecticide Resistance." *Journal of economic entomology* 81(1): 49–56.

Bustin, SA, V Benes, … T Nolan - Journal of molecular, and Undefined 2005. "Quantitative Real-Time RT-PCR–a Perspective." *jme.bioscientifica.com*. https://jme.bioscientifica.com/view/journals/jme/34/3/0340597.xml (March 3, 2023).

Byrom, Andrea E. et al. 2015. "Assessing Movements of Brushtail Possums (Trichosurus Vulpecula) in Relation to Depopulated Buffer Zones for the Management of Wildlife Tuberculosis in New Zealand." *PLoS ONE* 10(12): e0145636. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0145636 (March 20, 2023).

Campbell, Catriona D. et al. 2021. "Has the Introduction of Two Subspecies Generated Dispersal Barriers among Invasive Possums in New Zealand?" *Biological Invasions* 23(12): 3831–45. https://link.springer.com/article/10.1007/s10530-021-02609-1 (March 20, 2023).

Carmelet-Rescan, David, Mary Morgan-Richards, Nimeshika Pattabiraman, and Steven A. Trewick. 2022. "Time-Calibrated Phylogeny and Ecological Niche Models Indicate Pliocene Aridification Drove Intraspecific Diversification of Brushtail Possums in Australia." *Ecology and Evolution* 12(12): e9633. https://onlinelibrary.wiley.com/doi/full/10.1002/ece3.9633 (February 15, 2023).

Costa-Silva, Juliana, Douglas Domingues, and Fabricio Martins Lopes. 2017. "RNA-Seq Differential Expression Analysis: An Extended Review and a Software Tool." *PLoS ONE* 12(12): e0190152. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0190152 (July 5, 2022).

Cowan, P. 2016. "Characteristics and Behaviour of Brushtail Possums Initially Moving into a Depopulated Area." *New Zealand Journal of Zoology* 43(3): 223–33.

Cowan, P. E. 1992. "The Eradication of Introduced Australian Brushtail Possums, Trichosurus Vulpecula, from Kapiti Island, a New Zealand Nature Reserve." *Biological Conservation* 61(3): 217–26. https://www.sciencedirect.com/science/article/pii/000632079291119D (March 15, 2020).

Deakin, Janine E. et al. 2013. "Towards an Understanding of the Genetic Basis behind 1080 (Sodium Fluoroacetate) Tolerance and an Investigation of the Candidate Gene ACO2." *Australian Journal of Zoology* 61(1): 69–77. http://www.publish.csiro.au/?paper=ZO12108 (March 15, 2020).

Efford, Murray, Bruce Warburton, and Nick Spencer. 2000. "Home-Range Changes by Brushtail Possums in Response to Control." *Wildlife Research* 27(2): 117–27. http://www.publish.csiro.au/WR/WR99005 (March 15, 2020).

Franks, Jennifer M., Guoshuai Cai, and Michael L. Whitfield. 2018. "Feature Specific Quantile Normalization Enables Cross-Platform Classification of Molecular Subtypes Using Gene Expression Data." *Bioinformatics* 34(11): 1868–74. https://academic.oup.com/bioinformatics/article/34/11/1868/4816109 (January 20, 2023).

Goncharov, Nikolay V., Richard O. Jenkins, and Andrey S. Radilov. 2006. "Toxicology of Fluoroacetate: A Review, with Possible Directions for Therapy Research." *Journal of Applied Toxicology* 26(2): 148–61. https://onlinelibrary.wiley.com/doi/full/10.1002/jat.1118 (February 16, 2023).

Gupta, RC. 2015. "Handbook of Toxicology of Chemical Warfare Agents." https://books.google.com/books?hl=fr&lr=&id=MXKDBAAAQBAJ&oi=fnd&pg=PP1&dq=Goncharov+N,+et+al.+2015+Fluoracetate.+Chapter+16,+pages+193–214+in+Gupta+RC+(ed)+Handbook+of+Toxicology+of+Chemical+Warfare+Agents+(2nd+edition).+Academic+Press.+978-0-12-800159-2&ots=L0zowzfqi3&sig=zkRzHn12-_2tCGiRS4zNQ2OIWOM (March 13, 2020).

How, R A, and J A Kerle. 1995. *Common Brushtail Possum,. In 'The Mammals of Australia'.(Ed. R. Strahan.) Pp. 273--275*. Reed Books: Sydney.

Institute, Broad. 2019. "'Picard Toolkit', Broad Institute, GitHub Repository." *Picard Toolkit*.

Jenkins, Amy et al. 2015. "Differential Expression and Roles of Staphylococcus Aureus Virulence Determinants during Colonization and Disease." mBio 6(1). https://journals.asm.org/doi/10.1128/mbio.02272-14 (September 22, 2023).

Kerle, A., G. M. McKay, and G. B. Sharman. 1991. "A Systematic Analysis of the Brushtail Possum, Trichosurus Vulpecula (Kerr, 1792) (Marsupialia: Phalangeridae)." *Australian Journal of Zoology* 39(3): 263–71. https://www.publish.csiro.au/zo/zo9910313 (November 23, 2021).

Kerle, A. 1984. "Variation in the Ecology of Trichosurus: Its Adaptive Significance." *Possums and gliders*: 115–28.

Kim, Daehwan et al. 2019. "Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype." *Nature Biotechnology 2019 37:8* 37(8): 907–15. https://www.nature.com/articles/s41587-019-0201-4 (July 4, 2022).

Kuznetsova, Irina et al. 2019. "CirGO: An Alternative Circular Way of Visualising Gene Ontology Terms." *BMC Bioinformatics* 20(1): 1–7. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2671-2 (February 14, 2023).

Langfelder, Peter, and Steve Horvath. 2008. "WGCNA: An R Package for Weighted Correlation Network Analysis." *BMC Bioinformatics* 9(1): 1–13. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559 (December 19, 2022).

Larsson, Johan, and Peter Gustafsson. 2018. "A Case Study in Fitting Area-Proportional Euler Diagrams with Ellipses Using Eulerr." In *Proceedings of International Workshop on Set Visualization and Reasoning*, , 84–91. https://cran.r-project.org/package=eulerr.

Leong, Lex Ee Xiang et al. 2017. "Fluoroacetate in Plants - a Review of Its Distribution, Toxicity to Livestock and Microbial Detoxification." *Journal of Animal Science and Biotechnology* 8(1): 1–11. https://jasbsci.biomedcentral.com/articles/10.1186/s40104-017-0180-6 (April 22, 2022).

Li, Heng et al. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25(16): 2078–79. https://academic.oup.com/bioinformatics/article/25/16/2078/204688 (February 10, 2023).

Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. "FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features." *Bioinformatics* 30(7): 923–30. https://academic.oup.com/bioinformatics/article/30/7/923/232889 (July 4, 2022).

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15(12): 1–21. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8 (July

4, 2022).

Margres, Mark J. et al. 2017. "Quantity, Not Quality: Rapid Adaptation in a Polygenic Trait Proceeded Exclusively through Expression Differentiation." *Molecular Biology and Evolution* 34(12): 3099–3110. https://academic.oup.com/mbe/article/34/12/3099/4097632 (March 2, 2023).

Marletta, Michael A, Paul A Srere, and Christopher Walsh. 1981. "Stereochemical Outcome of Processing of Fluorinated Substrates by ATP Citrate Lyase and Malate Synthase." *Biochemistry* 20(13): 3719–23.

McIlroy, J. C. 1983. "The Sensitivity of the Brushtail Possum (Trichosurus Vulpecula) to 1080 Poison (Sodium Monfluoroacetate)." *New Zealand Journal of Ecology* 6: 125–31. https://www.jstor.org/stable/24052734 (March 15, 2020).

Mead, R. J., A. J. Oliver, D. R. King, and P. H. Hubach. 1985. "The Co-Evolutionary Role of Fluoroacetate in Plant-Animal Interactions in Australia." *Oikos* 44(1): 55.

Müllner, Daniel. 2013. "Fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python." *Journal of Statistical Software* 53(9): 1–18. https://cran.r-project.org/package=fastcluster (December 19, 2022).

Nugent, G. et al. 2012. "Bait Aggregation to Reduce Cost and Toxin Use in Aerial 1080 Baiting of Small Mammal Pests in New Zealand." *Pest Management Science* 68(10): 1374–79. https://onlinelibrary.wiley.com/doi/full/10.1002/ps.3315 (February 15, 2023).

Nugent, G., B. M. Buddle, and G. Knowles. 2015. "Epidemiology and Control of Mycobacterium Bovis Infection in Brushtail Possums (Trichosurus Vulpecula), the Primary Wildlife Host of Bovine Tuberculosis in New Zealand." *New Zealand Veterinary Journal* 63(sup1): 28–41. https://pubmed.ncbi.nlm.nih.gov/25290902/ (March 20, 2023).

Nugent, G., P. Sweetapple, J. Coleman, and P. Suisted. 2000. "Possum Feeding Patterns: Dietary Tactics of a Reluctant Folivore." In *The Brushtail Possum: Biology, Impact and Management of an Introduced Marsupial*, , 10–19.

Oliver, A. J., and D. R. King. 1979. "Fluoroacetate Tolerance, a Genetic Marker in Some Australian Mammals." *Australian Journal of Zoology* 27(3): 331–47. https://www.publish.csiro.au/zo/zo9790363 (April 22, 2022).

Pattabiraman, Nimeshika, Mary Morgan-Richards, Ralph Powlesland, and Steven A. Trewick. 2021. "Unrestricted Gene Flow between Two Subspecies of Translocated Brushtail Possums (Trichosurus Vulpecula) in Aotearoa New Zealand." *Biological Invasions* 24(1): 247–60. https://link.springer.com/article/10.1007/s10530-021-02635-z (May 31, 2022).

Powell, Gary L., and Harry Beevers. 1968. "Fluoroacetyl-CoA as a Substrate for Malate

Synthase." *BBA - Enzymology* 151(3): 708–10.

Pracy, L. 1974. "Opposums." *New Zealand Nature Heritage* 3(32): 873–82.

Ritchie, Matthew E. et al. 2015. "Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies." *Nucleic Acids Research* 43(7): e47. /pmc/articles/PMC4402510/ (December 19, 2022).

Robles, José A. et al. 2012. "Efficient Experimental Design and Analysis Strategies for the Detection of Differential Expression Using RNA-Sequencing." *BMC Genomics* 13(1): 1–14. https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-13-484 (July 4, 2022).

Ross, James G., Kathryn Bicknell, and G.J. Hickling. 1999. "COST-EFFECTIVE CONTROL OF 1080 BAIT-SHY POSSUMS."

Seyednasrollah, Fatemeh, Asta Laiho, and Laura L. Elo. 2013. "Comparison of Software Packages for Detecting Differential Expression in RNA-Seq Studies." Briefings in Bioinformatics 16(1): 59–70. https://academic.oup.com/bib/article/16/1/59/240754 (December 19, 2022).

Supek, Fran, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. 2011. "Revigo Summarizes and Visualizes Long Lists of Gene Ontology Terms." *PLoS ONE* 6(7): e21800. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021800 (February 14, 2023).

Taylor, Andrea C. et al. 2004. "High Microsatellite Diversity and Differential Structuring among Populations of the Introduced Common Brushtail Possum, Trichosurus Vulpecula, in New Zealand." *Genetical Research* 83(2): 101–11. https://www.cambridge.org/core/journals/genetics-research/article/high-microsatellite-diversity-and-differential-structuring-among-populations-of-the-introduced-common-brushtail-possum-trichosurus-vulpecula-in-new-zealand/AFB7EE781A5EF58A43F048B05EABEAEA (April 29, 2023).

Todd, Erica V., Michael A. Black, and Neil J. Gemmell. 2016. "The Power and Promise of RNA-Seq in Ecology and Evolution." *Molecular Ecology* 25(6): 1224–41. https://onlinelibrary.wiley.com/doi/full/10.1111/mec.13526 (July 12, 2022).

Twigg, L. E. et al. 1996. "Fluoroacetate Content of Some Species of the Toxic Australian Plant Genus, Gastrolobium, and Its Environmental Persistence." *Natural Toxins* 4(3): 122–27. http://doi.wiley.com/10.1002/19960403NT4 (March 15, 2020).

———. 2003. "Sensitivity of Some Australian Animals to Sodium Fluoroacetate (1080): Additional Species and Populations, and Some Ecological Considerations." *Australian*

Journal of Zoology 51(5): 515–31. https://www.publish.csiro.au/zo/zo03040 (March 23, 2022).

Twigg, L. E., and D. R. King. 1991. "The Impact of Fluoroacetate-Bearing Vegetation on Native Australian Fauna: A Review." *Oikos* 61(3): 412. https://www.jstor.org/stable/3545249?casa_token=wxSjPb9PE-QAAAAA:AqA216Efy0uOSzZJwBTgcg7Eo64RP0a9Hmm8Ua_rf4Faa-Qp-yot16PIkohRXPufbktoxRP5zU0RpZ0KIa9YOsEXHe5Co014rR-2rRyJe0hgP43FRiJi (March 15, 2020).

Wallenius, Kenneth T. 1963. "Biased Sampling: The Noncentral Hypergeometric Probability Distribution." https://apps.dtic.mil/sti/citations/AD0426243 (February 14, 2023).

Young, Matthew, Maintainer Nadia Davidson, Anthony Hawkins, and G O biocViews Sequencing. 2013. "Package 'Goseq.'"

Yu, Guangchuang. 2021. "Enrichplot: Visualization of Functional Enrichment Result." https://yulab-smu.top/biomedical-knowledge-mining-book/.

Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. "ClusterProfiler: An R Package for Comparing Biological Themes among Gene Clusters." *Omics: a journal of integrative biology* 16(5): 284–87.

Zhang, Zong Hong et al. 2014. "A Comparative Study of Techniques for Differential Expression Analysis on RNA-Seq Data." *PLoS ONE* 9(8): e103207. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0103207 (July 5, 2022).

Chapter 5: Evidence of positive selection in the western Australian brushtail possum genome

Abstract

This chapter documents the results of selection analysis on the genome of brushtail possums (*Trichosurus vulpecula*), comparing the western lineage (*T. v. hypoleucus*) to the eastern lineage (*T. v. fuliginosus* and *T. v. vulpecula*; represented by sampling from New Zealand). The study uses gene sequences constructed with mapped reads from RNA-Seq data to identify genes under positive selection and the associated biological mechanisms associated with them. The main goal was to identify genetic pathways that might be involved in the resistance shown by South-west Australian possums to the plant toxin sodium fluoroacetate. Here I used four different approaches to identify genes under positive selection and enrichment analysis to identify the biological pathways associated with them. The set of candidate genes under positive selection identified showed significant overlap and were notably associated with glycan degradation and fatty acid metabolism. I made hypotheses on the association of the discovered enriched pathways with fluoroacetate resistance but the list of candidate genes is still long.

Introduction

Evolution is a continuous process that shapes the genetic makeup of populations over time. Natural selection acting on individual survival and reproductive success drives the accumulation of advantageous alleles within populations. Positive selection is the process by which beneficial alleles within specific protein-coding genes increase in frequency, while most nucleotide mutations in these genes tend to be either neutral or harmful with respect to amino acid translation. Positive selection results in the fixation of functional amino acid substitutions and can be identified through a comparison of synonymous (dS – nucleotide substitutions that do not result in amino acid replacements) and non-synonymous (dN – nucleotide substitutions that do result in amino acid replacements) mutations. The ratio of non-synonymous to synonymous substitutions (ω= dN/dS) is commonly used as an estimator of selective pressure on a gene when it exceeds 1 as this implies natural selection is causing changes in protein sequence (Kimura 1977; Kryazhimskiy and Plotkin 2008). Using phylogenetic information and ω calculated for any codon site, branch-site models can detect selection at the lineage level (Yang 2007). The use of dN/dS ratio to detect positive selection has led to numerous insights into the evolution of genes and genomes. For example, it has been used to identify genes that are evolving rapidly in response to changing environments, such as those involved in pathogen defence or reproduction (Ford 2001; Yang 1998). Additionally, it has been used to uncover the

molecular basis of adaptation in various species, shedding light on the mechanisms by which genetic diversity contributes to evolutionary change (e.g. Kapralov and Filatov 2007; Souza et al. 2022; Vermaak, Henikoff, and Malik 2005).

Brushtail possums (*Trichosurus vulpecula*) are widespread and can be found in diverse habitats. Additionally, possums are known for their tolerance to secondary plant metabolites (Efford, Warburton, and Spencer 2000; Pracy 1974, Chapters 1 & 2). In western Australia, brushtail possums have evolved resistance of the potent mammal toxin sodium fluoroacetate from their exposure to this molecule in the plants they eat (i.e. *Gastrolobium*). Natural selection has led to the LD50 of fluoroacetate for brushtail possums in southwestern Australia being 160 times greater than that of their cousins in east Australia (Leong et al. 2017; Twigg et al. 1996). The potential for the emergence of resistance in the translocated population of this species could pose a challenge to conservation efforts in New Zealand. In the previous chapter (Chapter 4) I investigated the physiological and metabolic mechanisms underlying sodium fluoroacetate resistance in brushtail possums, using differential expression analysis of 15,674 genes. Several over-expressed pathways were associated with the tricarboxylic acid cycle (TCA) and could therefore be linked with sodium fluoroacetate detoxification (Dearing and Cork 1999). However, the genetic basis of sodium fluoroacetate resistance remains largely unknown. In this study, I identify genes that are under positive selection in western Australian brushtail possums, and in particular those genes that might contribute to their ability to detoxify sodium fluoroacetate. The geographic isolation of *T. v. hypoleucus* in western Australia from other possum populations, as demonstrated by high genetic distances (Carmelet-Rescan et al. 2022, Chapter 2) would have facilitated the evolution and maintenance of toxin resistance here (Twigg and King 1991). I used the transcriptomes of western and eastern Australian brushtail possums and computed several positive selection analyses, including CodeML branch model (BS1-BS2), HyPhy FUBAR, HyPhy MEME, Tassel trait association analysis, and FST/GST values (Bradbury et al. 2007; Jeffares et al. 2015; Monier et al. 2021; Murrell et al. 2012, 2013) on each of the expressed genes. The resulting identification of genes with signatures of positive selection can provide insights into the molecular mechanisms underlying adaptation and evolutionary changes in populations, as well as identifying candidate genes associated with traits of interest, such as resistance to sodium fluoroacetate. Furthermore, those results can inform conservation efforts by providing information on the genetic basis of this resistance, which may lead to the development of more targeted and effective pest control measures in New Zealand.

Material and Method

This chapter utilizes data derived from a subset of samples previously discussed in Chapter 4, and the extraction and sequencing procedures for these samples were detailed in the preceding chapter. Prior to mapping, the quality of the sequencing reads was assessed using FastQC (Andrews and others 2010) and adapter sequenced and quality-based trimming was performed using Trimmomatic (Bolger, Lohse, and Usadel 2014) removing reads ≤ 20 bp in length and reads with a quality score of ≤ 20. Sequences were then aligned to the brushtail possum (*Trichosurus vulpecula*) genome (Genebank : mTriVul1.pri - GCA_011100635.1) using HISAT2 (Kim et al. 2019), a spliced read aligner, with default parameters. Samtools (Li et al. 2009) was then used to sort and format the outputting .bam alignment files, and the Picards tools (Institute 2019) to collect alignment quality metrics. The haplotype variants were called using the software FreeBayes (Garrison and Marth 2012) generating variant calling format files. Variants are then filtered based on different quality metrics using VCFtools and BCFtools (Danecek et al. 2021). A flowchart detailing this part of the method of this study is represented in Figure 1.

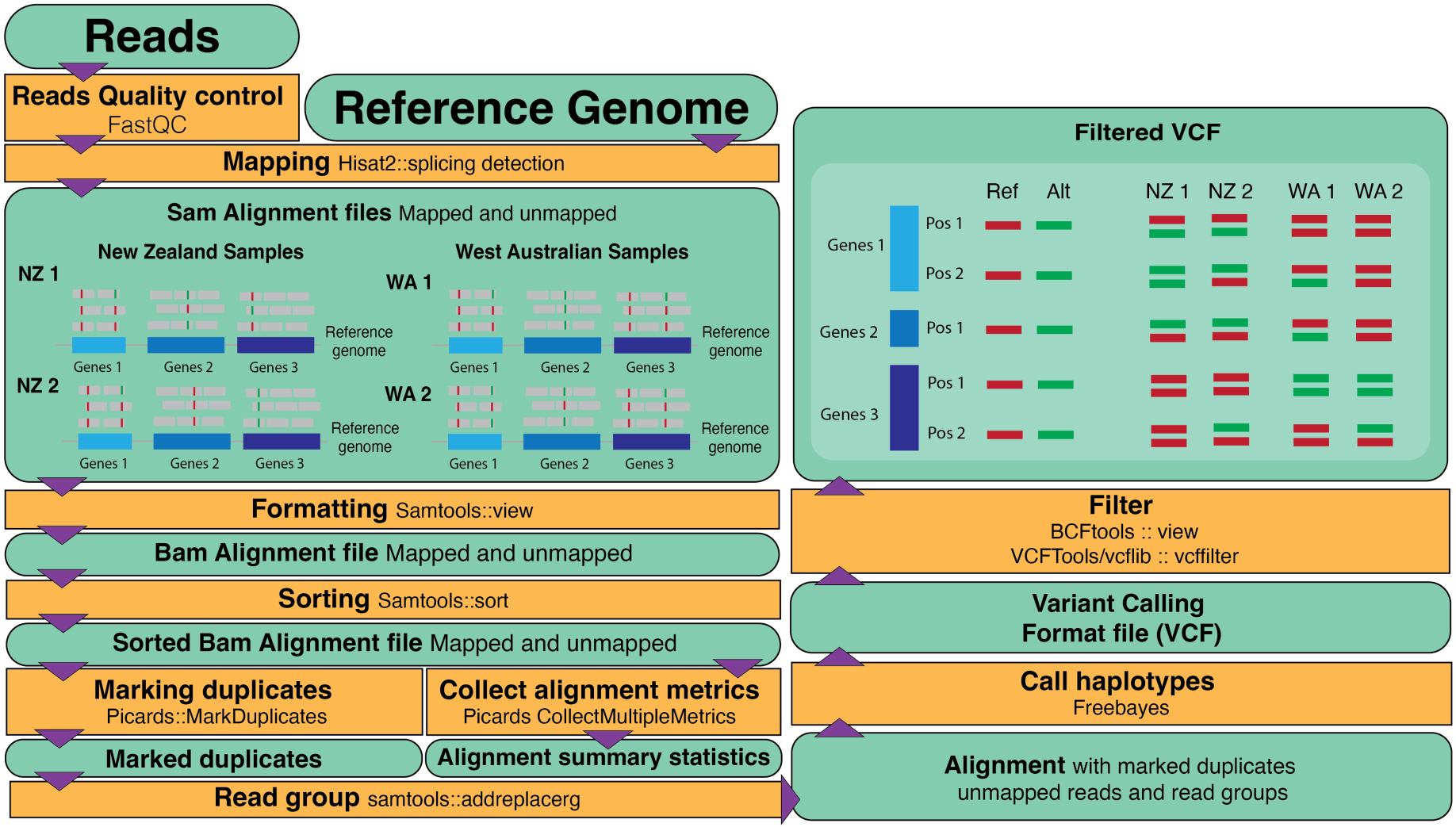
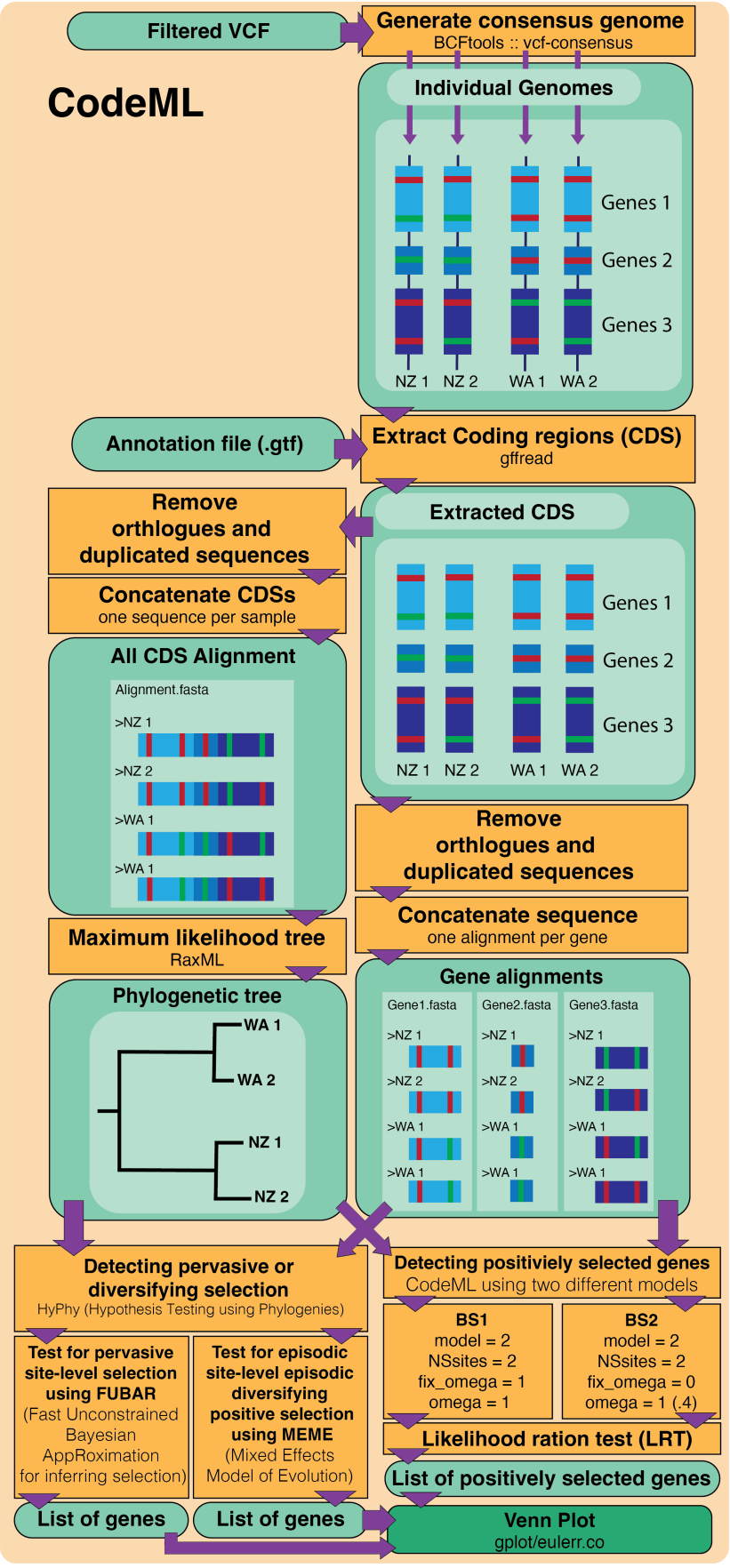


Figure 1: Strategy used to compute filtered variant calling files (VCF) for brushtail possums (*Trichosurus vulpecula*) gene sequences derived from RNASeq data of liver RNA.



In order to investigate genes under positive selection I used CodeML software from PAML using a branch-site model (Yang 2007). The branch-site model is a type of evolutionary analysis that allows for the detection of positive selection acting on specific lineages in a phylogeny. The model assumes that the ratio of the rate of nonsynonymous (dN) to synonymous (dS) substitutions can vary among branches in the phylogeny provided by the user. These are the branch-site neutral model (BS1) and the branch-site selection model (BS2) (Yang 2007). In this model, it is assumed that the dN/dS ratio may not only vary among different sites within the same coding sequence, but also across different lineages or branches of the given topology.

Figure 2: Strategy used to identify genes under positive selection in western (WA) and eastern (NZ) brushtail possum populations (*Trichosurus vulpecula*). The approach uses CodeML with variant data from transcriptomes derived from liver RNA.

A likelihood ratio test (LRT) is performed to see which of the two models is more probable for the given data. If the null hypothesis is rejected, it can be assumed that the dN/dS ratio is not the same when the branch of interest (foreground branch) is compared to other branches of the gene topology. In that case, the dN/dS ratio of the branch-of-interest will be checked to see if it is greater than 1, which would suggest positive selection acting on that lineage. In summary, by using CodeML and LRT tests, this study aimed to compare the likelihood scores of different models and assess the potential differences in selective pressures between western and eastern brushtail possums. Overall, the branch model is a powerful tool for investigating the evolutionary history of a group of organisms and detecting positive selection acting on specific lineages. The first step is to generate a consensus genome for each sample using BCFTools, and then extract only the coding region (CDS). A phylogeny is generated with RaxML (Stamatakis 2014) made with the alignment of concatenated of the CDS with each CDS define as a partition. JModeltest (Posada 2008) was used to determine the best substitution model to use. Using another approach, one alignment per gene was produced and used as input for the script to run CodeML using two different branch-model through all the alignments and then calculate a likelihood ratio to compute a p-value and identify positively selected genes. A flowchart detailing this part of the method of this study is represented in Figure 2.

In addition to CodeML the Hyphy platform (Kosakovsky Pond, Frost, and Muse 2005) was used to run two different models of selection: FUBAR and MEME (Murrell et al. 2012, 2013). The FUBAR model tests for pervasive (widespread across samples) site-level selection within each gene alignment, using a Bayesian Markov chain Monte Carlo (MCMC) algorithm to estimate the posterior probability of dN/dS ratios at each site in each gene alignment. Alignments containing sites showing evidence of selection are retained. MEME (Mixed Effects Model of Evolution) estimates site-wise synonymous and non-synonymous substitution rates in protein-coding DNA sequences and uses a two-category mixture model to detect positive selection on a subset of sites based on a likelihood ratio test and information aggregated over a proportion of branches at a site. Each gene alignment showing sites under selection with the MEME algorithm is retained. The R package eulerr (Larsson and Gustafsson 2018) was utilized to generate proportional Venn diagrams for the purpose of comparing the outcomes of the three distinct analyses (CodeML, Hyphy FUBAR and Hyphy MEME). This part of the methodology is summarised in Figure 2.

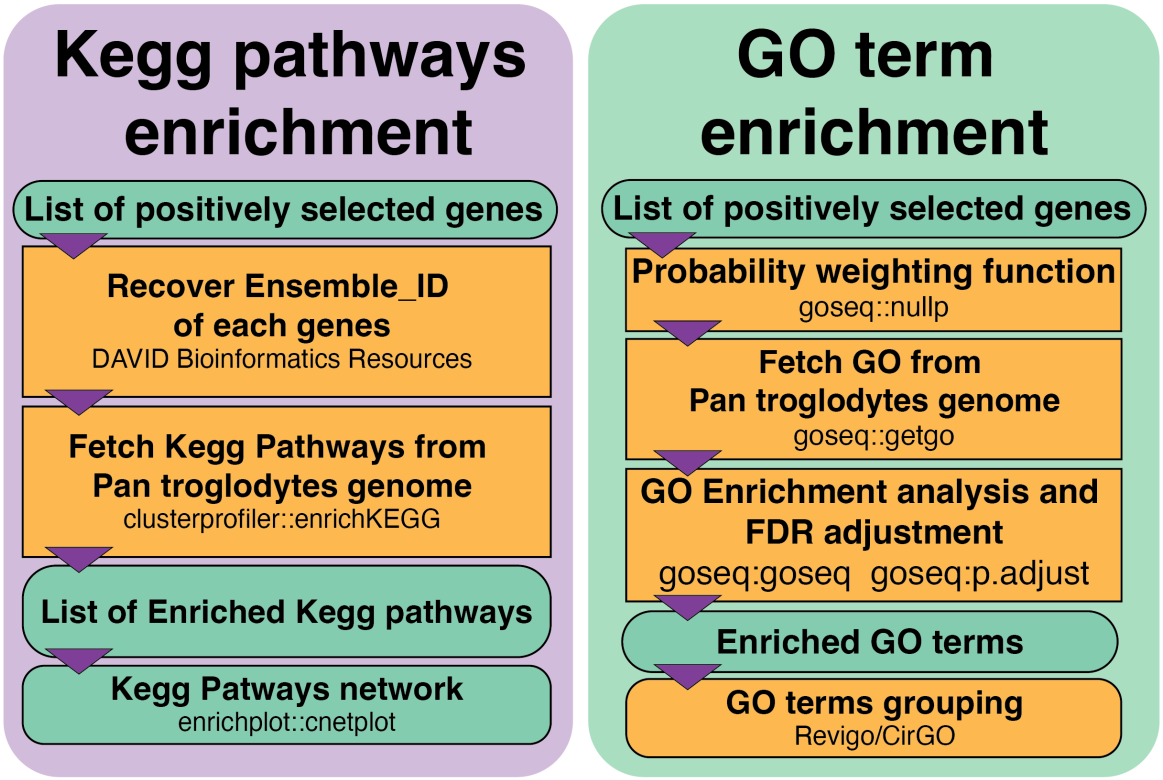


Figure 3: Strategy to identify genes under positive selection in western (WA) and eastern (NZ) brushtail possum populations (*Trichosurus vulpecula*).

As the *Trichosurus vulpecula* genome was not present in the Gene Ontologies (GOs) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway databases, I opted to use the genome of *Pan troglodytes* instead. This genome returned the highest number of identical genes (> 83%) allowing me to select the GO terms and KEGG pathways associated with candidate genes. To determine the enriched GO terms present in the list of significantly positively selected genes, I performed a Gene Ontology enrichment analysis using the "goseq" and "clusterProfiler" packages (Young et al. 2013; Yu et al. 2012), with the significance of enriched GO terms determined using the Wallenius non-central hypergeometric distribution (Wallenius 1963) and including Benjamini & Hochberg FDR control (Benjamini and Hochberg 1995). I then explored the resulting GO terms and their parent relations using REVIGO (Supek et al. 2011) and visualized them as circle charts using the program CirGO (Kuznetsova et al. 2019). To conduct the KEGG pathways enrichment analysis, I utilized the enrichKEGG function of the "clusterProfiler" R package (Yu et al. 2012), which includes FDR control. I then plotted the significantly enriched pathways (qvalue > 0.2), along with the associated under-expressed

and overexpressed genes, using the "enrichplot" R package (Yu 2021). Gene Ontology term and KeGG pathways enrichment analyses are also described using a flowchart in Figure 3.

I used F_{ST} as a marker of genetic differentiation of genes, with genes showing significantly strong differentiation likely to be under some kind of selection in one or the other population (Holsinger and Weir 2009). For this purpose, sliding windows F_{ST} computation was performed using vcftools and Weir and Cockerham's F_{ST} (Weir and Cockerham 1984). I used those results to compute a mean F_{ST} value per gene and used empirical distributions with the R package fistdistrplus (Delignette-Muller and Dutang 2015) to identify genes displaying significantly high F_{ST} values. In parallel to Weir and Cockerham's F_{ST}, Nei's Gst was calculated for each gene (Takahata and Nei 1984) using previously built fasta alignments and the R package mmod (Winter 2012) and significantly high Gst values were kept using an empirical distribution with the R package fistdistrplus.

Finally, trait analysis by association was performed using the software Tassel (Bradbury et al. 2007) and its R implementation rTassel (Monier et al. 2021). The analysis aims to associate loci with a population trait (quantitative trait loci: QTL) using a mixed linear model and linkage disequilibrium. The first step was to convert the vcf data into HapMap in order to import them into rTassel. After filtering out alleles with fewer than three representatives the mixed model analysis produced a list of significantly associated SNPs with population differences. The last step was to associate those QTL with specific genes. Tassel is a tool made to associate SNPs with traits, the analysis is designed to be associated with continuous traits. To fit the possum data to this analysis I associated the western lineage trait with a value of 1 and the other samples with a value of 0. A flowchart detailing this part of the method of this study is represented in Figure 4.

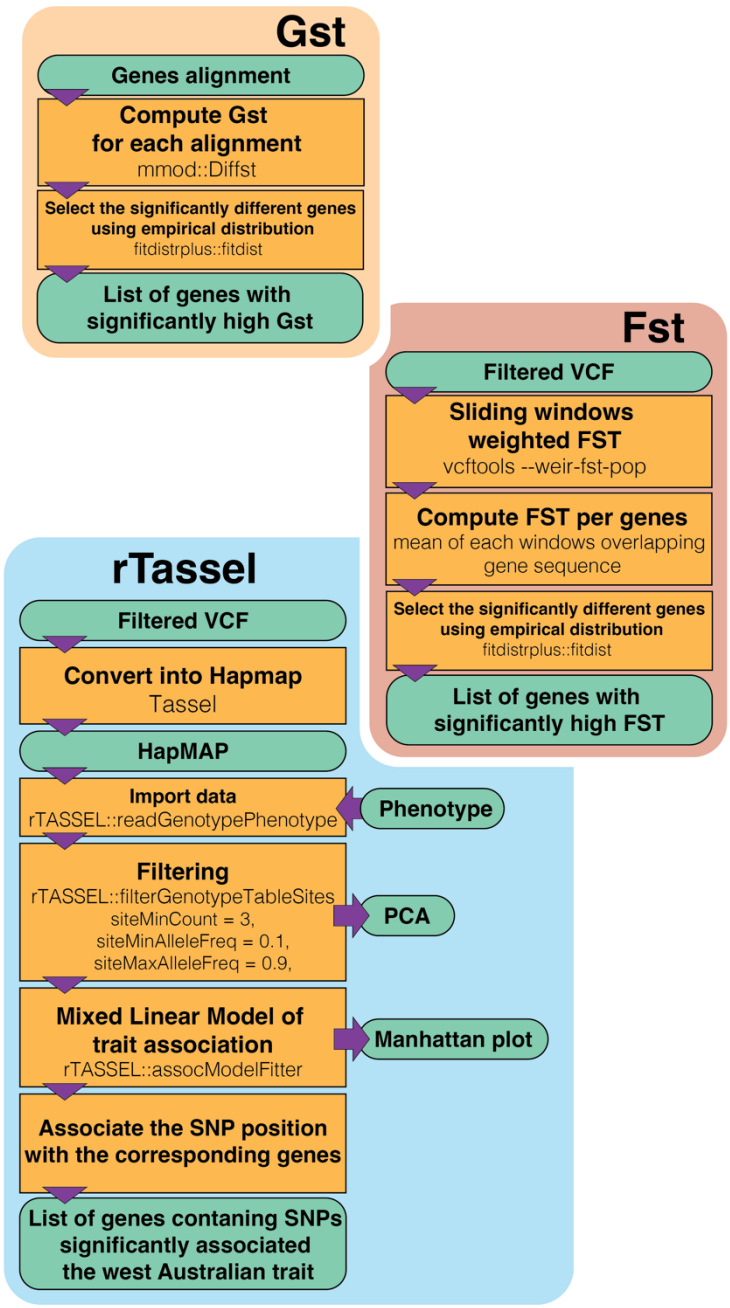


Figure 4: Strategy used to identify genes showing evidence of selection in variant data from western (WA) and eastern (NZ) brushtail possum lineage (*Trichosurus vulpecula*) derived from liver transcriptomes.

Results

Maximum likelihood phylogenetic analysis of 17355 genes revealed the major difference between western and eastern possums. Two western possums cluster together but a third (WA_3) was strongly differentiated from the others, leading to uncertainty in the tree topology (Figure 5). Four Stewart Island (ST) and three Manawatu (MN) samples in the analysis were each found to form monophyletic clusters. This tree topology was used for the subsequent branch-model analysis.

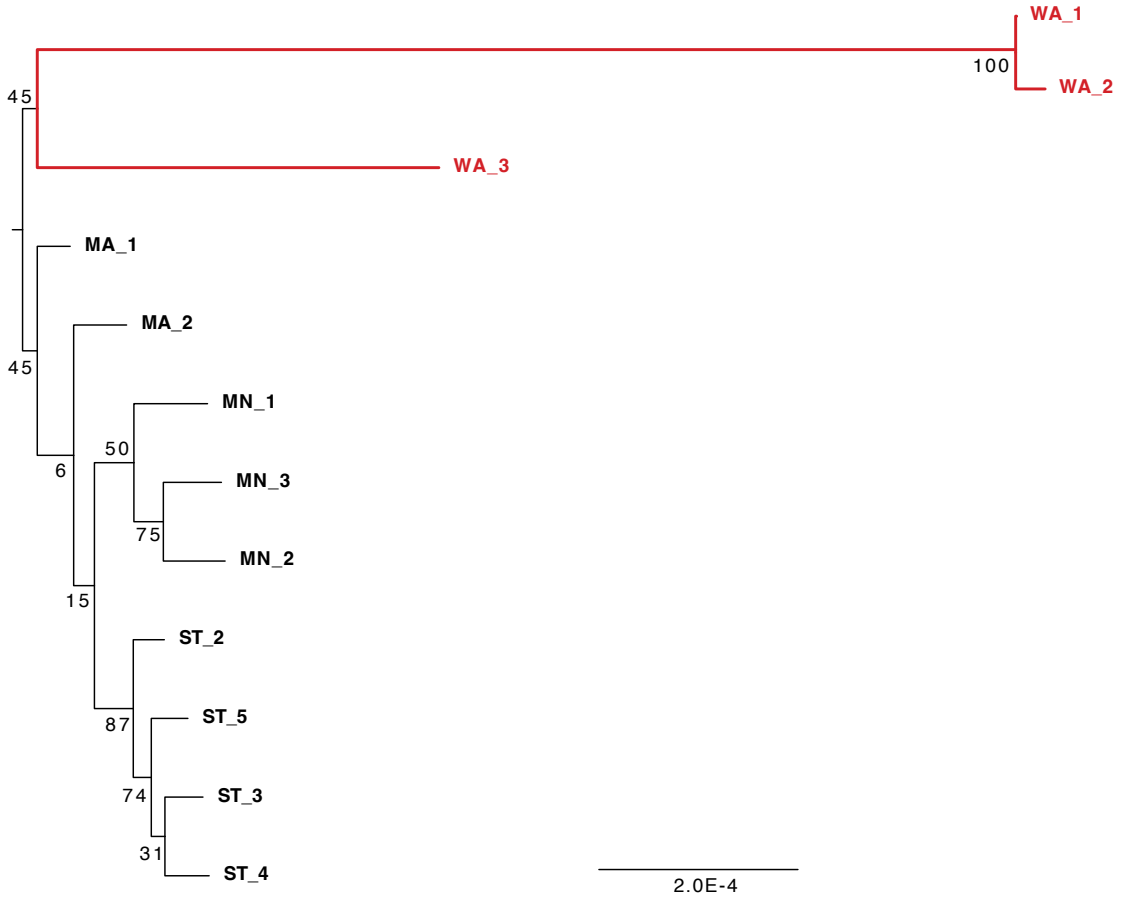


Figure 5: Maximum Likelihood (RAxML V.8) phylogeny of three western (red) and nine eastern (black) brushtail possums (*Trichosurus vulpecula*) based on 17600 concatenated genes sequences. Samples names correspond to the population of origin (WA: Western Australia, MA: Manaroa, ST: Stewart Island, MN: Manawatu) Bootstrap support from 1000 trees is indicated on each node.

CodeML

This analysis of selection of the 17355 genes using PAML CodeML with the BS1-BS2 branch-model revealed 47 genes showing evidence of positive selection along the genome (Figure 6, Appendix C5 Table 1). The LR value of PNPLA6 gene (Patatin Like Phospholipase Domain Containing 6) was much higher compared to all other genes (37.7).

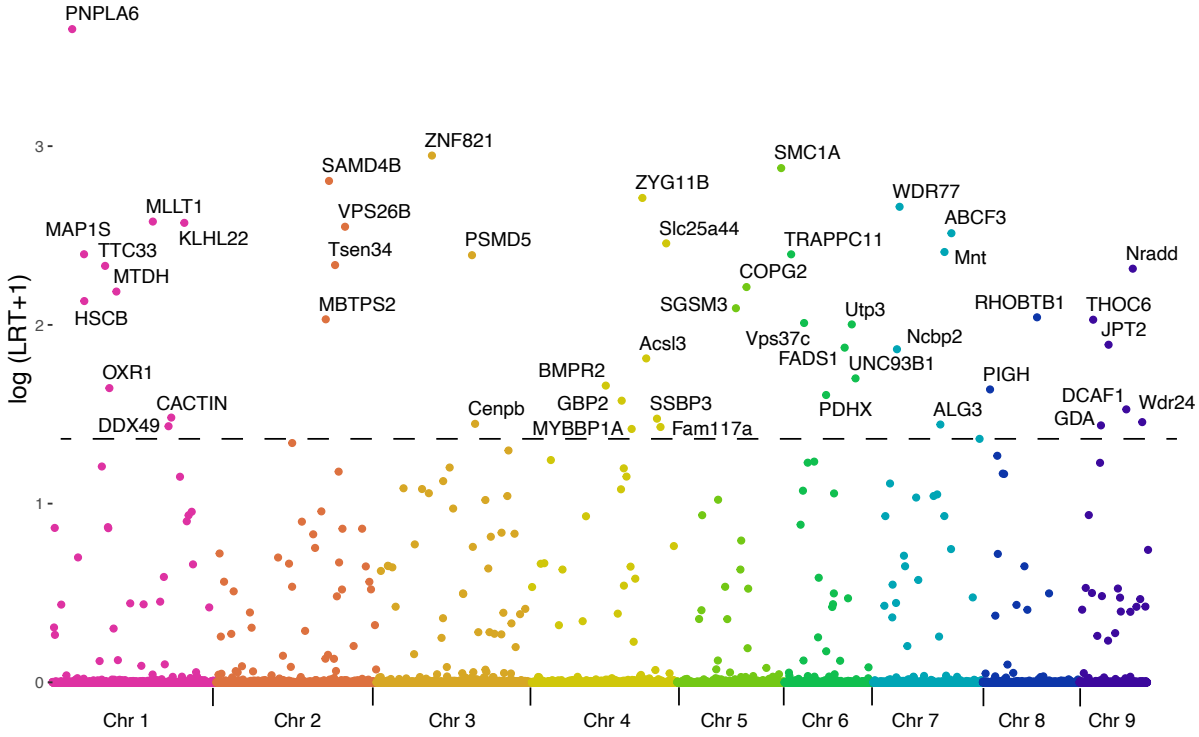


Figure 6: Likelihood ratio of each of the 17355 tested genes of CodeML analysis between two branch-models (BS1 and BS2) of *Trichosurus vulpecula* genes, with western possums as foreground branches and eastern possums as a background branch. This analysis displays evidence of positive selection in western possums. Colours represent the putative chromosomal mapping of the genes.

Because only a small set of genes were found to be under positive selection using CodeML branch model, no GO terms and only two KeGG pathways showed signs of enrichment (Fatty acid metabolism and Viral life cycle) with each only two associated genes.

FUBAR

HyPhy FUBAR analysis was performed on the same set of genes and allowed to identify sites under pervasive selection in 467 genes (Appendix C5 Table 2). Among those genes, 393 displayed only one site under pervasive selection, 54 displayed two sites and 20 genes more than two sites. Gene ontology on this set of genes enrichment analysis revealed 19 enriched terms with a significant FDR (adjusted p-value < 0.05) (Appendix C5 Table 3). The Revigo (Supek et al. 2011) grouping of the enriched terms is visualized using CirGO (Kuznetsova et al. 2019) (Figure 7) and shows that a significant number of genes are associated with small-molecule metabolic process and oxidoreductase activity.

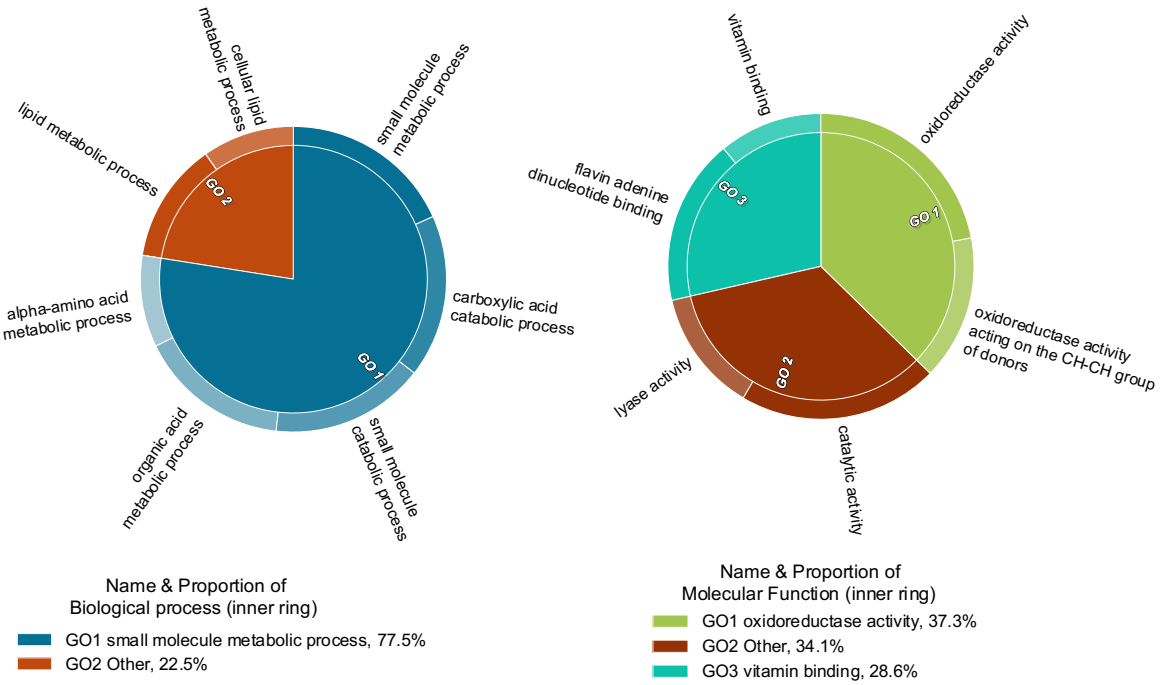


Figure 7: Gene Ontology (GOs) enrichment analysis among the genes showing evidence of pervasive selection using HyPhy FUBAR between western and eastern brushtail possum (*Trichosurus vulpecula*) genomes. Terms are grouped by hierarchical clustering. Parent terms are identified in the legend and their respective proportions, directly proportional to statistical significance. GO terms were first summarized based on a semantic similarity of 0.4 using REVIGO and visualized in CirGO. Circles correspond to one ontology group (BP: Biological process, MF: Molecular function and CC: Cellular component).

Seven significantly enriched KeGG pathways were associated with the genes discovered using FUBAR (Figure 8). The pathways associated with the greatest number of genes were the biosynthesis of cofactors, the lysosome and the Aminoacyl-tRNA biosynthesis. Other pathways were associated with molecules metabolism (Nicotinate, glycan and amino-sugar) and with base excision repair.

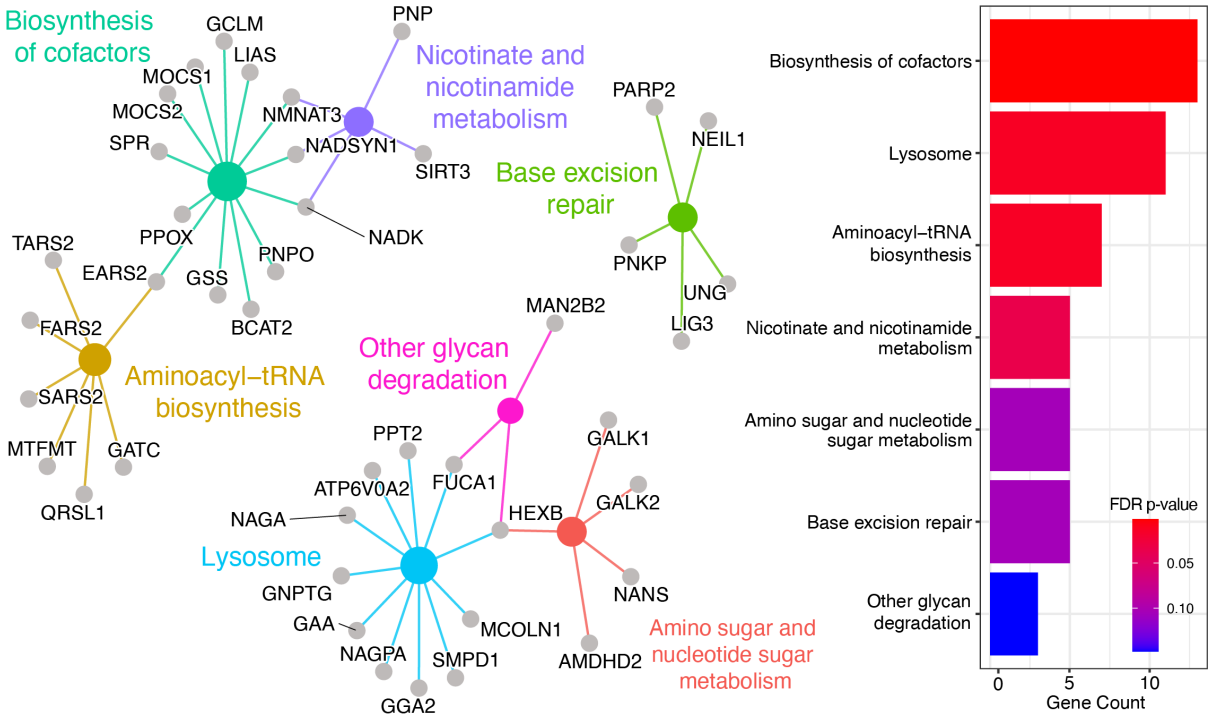


Figure 8: Significantly enriched KeGG pathways (qvalue ≤ 0.2) associated with the genes displaying sites under pervasive selection using HyPhy FUBAR. Large coloured nodes represent pathways and grey nodes indicate positively selected genes within western lineage associated with those pathways. Nodes are sized using enrichment analysis p-value. The number of genes and FDR-adjusted p-values associated with the enriched pathways are reported on the bar plot. Enrichment by pathway terms visualized using the cnetplot function from the "enrichplot" R package.

MEME

Similarly to HyPhy FUBAR, HyPhy MEME was performed on all the applicable genes of the genome looking for sites showing evidence of episodic diversifying positive selection. The result found 94 genes with one significant site under positive selection (Appendix C5 Table 4). Despite gene ontology analysis I was unable to reveal any significantly enriched GO terms because of the small set of genes, but the Kegg pathway analysis indicated five significantly enriched pathways: Nicotinate and nicotinamide metabolism and nucleotide excision repair are shared with HyPhy FUBAR results. The other three pathways are associated with molecule biosynthesis (Folate, cofactors and GPl-anchor, Figure 9).

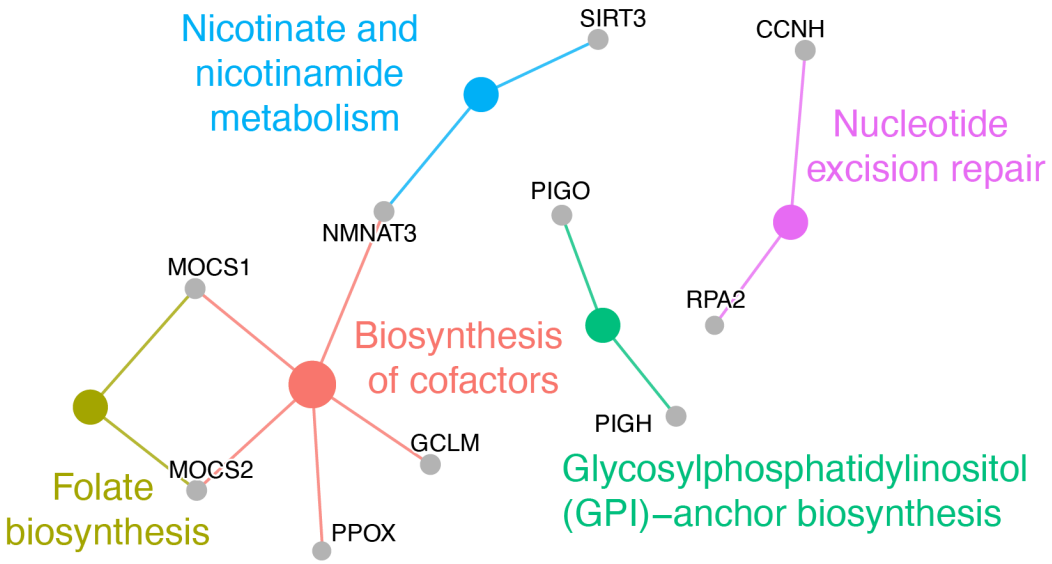


Figure 9: Significantly enriched KeGG pathways (qvalue ≤ 0.2) associated with the genes displaying bases under diversifying positive selection using HyPhy MEME. Big and coloured nodes represent the pathways and grey nodes are the positively selected genes associated with those pathways. Nodes are sized using enrichment analysis p-value. Enrichment by pathway terms visualized using the cnetplot function from the "enrichplot" R package

Genes identified as under positive selection and common to the different analyses were displayed using proportional venn plot (Figure 10).

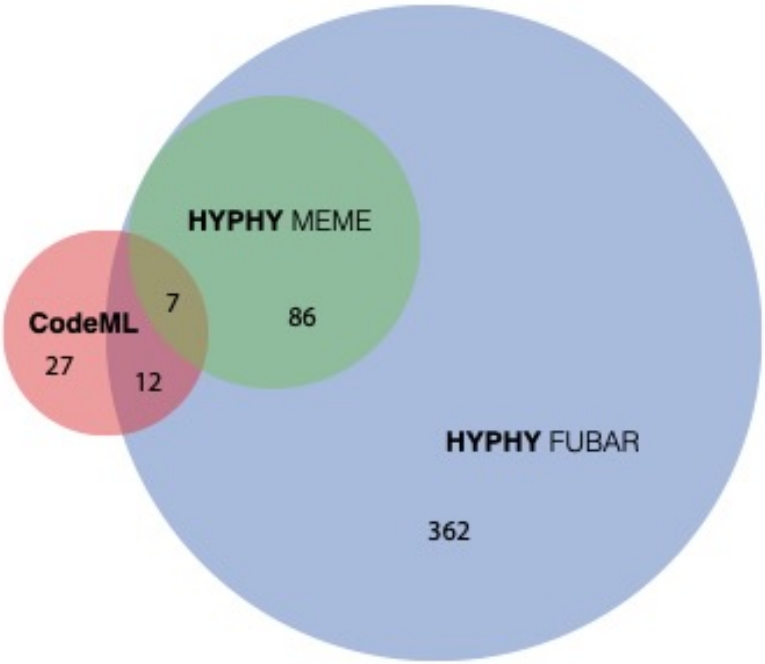


Figure 10: Proportional Venn plot displaying the number of genes showing evidence of positive selection between western and eastern brushtail possums (*Trichosurus vulpecula*) among three different tools (PaML CodeML, HyPhy MEME and HyPhy FUBAR). Ellipse size represents the number of genes and the overlap of the common genes discovered using the analysis. The number of genes associated with each region of the plot is indicated.

Seven candidate genes were found by all three analyses to be under positive selection (Table 1) and the all genes from HyPhy MEME were also discovered by HyPhy FUBAR. It seems that HyPhy FUBAR is less stringent being able to find sites under pervasive selection.

GENES	HYPHY FUBAR N SITES	HYPHY MEME N SITES	PAML CODEML PVALUE	DESCRITPION ENTREZ GENE SUMMARY
TRAPPC11	1	1	0.000866	The protein encoded by this gene is a subunit of the TRAPP (transport protein particle) tethering complex, which functions in intracellular vesicle trafficking. This subunit is involved in early stage endoplasmic reticulum-to-Golgi vesicle transport. Alternative splicing of this gene results in multiple transcript variants. [provided by RefSeq, Jan 2013]
PIGH	1	1	0.024598	This gene encodes an endoplasmic reticulum associated protein that is involved in glycosylphosphatidylinositol (GPI)-anchor biosynthesis. The GPI anchor is a glycolipid found on many blood cells and which serves to anchor proteins to the cell surface. The protein encoded by this gene is a subunit of the GPI N-acetylglucosaminyl (GlcNAc) transferase that transfers GlcNAc to phosphatidylinositol (PI) on the cytoplasmic side of the endoplasmic reticulum. [provided by RefSeq, Jul 2008]
MAP1S	2	1	0.000864	Enables DNA binding activity and cytoskeletal protein binding activity. Involved in microtubule bundle formation; neuron projection morphogenesis; and regulation of chromatin disassembly. Located in several cellular components, including microtubule cytoskeleton; nuclear lumen; and perinuclear region of cytoplasm. [provided by Alliance of Genome Resources, Apr 2022]
THOC6	1	1	0.005706	This gene encodes a subunit of the multi-protein THO complex, which is involved in coordination between transcription and mRNA processing. The THO complex is a component of the TREX (transcription/export) complex, which is involved in transcription and export of mRNAs. A missense mutation in this gene is associated with a neurodevelopmental disorder called Beaulieu-Boycott-Innes syndrome. [provided by RefSeq, Dec 2016]
PSMD5	2	1	0.000891	The 26S proteasome is a multicatalytic proteinase complex with a highly ordered structure composed of 2 complexes, a 20S core and a 19S regulator. The 20S core is composed of 4 rings of 28 non-identical subunits; 2 rings are composed of 7 alpha subunits and 2 rings are composed of 7 beta subunits. The 19S regulator is composed of a base, which contains 6 ATPase subunits and 2 non-ATPase subunits, and a lid, which contains up to 10 non-ATPase subunits. Proteasomes are distributed throughout eukaryotic cells at a high concentration and cleave peptides in an ATP/ubiquitin-dependent process in a non-lysosomal pathway. This gene encodes a non-ATPase subunit of the 19S regulator base that functions as a chaperone protein during 26S proteasome assembly. [provided by RefSeq, Jul 2012]
SAMD4B	1	1	0.000043	Enables RNA binding activity. Predicted to be involved in nuclear-transcribed mRNA poly(A) tail shortening. Predicted to act upstream of or within cerebellar neuron development. Located in cytosol. [provided by Alliance of Genome Resources, Apr 2022]
NRADD	1	1	0.001382	The neurotrophin receptor alike death domain proteins belong to the death domain superfamily and are involved in mediating apoptosis. This gene has been inactivated by mutation and is nonfunctional in humans. [provided by RefSeq, Oct 2008]

Table 1: Table of the seven genes displaying significant positive selection among CodeML, MEME and FUBAR analyses between western and eastern brushtail possums. For each gene the number of sites under selection for both MEME and FUBAR analysis and the CodeML LRT p-value and their respective entrez gene summary is indicated.

F-statistics

Exploring F_{ST} and Gst statistics across all genes of the genomes reveals many genes with high genetic differentiation across the alignment (Figure 11) with 215 genes with a Gst value over 0.54 and 165 genes with a F_{ST} value over 0.60 (Appendix C5 Table 5), those cutoff values were defined modelling empirical distribution and selection quantiles cutoff (0.75 for Gst and 0.95 for Fst). Only 25 genes are common among those list but plotting the genetic differentiation values across the genome reveal expected and similar pattern with a significant correlation between the two statistics (Pearson cor test : p-value < 0.01).

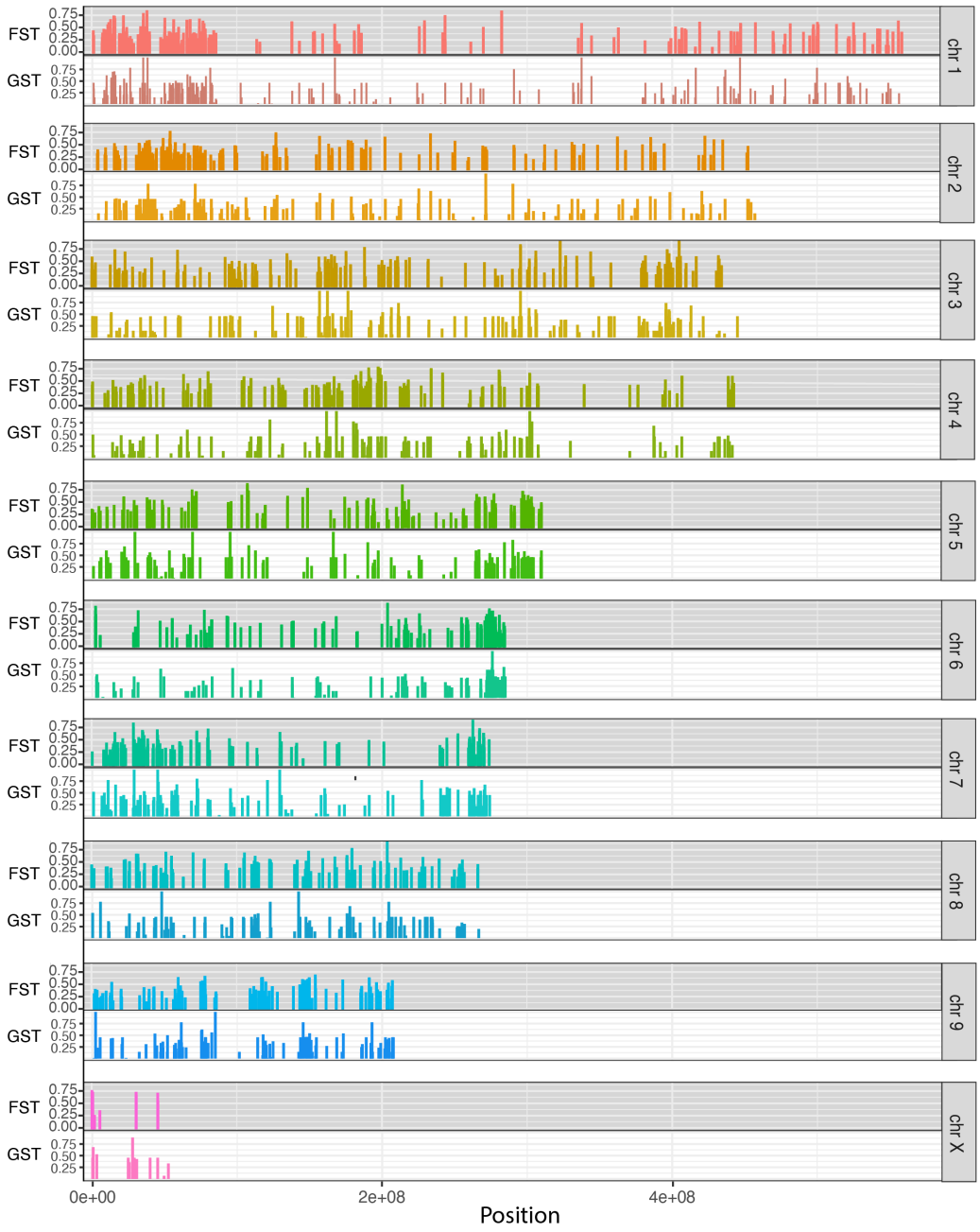


Figure 11: Whole transcriptome FST (top) and GST (bottom) between western Australian and New Zealand brushtail possums

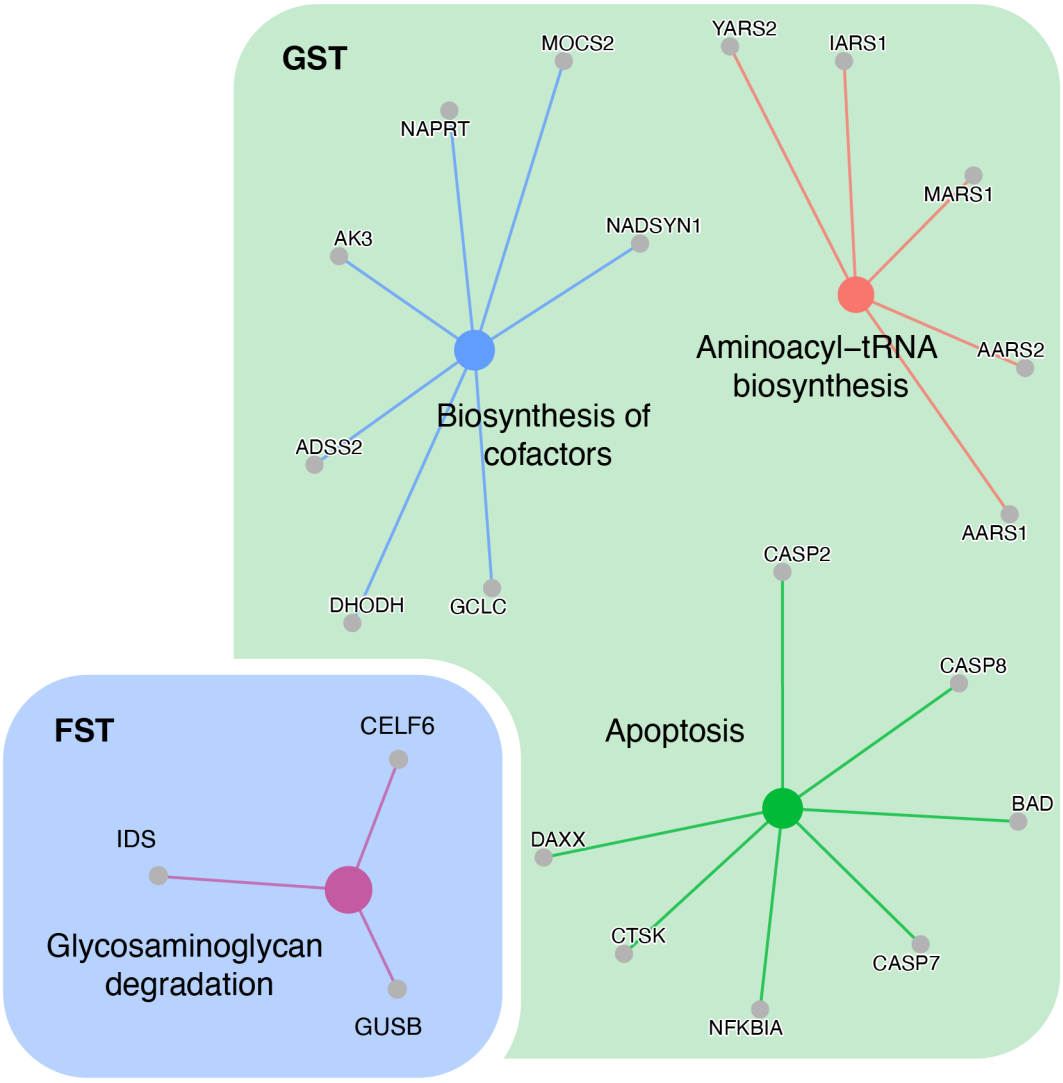


Figure 12: Significantly enriched KeGG pathways (qvalue ≤ 0.2) associated with the genes displaying significantly high Gst and Fst value. Big and coloured nodes represent the pathways and grey nodes are the positively selected genes associated with those pathways. Nodes are sized using enrichment analysis p-value. Enrichment by pathway terms visualized using the cnetplot function from the "enrichplot" R package.

Among the 215 genes with significantly high Gst value three KeGG functional pathways were significantly enriched (Biosynthesis of cofactors, Apoptosis and Aminoacyl-tRNA, Figure 12). For the 165 genes with significantly high F_{ST} only one pathway is enriched with three genes associated with Glycosaminoglycan degradation.

Tassel

The last way I decided to investigate genes under selection is using whole genome trait analysis by association. For this purpose, the rTassel package (Bradbury et al. 2007; Monier et al. 2021) was used. The multidimensional scaling on the distance matrix generated by rTassel from the SNPs reveals how one western sample tends to cluster closer to the other New Zealand samples (Figure 13). This pattern in the data should be considered when analysing the results. Among the 15516 SNPs, rTassel found 1126 of them significantly associated with western brushtail possums using mixed-linear model (Figure 14). Those SNPs were associated to 616 unique genes (Appendix C5 Table 6).

It is apparent that significant p-values tends to reach only some specific values and the reason of this pattern is discussed later in this chapter.

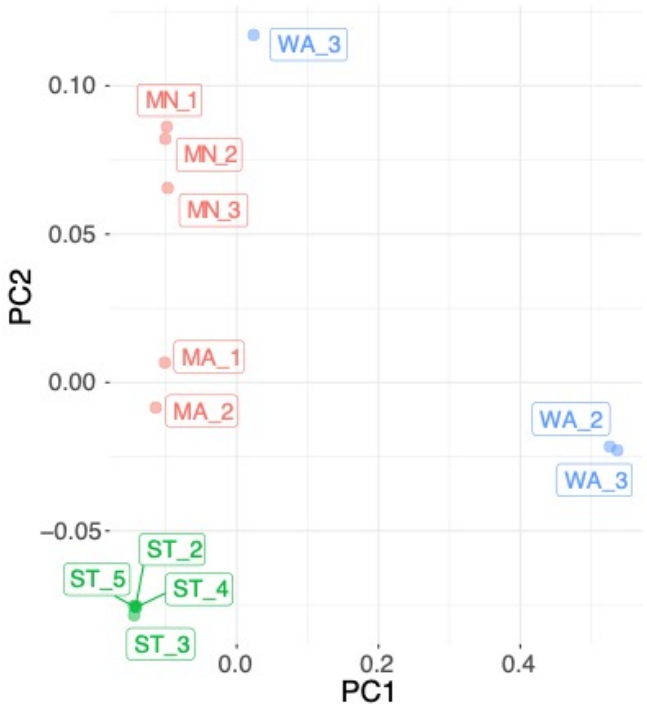


Figure 13: Multi-dimensional scaling plot from Tassel distance matrix of western and eastern brushtail possums SNPs. Samples are coloured from subpopulation (Blue: Western Australian, Red: New Zealand mainland, Green: Stewart Island).

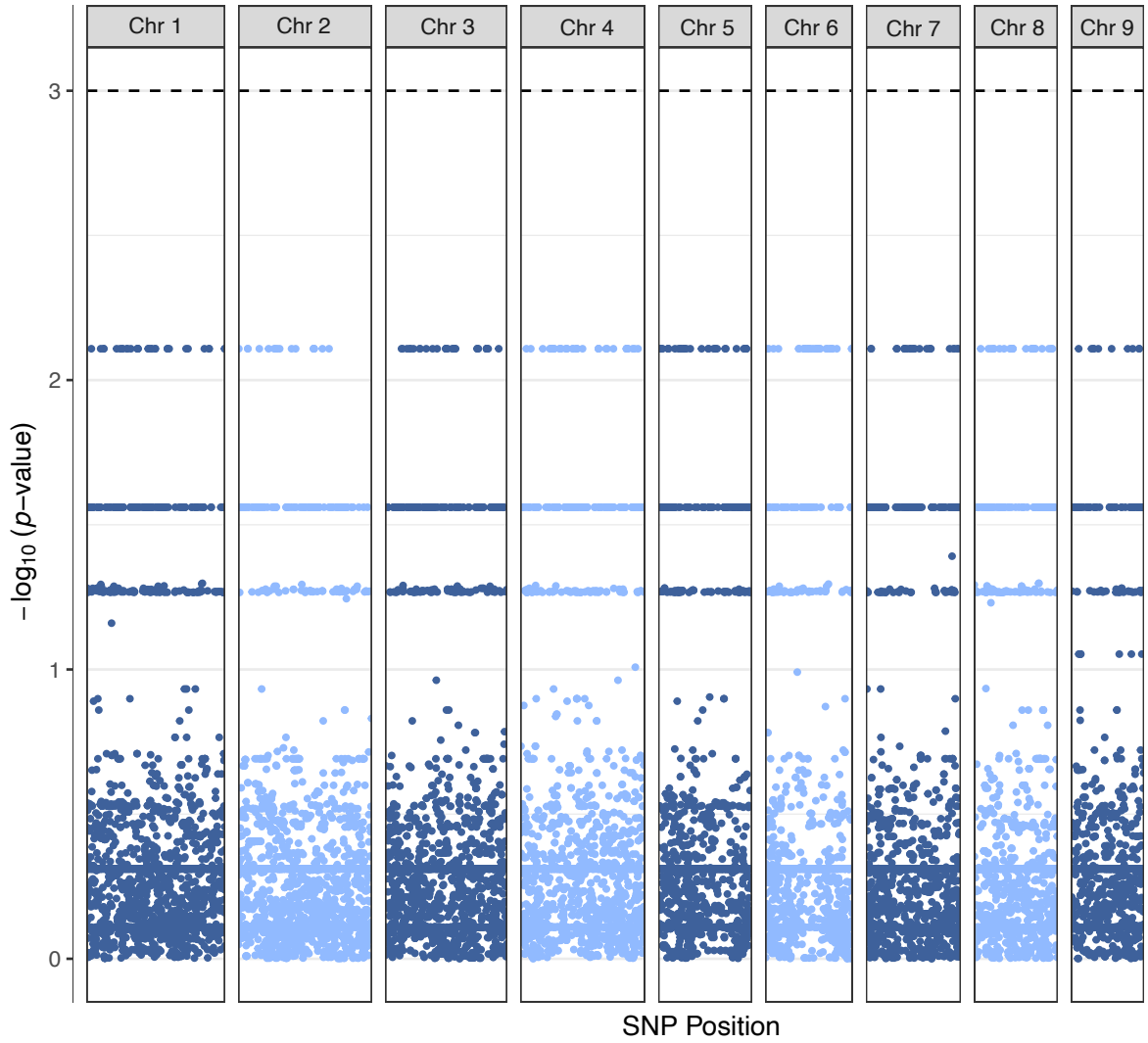


Figure 14: Manhattan type plot of -$\log_{10}$ of the p-value of each SNPs association with the western lineage trait in brushtail possums. Results are separated by chromosome and ordered from their position on the genome. The stratification of the p-values seems to be associated with the small amount of samples in one group.

Gene Ontology term enrichment analysis on the genes associated with significant SNPs (i.e. genes that correspond to SNPs that are significantly associated with western Australian possums) revealed thirteen significantly enriched terms with a significant FDR (adjusted p-value < 0.05) (Appendix C5 Table 7). The Revigo (Supek et al. 2011) grouping of the enriched terms is visualized using CirGO (Kuznetsova et al. 2019) (Figure 15) and reveals that several terms are associated with metabolic process, and catalytic activity, the significant GOs seems to also be associated with nucleolus cellular component.

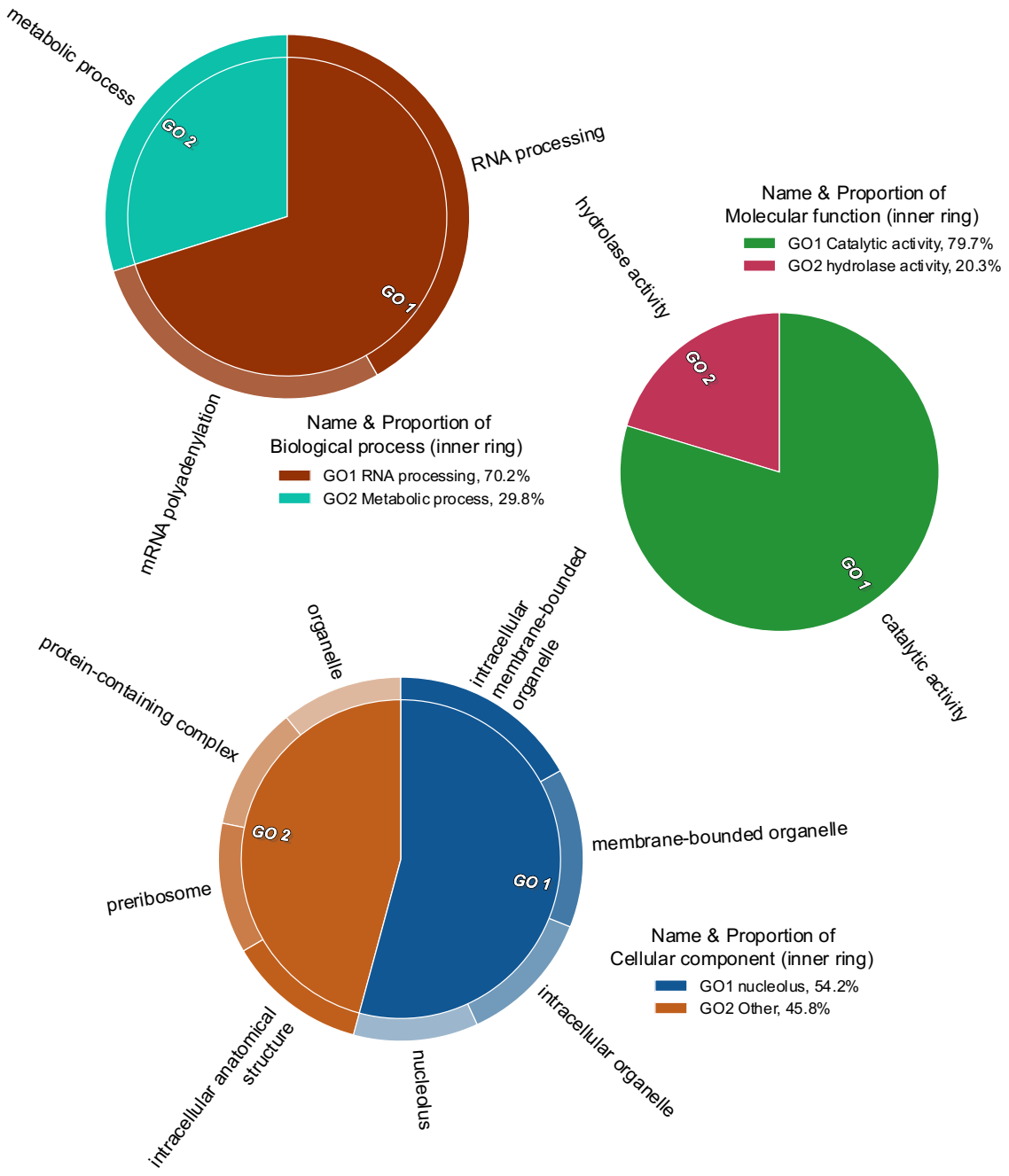


Figure 15: Gene Ontology (GOs) enrichment on the genes displaying SNPs significantly associated with the western brushtail possums' population. Terms are grouped by hierarchical clustering. Parent terms are identified in the legend and their respective proportions, directly proportional to statistical significance. GO terms were first summarized based on a semantic similarity of 0.4 using REVIGO and visualized in CirGO. Circles correspond to one ontology group (BP: Biological process, MF: Molecular function and CC: Cellular component).

Finally, nine significantly enriched KeGG pathways were associated with the genes holding SNPs significantly associated with western lineage. The pathways associated with the greatest number of genes are the biosynthesis of cofactors the lysosome and the nucleotide metabolisme. Other pathways are associated with molecule metabolism (N-glycan biosynthesis, vitamine digestion and Aminoacyl-tRNA biosynthesis) and autophagy.

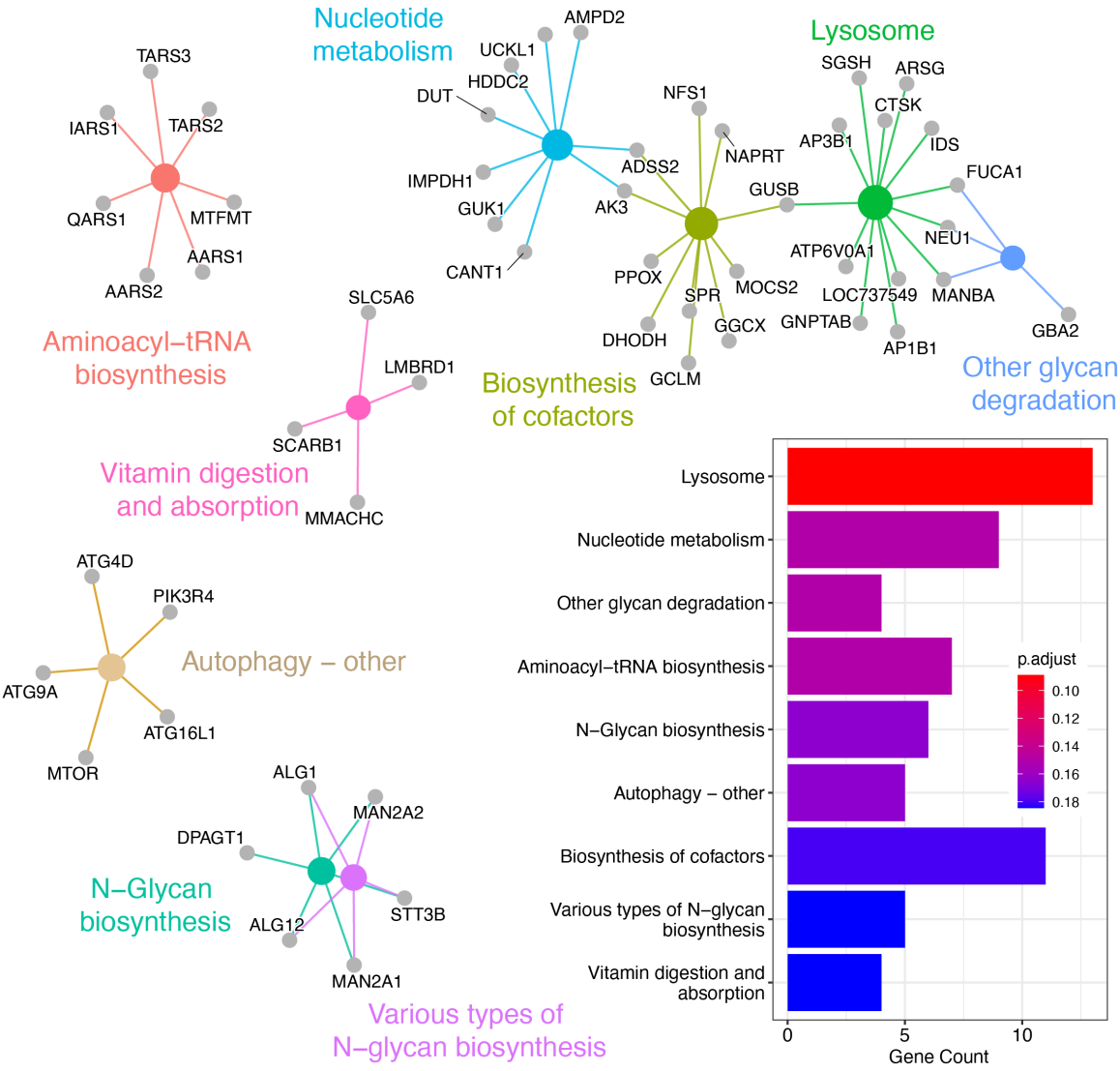


Figure 15: Significantly enriched KeGG pathways (qvalue ≤ 0.2) associated with the genes displaying SNPs significantly associated with western brushtail possums. Big and coloured nodes represent the pathways and grey nodes are the positively selected genes associated with those pathways. Nodes are sized using enrichment analysis p-value. The number of genes and FDR-adjusted p-values associated with the enriched pathways are reported on the bar plot. Enrichment by pathway terms visualized using the cnetplot function from the "enrichplot" R package.

Table 2: Five approaches to identifying genes under positive selection in the expressed genome of brushtail possums yielded this list of enriched KeGG pathways, and number of genes associated with each pathway. The different analysis used phylogenetic branch-site models (CodeML), DNA site-wise models (Fubar, MEME) and genetic differentiation per gene (F_{ST} and GST).

KEGG PATWHAYS	CODEML	HYPHY FUBAR	HYPHY MEME	TASSEL	GST/FST
VIRAL LIFE CYCLE - HIV-1	**2**				
FATTY ACID METABOLISM	**2**				
LYSOSOME		**11**		**13**	
BIOSYNTHESIS OF COFACTORS		**13**	**5**	**11**	**7**
AMINOACYL-TRNA BIOSYNTHESIS		**7**			**5**
NICOTINATE AND NICOTINAMIDE METABOLISM		**5**	**2**		
AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM		**5**			
BASE EXCISION REPAIR		**5**			
OTHER GLYCAN DEGRADATION		**3**		**4**	
GLYCOSYLPHOSPHATIDYLINOSITOL (GPI)-ANCHOR BIOSYNTHESIS			**2**		
FOLATE BIOSYNTHESIS			**2**		
GLYCOSAMINOGLYCAN DEGRADATION					**3**
AMINOACYL-TRNA BIOSYNTHESIS				**7**	
APOPTOSIS					**7**
NUCLEOTIDE METABOLISM				**9**	
N-GLYCAN BIOSYNTHESIS				**6**	
AUTOPHAGY-OTHER				**5**	
VARIOUS TYPE OF N-GLYCAN BIOSYNTHESIS				**5**	
VITAMIN DIGESTION AND ABSORPTION				**4**	
NUCLEOTIDE EXCISION REPAIR			**2**		

Tables of the detail results for each analysis for each gene can be found on the data repository: https://figshare.com/s/4e0f2dc7c9a05e441e03

Discussion

One of the first point to be discussed is how the alignments were generated from transcriptomic data, as it has several impacts on the conclusions we can draw from these analyses. The genome used to map the transcriptome to generate consensus sequences was generated using a brushtail possum from New Zealand, this means that some differences associated with western Australian samples might result from an artefact of the mapping generated by the genome. Using transcription data, only expressed genes are sequenced so the detection of variation is non-random. Using transcriptomic data has the advantage that both differential expression analysis and selection analysis and sequence comparison can be performed using the same sequencing dataset and we are sure that the alleles observed are expressed by the animals.

The data and particularly the genetic and geographic structure of the samples need to be considered to interpret the results. The first point to be acknowledged is the use of New Zealand and western Australian possums to represent the difference between the Eastern lineages (*T. v. fuliginosus* and *T. v. vulpecula*) and the South-Western lineage (*T. v. hypoleucus*). The Eastern lineages were confirmed as the source population of New Zealand possums in Chapter 2 (Carmelet-Rescan et al. 2022, Chapter 2) and correspond to the historic records of the introduction of brushtail possums to New Zealand (Pracy 1974). However, selection pressures on the western and eastern lineages will not be confined to the resistance of plant defence molecules as their environments differ in many other significant ways. In the set of samples used for the analysis in this chapter, there is a disproportion between the number of western Australian and New Zealand samples, but this disproportion is less important for the interpretation of the results than it could be in the expression analysis in the previous chapter. One last point to remember is that one western Australian sample generated ambiguities in the phylogeny (Figure 4). Those ambiguities might be associated with low read depth on some genes and the overall low read count could generate a consensus more similar to the mapped genome (corresponding to the New Zealand lineage). The issues associated with this sample were identified in the previous chapter (Chapter 4), but because I had data from just three western Australian samples this specimen was included here. The inclusion of the low-quality sample might result in failure to detect some genes but it should not increase the rate of false-positive discovery.

There are many different approaches to identifying genes under positive selection and the value of using multiple methods is that the set of genes found by each method can be compared to find candidates in common. In each case, I implemented the selection analysis by comparing DNA sequences from the South-Western possum lineage (*T. v. hypoleucus*) to the Eastern possum lineage (*T. v. fuliginosus* and *T. v. vulpecula*; represented by sampling from New Zealand). Genes under positive selection can be identified using a branch-site model based on phylogenetic analysis. CodeML analysis is based on the concept that the likelihood of a random mutation being non-synonymous is about three times higher than of it being synonymous, but this can vary depending on several factors, including codon composition and transition/transversion ratios. To calculate dS, we must first determine the proportion of synonymous mutations by using it as a baseline. As a rough estimate (assuming that most third positions are synonymous), each codon typically contains one synonymous site and two non-synonymous sites. CodeML successfully detected 47 genes under positive selection, with one enriched pathway (Fatty acid metabolism) associated with two genes (ACSL3, FADS1) that could be associated with western brushtail possums' different diet (Leong et al. 2017; Nugent et al. 2000) and both interact with fatty acyl-coenzyme A.

HyPhy FUBAR is an analysis that finds sites which have experienced diversifying selection while HyPhy MEME finds sites in a DNA alignment where some of the branches have experienced diversifying selection (Kosakovsky Pond, Frost, and Muse 2005; Murrell et al. 2012, 2013). These similarities explain the overlap between the genes identified by both analyses, all genes found using MEME were also found using FUBAR.

It is interesting to look at common genes of the selection analyses with the results from the differential expression analyses from Chapter 4. Among the 2706 differentially expressed genes using DESeq2 (Love, Huber, and Anders 2014) and the 47 genes significantly positively selected using CodeML, two are common and both are upregulated in western brushtail possums. The gene fatty acid desaturase 1 (FADS1) is associated with fatty acid metabolism as described above and the Family With Sequence Similarity 117 Member A (Fam117a) in which the function is not clearly defined. Overall from the 495 positively selected genes (found with HyPhy FUBAR and CodeML) 14 genes were common with the 959 differentially expressed genes identified using DESeq2 (Chapter 4). Despite the fact that this overlap is not significant (Hypergeometric test p-value > 0.05), genes showing both signs of positive selection and evidence of upregulation are relevant to look at trying to have a more broader

view of the genomic differences between western and eastern brushtail possum (*Trichosurus vulpecula*) (Table 3).

Table 3: Genes showing differential expression among brushtail possums (east and Western Australia) which also show significant positive selection in one of the three different analyses (CodeML, HyPhy FUBAR and HyPhy MEME)

GENES	Description
ZYG11B	Zyg-11 Family Member B, Cell Cycle Regulator
PAFAH2	Platelet Activating Factor Acetylhydrolase 2
TOP1MT	DNA Topoisomerase I Mitochondrial
CCL19	C-C Motif Chemokine Ligand 19
CYP2J	Cytochrome P450 Family 2 Subfamily J
SNTB1	Syntrophin Beta 1
LOC101867462	
CPVL	Carboxypeptidase Vitellogenic Like
GJA4	Gap Junction Protein Alpha 4
SEMA3F	Semaphorin 3F
IPP	Intracisternal A Particle-Promoted Polypeptide
DCTN5	Dynactin Subunit 5
ALG2	Alpha-1,3/1,6-Mannosyltransferase
PECAM1	Platelet And Endothelial Cell Adhesion Molecule 1

Using F_{ST} outliers is a common tool to look for selection across the genome (Barendse et al. 2009; Bryk et al. 2008; Holsinger and Weir 2009) but the result from this kind of analysis needs to be combined with other selection analyses. In my case, the results from F_{ST} and GST outliers provided few significant genes but the results of enrichment (Figure 11) were not contradictory with the other approaches. The low amount of samples from western lineage might make this approach a poor fit, the interpretation for these results is accounting for it.

To perform rTassel analysis I associated the western lineage trait with a value of 1 and the other samples with a value of 0. This pattern in the input data and the distance of WA_3 sample to the other western Australian samples is the main cause of the pattern observed in the results, the p-value generated by the analysis seems to take some specific values, when all three western Australian samples share SNP the p-value reach 0.0078, but when only two WA samples are associated with the SNP then the p-value is 0.0275. Despite this pattern, the results are coherent

with other analysis results with 196 found genes and three KeGG pathways shared with other analyses.

GO term and KeGG pathway enrichment analysis performed on genes isolated with each analysis tend to result in shared or similar terms and pathways. The enriched KEGG functions cover a wide range of biological processes (Table 2), such as lysosome, biosynthesis of co-factors, aminoacyl-tRNA biosynthesis, nicotinate and nicotinamide metabolism, amino sugar and nucleotide sugar metabolism, base excision repair, other glycan degradation, GPI-anchor biosynthesis, folate biosynthesis, glycosaminoglycan degradation, apoptosis, nucleotide metabolism, N-glycan biosynthesis, autophagy, various types of N-glycan biosynthesis, and vitamin digestion and absorption. Nicotinate and nicotinamide metabolism, amino sugar and nucleotide sugar metabolism, other glycan degradation, and glycosaminoglycan degradation are metabolic pathways that are involved in the breakdown and synthesis of different classes of molecules. Glycosylphosphatidylinositol (GPI)-anchor biosynthesis is a pathway that is involved in the post-translational modification of proteins. Base excision repair is a DNA repair pathway that is involved in the repair of DNA damage caused by oxidative stress. Apoptosis is a cellular process of programmed cell death that is involved in the elimination of damaged or unnecessary cells. It is unclear if there is any direct association between these functions and sodium fluoroacetate resistance in brushtail possums, but some of the functions listed may indirectly contribute to the resistance. For example, lysosome function may play a role in breaking down and eliminating toxic substances, including sodium fluoroacetate. Additionally, folate biosynthesis and vitamin digestion and absorption may be involved in providing necessary nutrients for the possums to combat the toxic effects of sodium fluoroacetate.

The use of statistical methods to look for positive selection in the genome especially when it is detached from any clear biological mechanism has been criticized and when the associated phylogeny is not properly supported (Hughes 2007). In this chapter I tried to associated different tools and look at the associate biological mechanisms to make sure the conclusions are supported by several results.

Overall, this chapter contributes to a better understanding of the adaptation and evolution of brushtail possum South-Western lineage (*T. v. hypoleucus*) compared to the Eastern lineage (*T. v. fuliginosus* and *T. v. vulpecula*).

References

Andrews, Simon, and others. 2010. "FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010." *Https://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/* 1(1): http://www.bioinformatics.babraham.ac.uk/projects/. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/%0Ahttp://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/ (March 17, 2020).

Barendse, William et al. 2009. "Genome Wide Signatures of Positive Selection: The Comparison of Independent Samples and the Identification of Regions Associated to Traits." *BMC Genomics* 10(1): 1–15. https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-10-178 (April 28, 2023).

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57(1): 289–300.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30(15): 2114–20. https://academic.oup.com/bioinformatics/article-abstract/30/15/2114/2390096 (March 17, 2020).

Bradbury, Peter J. et al. 2007. "TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples." *Bioinformatics* 23(19): 2633–35. https://academic.oup.com/bioinformatics/article/23/19/2633/185151 (March 10, 2023).

Bryk, Jarosław et al. 2008. "Positive Selection in East Asians for an EDAR Allele That Enhances NF-KB Activation." *PLoS ONE* 3(5): e2209. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0002209 (April 28, 2023).

Carmelet-Rescan, David, Mary Morgan-Richards, Nimeshika Pattabiraman, and Steven A. Trewick. 2022. "Time-Calibrated Phylogeny and Ecological Niche Models Indicate Pliocene Aridification Drove Intraspecific Diversification of Brushtail Possums in Australia." *Ecology and Evolution* 12(12): e9633. https://onlinelibrary.wiley.com/doi/full/10.1002/ece3.9633 (February 15, 2023).

Danecek, Petr et al. 2021. “Twelve Years of SAMtools and BCFtools.” *GigaScience* 10(2): 1–4. https://academic.oup.com/gigascience/article/10/2/giab008/6137722 (March 10, 2023).

Dearing, M. Denise, and Steven Cork. 1999. “Role of Detoxification of Plant Secondary Compounds on Diet Breadth in a Mammalian Herbivore, Trichosurus, Vulpecula.” *Journal of Chemical Ecology* 25(6): 1205–19. https://link.springer.com/article/10.1023/A:1020958221803 (May 8, 2023).

Delignette-Muller, Marie Laure, and Christophe Dutang. 2015. “Fitdistrplus: An R Package for Fitting Distributions.” *Journal of Statistical Software* 64(4): 1–34. https://www.jstatsoft.org/index.php/jss/article/view/v064i04 (March 10, 2023).

Efford, Murray, Bruce Warburton, and Nick Spencer. 2000. “Home-Range Changes by Brushtail Possums in Response to Control.” *Wildlife Research* 27(2): 117–27. http://www.publish.csiro.au/WR/WR99005 (March 15, 2020).

Ford, Michael J. 2001. “Molecular Evolution of Transferrin: Evidence for Positive Selection in Salmonids.” *Molecular Biology and Evolution* 18(4): 639–47. https://academic.oup.com/mbe/article/18/4/639/980136 (April 25, 2023).

Garrison, Erik, and Gabor Marth. 2012. “Haplotype-Based Variant Detection from Short-Read Sequencing.” https://arxiv.org/abs/1207.3907v2 (March 10, 2023).

Holsinger, Kent E., and Bruce S. Weir. 2009. “Genetics in Geographically Structured Populations: Defining, Estimating and Interpreting FST.” *Nature Reviews Genetics* 10(9): 639–50. /pmc/articles/PMC4687486/ (May 8, 2023).

Hughes, A. L. 2007. “Looking for Darwin in All the Wrong Places: The Misguided Quest for Positive Selection at the Nucleotide Sequence Level.” *Heredity* 99(4): 364–73. https://www.nature.com/articles/6801031 (April 28, 2023).

Institute, Broad. 2019. “‘Picard Toolkit’, Broad Institute, GitHub Repository.” *Picard Toolkit*.

Jeffares, Daniel C., Bartłomiej Tomiczek, Victor Sojo, and Mario dos Reis. 2015. “A Beginners Guide to Estimating the Non-Synonymous to Synonymous Rate Ratio of All Protein-Coding Genes in a Genome.” *Methods in Molecular Biology* 1201: 65–90.

Kapralov, Maxim V., and Dmitry A. Filatov. 2007. “Widespread Positive Selection in the Photosynthetic Rubisco Enzyme.” *BMC Evolutionary Biology* 7(1): 1–10. https://link.springer.com/articles/10.1186/1471-2148-7-73 (April 25, 2023).

Kim, Daehwan et al. 2019. “Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype.” *Nature Biotechnology 2019 37:8* 37(8): 907–15. https://www.nature.com/articles/s41587-019-0201-4 (July 4, 2022).

Kimura, Motoo. 1977. “Preponderance of Synonymous Changes as Evidence for the Neutral Theory of Molecular Evolution [33].” *Nature* 267(5608): 275–76.

Kosakovsky Pond, Sergei L., Simon D.W. Frost, and Spencer V. Muse. 2005. “HyPhy: Hypothesis Testing Using Phylogenies.” *Bioinformatics* 21(5): 676–79. https://academic.oup.com/bioinformatics/article/21/5/676/220389 (April 5, 2023).

Kryazhimskiy, Sergey, and Joshua B. Plotkin. 2008. “The Population Genetics of DN/DS.” *PLoS Genetics* 4(12).

Kuznetsova, Irina et al. 2019. “CirGO: An Alternative Circular Way of Visualising Gene Ontology Terms.” *BMC Bioinformatics* 20(1): 1–7. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2671-2 (February 14, 2023).

Larsson, Johan, and Peter Gustafsson. 2018. “A Case Study in Fitting Area-Proportional Euler Diagrams with Ellipses Using Eulerr.” In *Proceedings of International Workshop on Set Visualization and Reasoning*, , 84–91. https://cran.r-project.org/package=eulerr.

Leong, Lex Ee Xiang et al. 2017. “Fluoroacetate in Plants - a Review of Its Distribution, Toxicity to Livestock and Microbial Detoxification.” *Journal of Animal Science and Biotechnology* 8(1): 1–11. https://jasbsci.biomedcentral.com/articles/10.1186/s40104-017-0180-6 (April 22, 2022).

Li, Heng et al. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25(16): 2078–79. https://academic.oup.com/bioinformatics/article/25/16/2078/204688 (February 10, 2023).

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15(12): 1–21. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8 (July 4, 2022).

Monier, Brandon, Terry M Casstevens, Peter J Bradbury, and Edward S Buckler. 2021. “RTASSEL: An R Interface to TASSEL for Association Mapping of Complex Traits.” *bioRxiv*: 2020.07.21.209114. http://biorxiv.org/content/early/2021/03/31/2020.07.21.209114.abstract.

Murrell, Ben et al. 2012. “Detecting Individual Sites Subject to Episodic Diversifying Selection.” *PLoS Genetics* 8(7): e1002764. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002764 (April 5, 2023).

———. 2013. “FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring

Selection." *Molecular Biology and Evolution* 30(5): 1196–1205. https://academic.oup.com/mbe/article/30/5/1196/998247 (April 5, 2023).

Nugent, G., P. Sweetapple, J. Coleman, and P. Suisted. 2000. "Possum Feeding Patterns: Dietary Tactics of a Reluctant Folivore." In *The Brushtail Possum: Biology, Impact and Management of an Introduced Marsupial*, , 10–19.

Posada, David. 2008. "JModelTest: Phylogenetic Model Averaging." *Molecular Biology and Evolution* 25(7): 1253–56. https://academic.oup.com/mbe/article/25/7/1253/1045159 (April 24, 2023).

Pracy, L. 1974. "Opposums." *New Zealand Nature Heritage* 3(32): 873–82.

Souza, Stephanie S.R. et al. 2022. "Population Analysis of Heavy Metal and Biocide Resistance Genes in Salmonella Enterica from Human Clinical Cases in New Hampshire, United States." *Frontiers in Microbiology* 13: 4163.

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30(9): 1312–13. https://academic.oup.com/bioinformatics/article-abstract/30/9/1312/238053 (March 16, 2020).

Supek, Fran, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. 2011. "Revigo Summarizes and Visualizes Long Lists of Gene Ontology Terms." *PLoS ONE* 6(7): e21800. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021800 (February 14, 2023).

Takahata, N, and M Nei. 1984. "F and g Statistics in the Finite Island Model." *Genetics* 107(3): 501–4. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1202339/ (April 18, 2023).

Twigg, L. E. et al. 1996. "Fluoroacetate Content of Some Species of the Toxic Australian Plant Genus, Gastrolobium, and Its Environmental Persistence." *Natural Toxins* 4(3): 122–27. http://doi.wiley.com/10.1002/19960403NT4 (March 15, 2020).

Twigg, L. E., and D. R. King. 1991. "The Impact of Fluoroacetate-Bearing Vegetation on Native Australian Fauna: A Review." *Oikos* 61(3): 412. https://www.jstor.org/stable/3545249?casa_token=wxSjPb9PE-QAAAAA:AqA216Efy0uOSzZJwBTgcg7Eo64RP0a9Hmm8Ua_rf4Faa-Qp-yot16PIkohRXPufbktoxRP5zU0RpZ0KIa9YOsEXHe5Co014rR-2rRyJe0hgP43FRiJi (March 15, 2020).

Vermaak, Danielle, Steven Henikoff, and Harmit S. Malik. 2005. "Positive Selection Drives the Evolution of Rhino, a Member of the Heterochromatin Protein 1 Family in Drosophila." *PLoS Genetics* 1(1): 0096–0108.

https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.0010009 (April 25, 2023).

Wallenius, Kenneth T. 1963. "Biased Sampling: The Noncentral Hypergeometric Probability Distribution." https://apps.dtic.mil/sti/citations/AD0426243 (February 14, 2023).

Weir, B. S., and C. C. Cockerham. 1984. "Estimating F-Statistics for the Analysis of Population Structure." *Evolution* 38(6): 1358–70. https://www.jstor.org/stable/2408641 (April 18, 2023).

Winter, David J. 2012. "MMOD: An R Library for the Calculation of Population Differentiation Statistics." *Molecular Ecology Resources* 12(6): 1158–60. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1755-0998.2012.03174.x (April 18, 2023).

Yang, Ziheng. 1998. "Likelihood Ratio Tests for Detecting Positive Selection and Application to Primate Lysozyme Evolution." *Molecular Biology and Evolution* 15(5): 568–73. https://academic.oup.com/mbe/article/15/5/568/987857 (April 25, 2023).

———. 2007. "PAML 4: Phylogenetic Analysis by Maximum Likelihood." *Molecular Biology and Evolution* 24(8): 1586–91. https://academic.oup.com/mbe/article/24/8/1586/1103731 (March 10, 2023).

Young, Matthew, Maintainer Nadia Davidson, Anthony Hawkins, and G O biocViews Sequencing. 2013. "Package 'Goseq.'"

Yu, Guangchuang. 2021. "Enrichplot: Visualization of Functional Enrichment Result." https://yulab-smu.top/biomedical-knowledge-mining-book/.

Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. "ClusterProfiler: An R Package for Comparing Biological Themes among Gene Clusters." *Omics: a journal of integrative biology* 16(5): 284–87.

Chapter 6: General conclusion

Brushtail possums (*Trichosurus vulpecula*) were introduced to New Zealand, where they have significantly impacted native ecosystems. Introductions in the 19th century consisted of translocations of individuals of the *T. v. fuliginosus* and *T. v. vulpecula* subspecies from Tasmanian and south-eastern lineages, respectively (Pracy, 1974, see Chapters 1 and 2). Due to their dependence on vegetative quality for foraging opportunities, possum density in New Zealand is high as the abundant and diverse evergreen vegetation provides a rich habitat for the animal (Efford, Warburton, and Spencer 2000). Unfortunately, these possums pose a serious threat to the country's biodiversity since they feed on native endangered plant species, and sometimes display opportunistic feeding habits with native bird eggs and insects (Nugent et al. 2000). Moreover, possums act as vectors for bovine tuberculosis (TB), which can exert a significant cost on the agricultural sector. As part of pest control measures, sodium fluoroacetate poison (widely known as compound 1080) is used extensively to reduce local populations of these possums (McIlroy 1983; Ross, Bicknell, and Hickling 1999).

The key driver of the research presented in this thesis was to explore molecular genomic and expression variation to better understand functional differences between populations subject to different natural selection. In particular the mechanism of toxin resistance that is apparent in the phenotype of *T. v. hypoleucus* in western Australia. The primary hypothesis was that the toxin resistance observed in western Australian brushtail possums is linked to lineage-specific differences in gene expression and genes experiencing positive selection. Additionally, other phenotypic characteristics associated with various brushtail possum lineages have been examined. To test these hypotheses, the first step involved determining the evolutionary history of brushtail possums throughout their habitat range. Subsequently, transcriptomics were generated to identify differential expression of genes and associated pathways specific in possum lineages. Finally, the distinct coding regions of genomes have been assembled and compared to test for a signal of positive selection.

The first part of the thesis has, with a sufficient sampling of Australian and New Zealand brushtail possums, corroborated the previously stated subspecies delimitation (Kerie, McKay, and Sharman 1991) and the use of niche modelling and molecular clock tools has provided an explanation of this delimitation linked to Pliocene climate change (Chapter2; Carmelet-Rescan et al. 2022). This provided crucial information on the subspecies / lineages from which the New Zealand introduction originated and information about the genetic distance of the New Zealand subspecies from the others, especially *T. v. hypoleucus* being the subspecies known to

be most resistant to sodium fluoroacetate (Mead, Oliver, and King 1979; Twigg et al. 2003). Extensive sampling and extensive single-copy nuclear markers would be a great follow-up to this study to provide a more detailed geographic and time resolution of the splits and degree of gene flow among subspecies and populations.

The third chapter focuses on gene expression analysis and utilised data from the liver of juvenile and adult brushtail possums. The study explored the differences in gene expression patterns related to tissue development, metabolism, and potential dietary changes. This stage allowed development and testing of analytical protocols, identification of the gene set involved and the pathway behind brushtail possums' development.

The rest of the thesis focused on the link between *T. v. hypoleucus* phenotype and its expressed genotype compared to *T. v. fuliginosus* and *T. v. vulpecula* subspecies from Tasmanian and south-eastern lineages. Gene expression associated with the western Australian phenotype and its possible involvement in toxin resistance was then investigated in detail. The study identified differentially expressed genes related to various pathways, highlighting the adaptive responses of brushtail possums to distinct ecological conditions. The research suggests that gene expression differences contribute to physiological and morphological adaptations to environmental factors, including food availability and temperature fluctuations, and allows to formulate hypotheses on the brushtail possums' pathways of resistance to fluoroacetate with genes involved in carbon metabolism, glyoxylate and dicarboxylate metabolism, glycolysis, lactate dehydrogenase, and acetyl-CoA synthesis.

The final research chapter focused on the selection analysis of transcriptomic data. The study utilized different methods to identify genes under positive selection and highlights the genomic basis for metabolic differences between the possum lineages. It resulted in a list of positively selected genes and pathways.

Several genes associated with fatty acid metabolism appear to be over-expressed and show evidence of positive selection (Acyl-CoA dehydrogenase, Fatty acid desaturase) and other genes related to the carbon cycle confirming the hypothesis formulated in Chapter 4. The main hypothesis, as formulated by Goncharov, Jenkins, and Radilov (2006), is the over-expression of pathways competing for CoA, an element crucial of the toxicity of sodium fluoroacetate.

The over-expression of fatty acid metabolism can also be an alternative source of energy as the TCA cycle would be stopped by the sodium fluoroacetate activity.

Another hypothesis would be that the modified Acyl-CoA dehydrogenase (member 10) would be able to metabolise Fluoroacetyl-CoA and incorporate it into the TCA cycle or other pathways preventing Fluorocitrate to be produced and associated with Aconitase. The aconitase gene itself didn't show any evidence of under-expression or any sign of selection discarding the most obvious hypothesis of resistance but corresponding to other findings on the subject (Deakin et al. 2013).
Overall, these studies contribute to our understanding of the population structure, evolutionary history, and gene expression patterns of brushtail possums, shedding light on their adaptation to various environments and providing potential applications for further research and conservation studies associated with *Trichosurus vulpecula* in New Zealand.

An alternative hypothesis that could be explored in future studies would be that *Trichosurus vulpecula* tolerance to 1080 could be associated with gut microbes (Kohl et al. 2014).

More in-depth analysis of the highlighted genes focusing on other phenotypic differences between *T. v. hypoleucus, T. v. fuliginosus* and *T. v. vulpecula* would help to isolate which might be more directly associated with sodium fluoroacetate resistance (How and Hillcox 2000; Wayne et al. 2005; Wilson 2001).

Detailed examination of candidate genes among New Zealand possum populations that have experienced 1080 exposure to differing degrees would be a valuable test, especially between toxin naïve Stewart Island and mainland New Zealand populations. Use of a primer multiplex approach featuring candidate genes, constraint genes and neutral markers would allow monitoring of deviations in these genes in relation to population structure and management, to highlight signs of selection. Assembling the complete western Australian *T. v. hypoleucus* genome would be important to investigate large scale genomic variation among evolutionary lineages and selection among non-coding regions such as promoters (Bond et al. 2023).

Following up the result of differential expression analysis with RT-qPCR experiments is the usual next step, developing primers of candidate genes and monitoring the amount of

expression of the chosen genes within each sample of a chosen population (Heid et al. 1996; Orlando, Pinzani, and Pazzagli 1998).

In conclusion, this comprehensive research on brushtail possums (*Trichosurus vulpecula*) over their range has yielded valuable insights into their population structure, evolutionary history, and gene expression patterns. The study confirmed subspecies delimitation and provided a historical context of current population structures. It has also identified specific genes and pathways associated with the possums' adaptation to their diverse environments, including their resistance to naturally occurring sodium fluoroacetate poison. Notably, genes associated with fatty acid metabolism and CoA utilization and more globally genes associated with the TCA cycle appear to play a crucial role in sodium fluoroacetate resistance. Further investigation is needed to explore the potential role of gut microbes and other phenotypic differences. These findings open avenues for further investigations and potential conservation strategies.

Overall, this study contributes to our understanding of brushtail possums' population structure, evolutionary history, and gene expression patterns. It provides a foundation for future research and conservation efforts aimed at mitigating the impact of these possums on New Zealand's ecosystems and agriculture. Further studies involving candidate genes, controlled experiment, and promoting region variation analysis will be essential for a comprehensive understanding of this complex issue.

Reference

Bond, Donna M. et al. 2023. “The Admixed Brushtail Possum Genome Reveals Invasion History in New Zealand and Novel Imprinted Genes.” *Nature Communications 2023 14:1* 14(1): 1–17. https://www.nature.com/articles/s41467-023-41784-8 (October 18, 2023).

Carmelet-Rescan, David, Mary Morgan-Richards, Nimeshika Pattabiraman, and Steven A. Trewick. 2022. “Time-Calibrated Phylogeny and Ecological Niche Models Indicate Pliocene Aridification Drove Intraspecific Diversification of Brushtail Possums in Australia.” *Ecology and Evolution* 12(12): e9633. https://onlinelibrary.wiley.com/doi/full/10.1002/ece3.9633 (February 15, 2023).

Deakin, Janine E. et al. 2013. “Towards an Understanding of the Genetic Basis behind 1080 (Sodium Fluoroacetate) Tolerance and an Investigation of the Candidate Gene ACO2.”

Australian Journal of Zoology 61(1): 69–77. http://www.publish.csiro.au/?paper=ZO12108 (March 15, 2020).

Efford, Murray, Bruce Warburton, and Nick Spencer. 2000. "Home-Range Changes by Brushtail Possums in Response to Control." *Wildlife Research* 27(2): 117–27. http://www.publish.csiro.au/WR/WR99005 (March 15, 2020).

Goncharov, Nikolay V., Richard O. Jenkins, and Andrey S. Radilov. 2006. "Toxicology of Fluoroacetate: A Review, with Possible Directions for Therapy Research." *Journal of Applied Toxicology* 26(2): 148–61. https://onlinelibrary.wiley.com/doi/full/10.1002/jat.1118 (February 16, 2023).

Heid, Christian A., Junko Stevens, Kenneth J. Livak, and P. Mickey Williams. 1996. "Real Time Quantitative PCR." *Genome Research* 6(10): 986–94. https://genome.cshlp.org/content/6/10/986 (May 25, 2023).

How, R. A., and S. J. Hillcox. 2000. "Brushtail Possum, Trichosurus Vulpecula, Populations in South-Western Australia: Demography, Diet and Conservation Status." *Wildlife Research* 27(1): 81–89. https://www.publish.csiro.au/wr/wr98064 (November 8, 2021).

Kerie, J. A., G. M. McKay, and G. B. Sharman. 1991. "A Systematic Analysis of the Brushtail Possum, Trichosurus-Vulpecula (Kerr, 1792) (Marsupialia, Phalangeridae)." *Australian Journal of Zoology* 39(3): 313–31. https://www.publish.csiro.au/zo/zo9910313 (May 31, 2022).

Kohl, Kevin D. et al. 2014. "Gut Microbes of Mammalian Herbivores Facilitate Intake of Plant Toxins." *Ecology Letters* 17(10): 1238–46. https://onlinelibrary.wiley.com/doi/full/10.1111/ele.12329 (May 24, 2023).

McIlroy, J. C. 1983. "The Sensitivity of the Brushtail Possum (Trichosurus Vulpecula) to 1080 Poison (Sodium Monfluoroacetate)." *New Zealand Journal of Ecology* 6: 125–31. https://www.jstor.org/stable/24052734 (March 15, 2020).

Mead, R. J., A. J. Oliver, and D. R. King. 1979. "Metabolism and Defluorination of Fluoroacetate in the Brush-Tailed Possum (Trichosurus Vulpecula)." *Australian Journal of Biological Sciences* 32(1): 15–26.

Nugent, G., P. Sweetapple, J. Coleman, and P. Suisted. 2000. "Possum Feeding Patterns: Dietary Tactics of a Reluctant Folivore." In *The Brushtail Possum: Biology, Impact and Management of an Introduced Marsupial*, , 10–19.

Orlando, Claudio, Pamela Pinzani, and Mario Pazzagli. 1998. "Developments in Quantitative PCR." *Clinical Chemistry and Laboratory Medicine* 36(5): 255–69. https://www.degruyter.com/document/doi/10.1515/CCLM.1998.045/html (May 25,

2023).

Pracy, L. 1974. “Opposums.” *New Zealand Nature Heritage* 3(32): 873–82.

Ross, James G., Kathryn Bicknell, and G.J. Hickling. 1999. “Cost-effective control pf 1080 bait-shay possums (Trichosurus vulpecula)”

Twigg, L. E. et al. 2003. “Sensitivity of Some Australian Animals to Sodium Fluoroacetate (1080): Additional Species and Populations, and Some Ecological Considerations.” *Australian Journal of Zoology* 51(5): 515–31. https://www.publish.csiro.au/zo/zo03040 (March 23, 2022).

Wayne, A. F. et al. 2005. “The Life History of Trichosurus Vulpecula Hypoleucus (Phalangeridae) in the Jarrah Forest of South-Western Australia.” *Australian Journal of Zoology* 53(4): 265–78. https://www.publish.csiro.au/zo/ZO05008 (May 25, 2023).

Wilson, T. G. 2001. “Resistance of Drosophila to Toxins.” *Annual Review of Entomology* 46: 545–71. https://www.annualreviews.org/doi/abs/10.1146/annurev.ento.46.1.545 (May 3, 2023).

Appendices

Chapter 2

List of datasets used for presence: https://doi.ala.org.au/doi/d542a045-55c0-4cd3-868f-83ebba172bbd

Appendix C2 Table 1 MaxEnt **(a)**, Random forest **(b)**, ANN **(c)** and GBM **(d)** SDMTune results

(a)

fc	reg	iter	train_AUC	test_AUC	diff_AUC
lqp	2.6	130	0.95664297	0.9411	0.01554297
lqp	2.6	130	0.95664297	0.9411	0.01554297
lqp	2.6	130	0.95664297	0.9411	0.01554297
lqp	2.6	130	0.95664297	0.9411	0.01554297
lqp	2.6	130	0.95664297	0.9411	0.01554297
lq	2.6	130	0.95508359	0.93955	0.01553359
lqp	1.1	130	0.95085703	0.934575	0.01628203
lqp	0.9	130	0.94298984	0.925425	0.01756484
lqph	2.6	130	0.91173125	0.89985	0.01188125
lqp	0.9	290	0.82	0.8175	0.0025

(b)

mtry	ntree	nodesize	train_AUC	test_AUC	diff_AUC
3	500	1	0.99999922	0.9772625	0.02273672
3	600	1	0.99999922	0.977225	0.02277422
3	800	1	0.99999922	0.9771	0.02289922
3	600	1	0.99999922	0.977	0.02299922
3	600	1	0.99999922	0.976525	0.02347422
4	600	1	0.99999922	0.9763125	0.02368672
3	800	1	0.99999922	0.9760125	0.02398672
5	600	1	0.99999922	0.97595	0.02404922
3	600	1	0.99999922	0.975825	0.02417422
5	500	1	0.99999922	0.9757375	0.02426172

(c)

mtry	ntree	nodesize	train_AUC	test_AUC	diff_AUC
3	500	1	0.99999922	0.9772625	0.02273672
3	600	1	0.99999922	0.977225	0.02277422
3	800	1	0.99999922	0.9771	0.02289922
3	600	1	0.99999922	0.977	0.02299922
3	600	1	0.99999922	0.976525	0.02347422
4	600	1	0.99999922	0.9763125	0.02368672
3	800	1	0.99999922	0.9760125	0.02398672
5	600	1	0.99999922	0.97595	0.02404922
3	600	1	0.99999922	0.975825	0.02417422
5	500	1	0.99999922	0.9757375	0.02426172

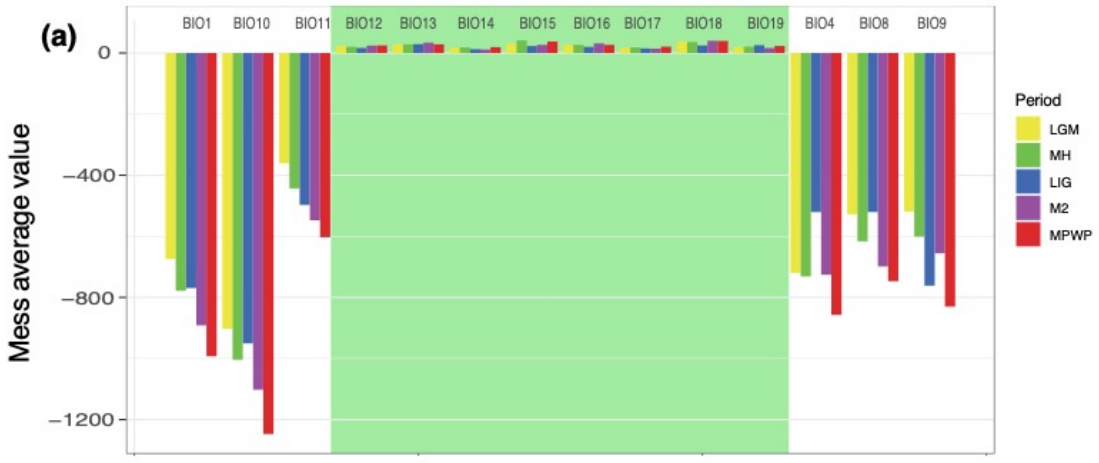
(d)

distribution	n.trees	interaction.depth	shrinkage	bag.fraction	train_AUC	test_AUC	diff_AUC
adaboost	1800	7	0.028	0.2	0.99716484	0.98	0.01716484
adaboost	2400	4	0.028	0.2	0.99517734	0.979775	0.01540234
adaboost	2400	4	0.028	0.2	0.99527734	0.97955	0.01572734
adaboost	2400	4	0.028	0.2	0.99516484	0.979525	0.01563984
adaboost	2400	7	0.028	0.2	0.99873672	0.979475	0.01926172
adaboost	1800	7	0.028	0.2	0.99746328	0.978725	0.01873828
adaboost	1800	4	0.028	0.2	0.99354453	0.978475	0.01506953
adaboost	2400	4	0.028	0.2	0.99549141	0.97755	0.01794141
adaboost	2400	4	0.028	0.2	0.99573359	0.9768	0.01893359
adaboost	1800	4	0.028	0.2	0.99359297	0.976525	0.01706797

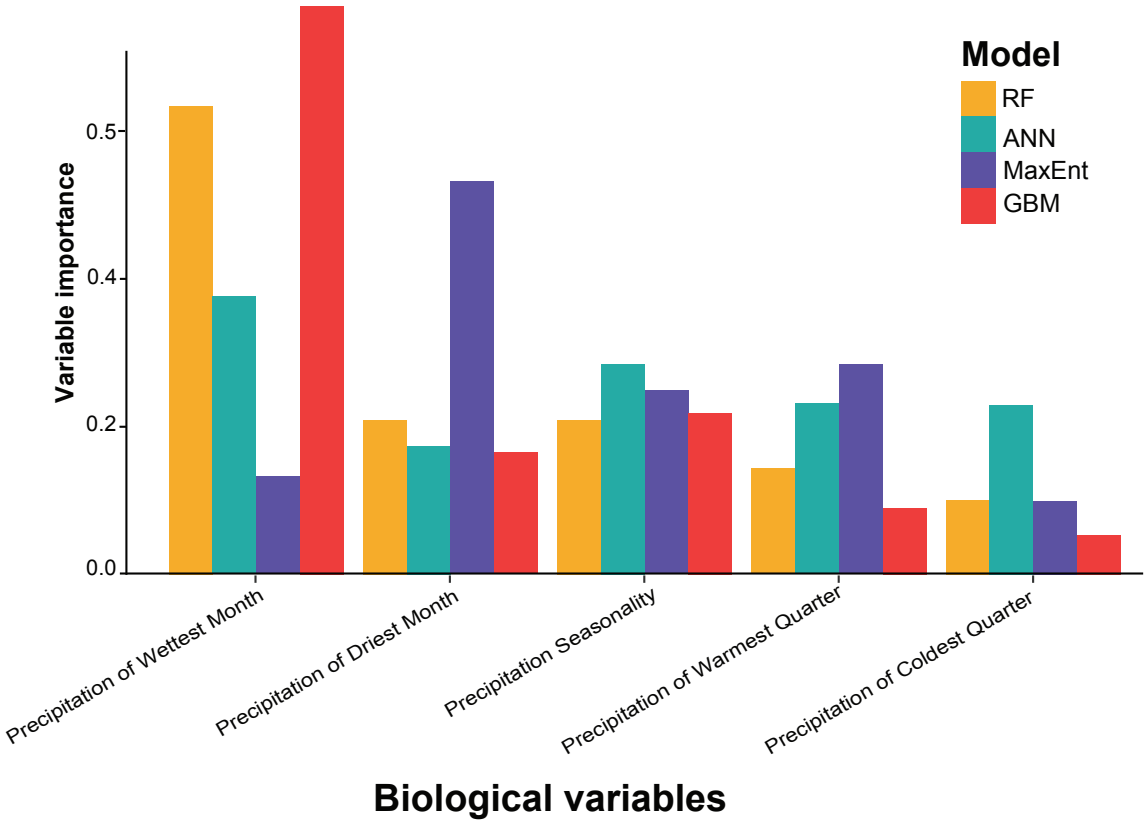
Table S2 ROC score of each model for each run with the associated block. Run with a ROC score over 0.8 (in bold) were retained for the ensemble model.

Block	1	2	3	4	1	2	3	4
Run	1	1	1	1	2	2	2	2
RF	**0.864**	**0.963**	**0.924**	**0.966**	**0.864**	**0.963**	**0.924**	**0.966**
ANN	0.702	**0.909**	**0.925**	**0.935**	**0.903**	**0.902**	**0.899**	**0.943**
MAXENT	0.592	**0.891**	0.770	**0.939**	0.592	**0.891**	0.770	**0.939**
GBM	**0.888**	**0.955**	**0.890**	**0.971**	**0.888**	**0.955**	**0.890**	**0.971**
Run	3	3	3	3	4	4	4	4
RF	**0.864**	**0.963**	**0.924**	**0.966**	**0.864**	**0.963**	**0.924**	**0.966**
ANN	**0.887**	**0.911**	**0.925**	**0.954**	**0.885**	**0.935**	**0.890**	**0.925**
MAXENT	0.592	**0.891**	0.770	**0.939**	0.592	**0.891**	0.770	**0.939**
GBM	**0.888**	**0.955**	**0.890**	**0.971**	**0.888**	**0.955**	**0.890**	**0.971**

Appendix C2 Figure 1 A) Result of MESS analysis for each set of bioclimatic variables for LGM, Mid-Holocene, Last Inter-Glacial period M2 and MPWP.



Appendix C2 Figure 2 Variable importance from the ENM.



Appendix C2 Table 3: Individual *Trichosurus* used for the different phylogenetic analyses with their associated location name, longitude, latitude and species and subspecies name. Australian samples came from the Australian National Wildlife Collection ANWC, and the Australian Biological Tissue Collection ABTC.

SAMPLE NAME	LOCATION	LONGITUDE	LATITUDE	*Trichosurus* SPECIES	*Trichosurus vulpecula* SUBSPECIES	*MITOCHONDRION* GENBANK	AUSTRALIAN SAMPLE ID	SAMPLE CODES
NSW_TC_1	New South Wales, Australia	152.995	-30.165	*caninus*	NA	ON400096	ANWCM16324	1Tc1b
NSW_TC_2	New South Wales, Australia	152.995	-30.165	*caninus*	NA	ON400092	ANWCM24178	2Tc2b
BI_TVH_1	Barrow Island, Western Australia	115.4	-20.79	*vulpecula*	*hypoleucus*	ON400081	ABTC62933	6WA12
CO_TVH_1	Collie, Western Australia	116.15	-33.35	*vulpecula*	*hypoleucus*	ON400079	ABTC83707	7WA13
CO_TVH_2	Collie, Western Australia	116.15	-33.35	*vulpecula*	*hypoleucus*	ON400080	ABTC56745	10WA4
PZ_TVH_1	Perth Zoo, Western Australia	115.85	-31.95	*vulpecula*	*hypoleucus*	ON400082	ABTC18396	11WA7
PZ_TVH_2	Perth Zoo, Western Australia	115.85	-31.95	*vulpecula*	*hypoleucus*	ON400083	ABTC18394	Pos18394
PZ_TVH_3	Perth Zoo, Western Australia	115.85	-31.95	*vulpecula*	*hypoleucus*	ON400078	ABTC18395	Pos18395
TU_TV_1	Turitea, New Zealand	175.6535	-40.415	*vulpecula*	***vulpecula/fulginosus***	ON400086	NA	PosTuGr2
TU_TV_2	Turitea, New Zealand	175.6535	-40.415	*vulpecula*	***vulpecula/fulginosus***	ON400087	NA	3Tur9b
TU_TV_3	Turitea, New Zealand	175.6535	-40.415	*vulpecula*	***vulpecula/fulginosus***	ON400085	NA	4Tur10b
TU_TV_4	Turitea, New Zealand	175.6535	-40.415	*vulpecula*	***vulpecula/fulginosus***	ON400088	NA	5Tur11
AC_TVV_1	Australian Capital Territory	149.1333	-35.217	*vulpecula*	*vulpecula*	ON400084	ANWCM28611	Pos28611
TA_TVF_1	Launceston, Tasmania	147.08	-41.26	*vulpecula*	*fulginosus*	ON400091	ABTC50745	Pos56745
DA_TVA_1	Darwin, Northern Territory	130.84	-12.46	*vulpecula*	*arnhemensis*	ON400093	ABTC50841	POS_M1
DA_TVA_2	Darwin, Northern Territory	130.84	-12.46	*vulpecula*	*arnhemensis*	ON400095	ABTC50842	POS_Wa_6
DA_TVA_3	Darwin, Northern Territory	130.84	-12.46	*vulpecula*	*arnhemensis*	ON400089	ABTC50839	8NT14
DA_TVA_4	Darwin, Northern Territory	130.84	-12.46	*vulpecula*	*arnhemensis*	ON400090	ABTC50840	9NT15
KI_TVA_1	Kimberley,Western Australia	125.066	-16.983	*vulpecula*	*arnhemensis*	ON400094	ABTC7828	Pos7828

Appendix C2 Table 4 Table of samples information, the majority of the sample are coming from Australian National Wildlife Collection of Mammals (2015-01-08).

Reg Number	*State*	*Location*	*Latitude*	*Longitude*	*Collector*	*Catalogue record*	*Genebank Accession #*	*Subspecies*
50839 DA_TVA_1 DA_TVA_2 DA_TVA_3 DA_TVA_4 50840 50841 50842 50844 50845 50846 50847 50848	NT	Darwin	-12.381	130.987	NA	554685 554686 554687 554688 554690 554691 554692 554693 554694	ON342631 ON342630 ON342628 ON342618 ON342629 ON342622 ON342621 ON342620 ON342619 ON342625 ON342624 ON342626 ON342627	*arnhemensis*
KI_TVA_1	WA	Kimberley	-16.983	125.066	NA	NA	ON342623	*arnhemensis*
50750 50751 50752 50753 50754 50746 50747 50748 50749 TA_TVF_1	Tas	Launceston	-41.433	147.133	Coman, Brian	554592 554593 554594 554595 554596 554588 554589 554590 554591	ON342655 ON342677 ON342676 ON342653 ON342678 ON342680 ON342679 ON342654 ON342681 ON342644	*fulginosus*
BI_TvH_1	WA	Barrow Island	-20.790	115.400	NA	NA	ON342722	*hypoleucus*
CO_TvH_1 CO_TvH_2	WA	Collie	-33.350	116.150	NA	NA	ON342721 ON342723	*hypoleucus*
PZ_TvH_1 PZ_TvH_2 PZ_TvH_3	WA	Perth Zoo	-31.950	115.850	NA	NA	ON342718 ON342719 ON342720	*hypoleucus*
50815 50825 50816 50817 50819 50823 50824	Qld	Townsville	-19.215	146.771	NA	554661 554662 554663 554665 554669 554670	ON342685 ON342686 ON342687 ON342688 ON342640 ON342642	*johnsoni*
Ac_TVV_1	ACT	Australian Capital Territory	-35.217	149.133	NA	NA	ON342643	*vulpecula*
50850 50851 50852 50853 50855 50857	NSW	Armidale	-30.442	151.608	NA	554696 554697 554698 554699 554701 554703	ON342701 ON342703 ON342704 ON342702 ON342700 ON342682	*vulpecula*

50827 50828 50829 50830 50831 50833 50834 50835	Qld	Moggill area Brisbane	-27.593	152.863	NA	554673 554674 554675 554676 554677 554679 554680 554681	ON342693 ON342694 ON342689 ON342708 ON342691 ON342690 ON342692 ON342632	*vulpecula*
50784 50785 50786 50787 50788 50789 50790 50791 50792	SA	Kingscote, Kangaroo Island	-35.650	137.633	Coman, Brian	554626 554627 554628 554629 554630 554631 554632 554633 554634	ON342713 ON342714 ON342709 ON342710 ON342716 ON342717 ON342711 ON342712 ON342715	*vulpecula*
73756 7829 7830 132145 132163 132202 27443 10734	SA	Adelaide	-34.751	138.897	NA	NA	ON342707 ON342699 ON342705 ON342698 ON342696 ON342706 ON342695 ON342697	*vulpecula*
50707 50708 50710 50711 50712 50713 50714 50715 50716	Vic	Sutton Grange	-36.983	144.350	Coman, Brian	554542 554543 554546 554547 554548 554549 554550 554551 554552	ON342657 ON342660 ON342659 ON342661 ON342649 ON342650 ON342651 ON342652 ON342648	*vulpecula*
50728 50729 50730 50731 50732 50733 50734 50735 50736 50737	Vic	Tang Tang Swamp, Bendigo	-36.367	144.300	Coman, Brian	554564 554565 554566 554567 554568 554569 554570 554571 554572 554573	ON342636 ON342638 ON342683 ON342635 ON342634 ON342658 ON342684 ON342637 ON342656 ON342639	*vulpecula*
TU_Tv_1 TU_Tv_2 TU_Tv_3 TU_Tv_4	NZ	Turitea	-40.415	175.654	NA	NA	ON342647 ON342645 ON342633 ON342646	*vulpecula & fulginosus*
Tvul_Tara_10 Tvul_Tara_12 Tvul_Tara_13 Tvul_Tara_15 Tvul_Tara_16 Tvul_Tara_17 Tvul_Tara_19 Tvul_Tara_20 Tvul_Tara_21 Tvul_Tara_22 Tvul_Tara_23 Tvul_Tara_24 Tvul_Tara_25 Tvul_Tara_26	NZ	Taranaki	-39.166	173.960	NA	NA	ON342672 ON342668 ON342674 ON342665 ON342667 ON342666 ON342662 ON342669 ON342663 ON342671 ON342673 ON342664 ON342670 ON342675	*vulpecula & fulginosus*

Appendix C2 Table 5 Marsupial mtDNA genomes used for the molecular clock analysis with the species names, Genbank accession numbers and the associated publications.

Species name	Genbank accession number	Publication
Macropus robustus	NC_001794	(Janke et al., 1997)
Macropus giganteus	KY996502	(Nilsson et al., 2018)
Lagorchestes hirsutus	NC_008136	(Munemasa et al., 2006)
Lagostrophus fasciatus	NC_008447	(Nilsson, 2006)
Potorous tridactylus	NC_006524	(Nilsson et al., 2004)
Phalanger vestitus	NC_008137	(Munemasa et al., 2006)
Sminthopsis crassicaudata	AY795974	(Phillips et al., 2006)
Neophascogale lorentzi	KJ868130	(Mitchell et al., 2014)
Dasyurus hallucatus	AY795973	(Phillips et al., 2006)
Sarcophilus harrisii	NC_018788	(Miller et al., 2011)
Paramurexia rothschildi	KJ868134	(Mitchell et al., 2014)

Janke, A., Xu, X., & Arnason, U. (1997). The complete mitochondrial genome of the wallaroo (Macropus robustus) and the phylogenetic relationship among Monotremata, Marsupialia, and Eutheria. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(4), 1276–1281. https://doi.org/10.1073/pnas.94.4.1276

Miller, W., Hayes, V. M., Ratan, A., Petersen, D. C., Wittekindt, N. E., Miller, J., Walenz, B., Knight, J., Qi, J., Zhao, F., Wang, Q., Bedoya-Reina, O. C., Katiyar, N., Tomsho, L. P., Kasson, L. M. C., Hardie, R. A., Woodbridge, P., Tindall, E. A., Bertelsen, M. F., … Schuster, S. C. (2011). Genetic diversity and population structure of the endangered marsupial Sarcophilus harrisii (Tasmanian devil). *Proceedings of the National Academy of Sciences of the United States of America*, *108*(30), 12348–12353. https://doi.org/10.1073/pnas.1102838108

Mitchell, K. J., Pratt, R. C., Watson, L. N., Gibb, G. C., Llamas, B., Kasper, M., Edson, J., Hopwood, B., Male, D., Armstrong, K. N., Meyer, M., Hofreiter, M., Austin, J., Donnellan, S. C., Lee, M. S. Y., Phillips, M. J., & Cooper, A. (2014). Molecular phylogeny, biogeography, and habitat preference evolution of Marsupials. *Molecular Biology and Evolution*, *31*(9), 2322–2330. https://doi.org/10.1093/molbev/msu176

Munemasa, M., Nikaido, M., Donnellan, S., Austin, C. C., Okada, N., & Hasegawa, M. (2006). Phylogenetic analysis of diprotodontian marsupials based on complete mitochondrial genomes. *Genes and Genetic Systems*, *81*(3), 181–191. https://doi.org/10.1266/ggs.81.181

Nilsson, M. A. (2006). Phylogenetic relationships of the Banded Hare wallaby (Lagostrophus fasciatus) and a map of the kangaroo mitochondrial control region. *Zoologica Scripta*, *35*(4), 387–393. https://doi.org/10.1111/j.1463-6409.2006.00237.x

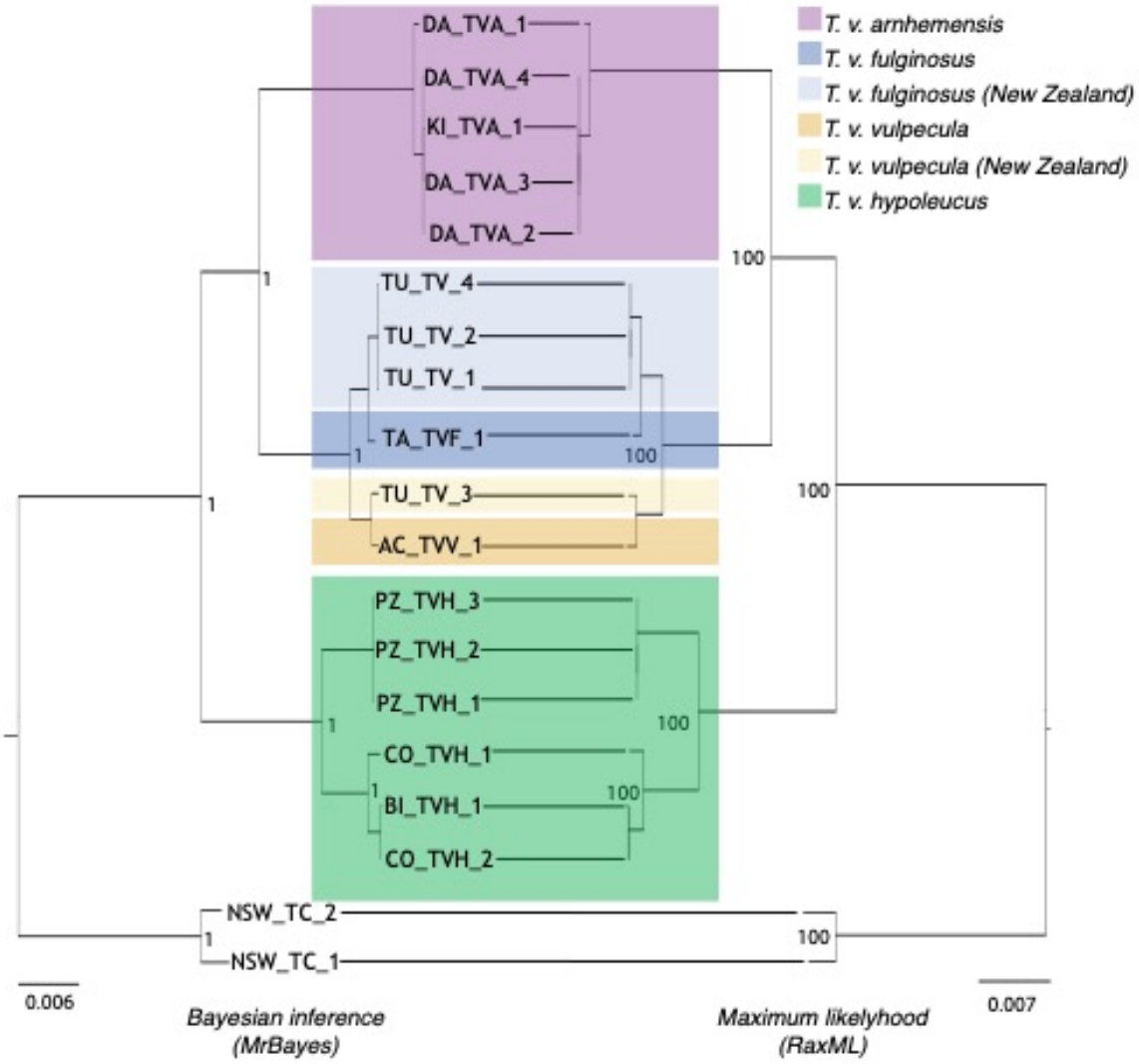
Nilsson, M. A., Arnason, U., Spencer, P. B. S., & Janke, A. (2004). Marsupial relationships and a timeline for marsupial radiation in South Gondwana. *Gene*, *340*(2), 189–196. https://doi.org/10.1016/j.gene.2004.07.040

Nilsson, M. A., Zheng, Y., Kumar, V., Phillips, M. J., & Janke, A. (2018). Speciation generates mosaic genomes in Kangaroos. *Genome Biology and Evolution*, *10*(1), 33–44. https://doi.org/10.1093/gbe/evx245

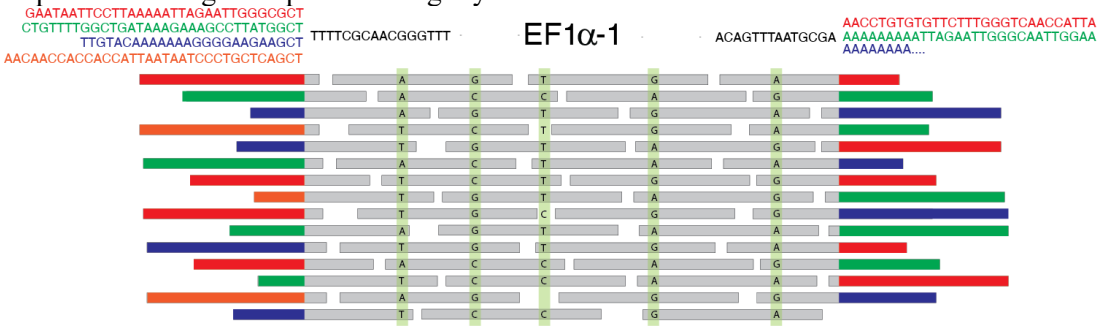
Phillips, M., McLenachan, P., Down, C., Gibb, G., & Penny, D. (2006). Combined

mitochondrial and nuclear DNA sequences resolve the interrelations of the major Australasian marsupial radiations. *Systematic Biology*, *55*(1), 122–137. https://doi.org/10.1080/10635150500481614

Appendix C2 Figure 3: Phylogenetic tree of the different MtDNA lineages inferred using Bayesian inference methods (MrBayes) and maximum likelihood methods (RaXML). Numbers at nodes indicate posterior probabilities for the Bayesian inference analysis and bootstrap values for maximum likelihood analysis.



Appendix C2 Figure 4: Schematic representation of consistent nucleotide variants in the nuclear sequences leading to sequence ambiguity.



Appendix C2 Table 6: Three examples of sequence ambiguity in *Trichosurus vulpecula*. Proportion (%) of the two most present nucleotides in the stack of read of a sample of variables sites of the ITS1, ITS2 and EF1α-1 loci.

ITS1

Sample		Site 8	Site 321	Site 345	Site 375	Site 424	Site 976	Site 1095	Site 1695	Site 1733	Site 1765
PosWA_6	Variant 1	**G** 61.5	**T** 70.6	**A** 27.4	**C** 24.2	**A** 79.2	**A** 48.9	**C** 24.4	**A** 42.8	**C** 55.4	**A** 28.6
	Variant 2	**C** 38.0	**C** 29.4	**G** 72.6	**T** 75.3	**G** 20.8	**C** 51.1	**G** 75.1	**G** 57.2	**T** 44.6	**G** 71.0
	Read depth	186	203	225	220	244	132	220	172	194	261
Pos56745	Variant 1	**G** 50	**T** 99.2	**A** 100	**C** 2.3	**A** 58.0	**C** 99.0	**G** 100.0	**A** 48.8	**C** 46.2	**A** 56.6
	Variant 2	**C** 48.7	**C** 0.8	**G** 0	**T** 97.7	**G** 42.0	**T** 1.0	**T** 0.0	**G** 51.2	**T** 53.8	**G** 43.4
	Read depth	89	118	120	132	155	100	103	127	137	142

ITS2

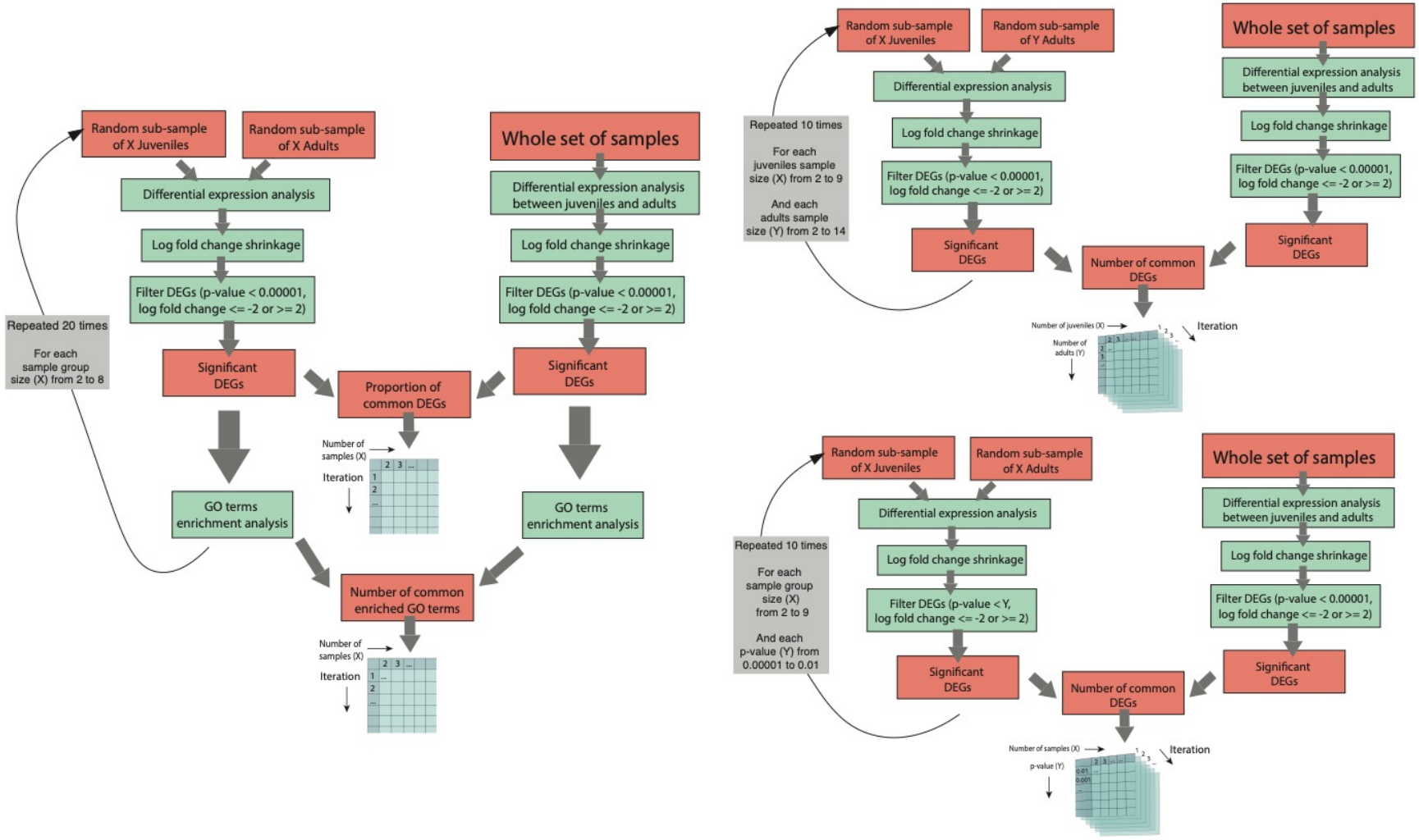
Sample		Site 150	Site 210	Site 321	Site 467	Site 567
PosWA_6	Variant 1	**A** 26.4	**A** 52	**C** 69.7	**C** 68.2	**G** 46.1
	Variant 2	**C** 73.6	**T** 48	**T** 29.3	**T** 31.1	**T** 53.9
	Read depth	260	97	207	150	101
Pos56745	Variant 1	**A** 12.7	**C** 54.5	**C** 93.8	**C** 33.7	**G** 100.0
	Variant 2	**C** 87.3	**T** 45.5	**T** 5.2	**T** 66.3	**T** 0.0
	Read depth	67	25	72	85	75

EF1α-1

Sample		Site 111	Site 271	Site 717	Site 1333	Site 1505
PosM1	Variant 1	**A** 39.5	**G** 59.5	**C** 54.8	**A** 66.7	**G** 50
	Variant 2	**C** 55.3	**A** 38.1	**T** 38.7	**G** 31.6	**A** 39.6
	Read depth	37	41	61	56	47
PosWA_6	Variant 1	**A** 40.9	**G** 41.8	**C** 67.7	**A** 72.8	**G** 52.8
	Variant 2	**C** 59.1	**A** 50.9	**T** 21.0	**G** 25.9	**A** 27.8
	Read depth	43	54	61	80	37

Chapter 3

Appendix C3 Figure 1: Flowchart of the algorithms used to assess the effect of sample size on DEGs and enriched GOs discovery.



Appendix C3 Table 1: List of differentially expressed genes between adults and juveniles identified using **DESeq2** with associated log2 FoldChange adjusted p-value using False Discovery Rate (FDR) calculated using Benjamini and Hochberg method.

GENES	GENES NAME	SHRUNK FOLDCHANGE	FDR ADJUSTED P-VALUE
SLC4A1	solute carrier family 4 member 1(SLC4A1)	5.156	1.56E-12
ANKEF1	ankyrin repeat and EF-hand domain containing 1(ANKEF1)	4.592	1.10E-14
WISP1	cellular communication network factor 4	4.254	1.27E-12
AHSP	alpha hemoglobin stabilizing protein(AHSP)	4.209	1.67E-13
KEL	Kell metallo-endopeptidase(KEL)	4.168	1.69E-06
MYRIP	myosin VIIA and Rab interacting protein(MYRIP)	4.099	2.76E-15
SPTA1	spectrin alpha, erythrocytic 1(SPTA1)	4.098	1.56E-09
CCDC3	coiled-coil domain containing 3(CCDC3)	4.005	6.26E-11
EPO	erythropoietin(EPO)	3.986	6.55E-12
MMP16	matrix metallopeptidase 16(MMP16)	3.971	8.08E-18
CBX2	chromobox 2(CBX2)	3.961	2.56E-18
TRIM63	tripartite motif containing 63(TRIM63)	3.883	2.15E-11
FBN2B	fibrillin 2B	3.832	2.37E-10
RBP7	retinol binding protein 7(RBP7)	3.793	5.22E-10
RAG1	recombination activating 1(RAG1)	3.724	3.82E-09
LMO3	LIM domain only 3(LMO3)	3.616	7.52E-09
PTCHD4	patched domain containing 4(PTCHD4)	3.613	1.51E-10
MDK	midkine(MDK)	3.606	7.58E-10
CCK	cholecystokinin(CCK)	3.605	1.59E-11
STK31	serine/threonine kinase 31(STK31)	3.595	7.24E-09
PAQR8	progestin and adipoQ receptor family member 8(PAQR8)	3.569	1.25E-16
PITX3	paired like homeodomain 3(PITX3)	3.525	2.69E-10
TRIM71	tripartite motif containing 71(TRIM71)	3.429	4.49E-11
LRRC17	leucine rich repeat containing 17(LRRC17)	3.406	4.54E-14
TNFRSF11B	TNF receptor superfamily member 11b(TNFRSF11B)	3.393	2.48E-09
FOXJ1	forkhead box J1(FOXJ1)	3.367	7.62E-13
EPOR	erythropoietin receptor(EPOR)	3.360	1.42E-07
VILL	villin like(VILL)	3.342	1.47E-13
OLFM2	olfactomedin 2(OLFM2)	3.336	4.01E-14
CASS4	Cas scaffold protein family member 4(CASS4)	3.323	6.32E-09
CYSLTR1	cysteinyl leukotriene receptor 1(CYSLTR1)	3.318	4.23E-14
RUNDC3A	RUN domain containing 3A(RUNDC3A)	3.295	1.90E-08
SLAIN1	SLAIN motif family member 1(SLAIN1)	3.275	1.90E-13
HCAR1	hydroxycarboxylic acid receptor 1(HCAR1)	3.269	3.14E-09
CLDN19.S	Claudin 19	3.248	3.73E-07
ADAMTS17	ADAM metallopeptidase with thrombospondin type 1 motif 17(ADAMTS17)	3.242	1.58E-08
COL1A1	collagen type I alpha 1 chain(COL1A1)	3.230	2.02E-11
UPK3A	uroplakin 3A(UPK3A)	3.210	1.48E-09
PRSS35	serine protease 35(PRSS35)	3.209	4.45E-08
PHLDA2	pleckstrin homology like domain family A member 2(PHLDA2)	3.209	2.58E-10
ALAS2	5'-aminolevulinate synthase 2(ALAS2)	3.200	3.93E-10
HAPLN1	hyaluronan and proteoglycan link protein 1(HAPLN1)	3.191	2.46E-07
SPON2	spondin 2(SPON2)	3.190	4.28E-09
S100P	S100 calcium binding protein P(S100P)	3.188	2.86E-11
HIPK4	homeodomain interacting protein kinase 4(HIPK4)	3.180	2.00E-07
GPC5	glypican 5(GPC5)	3.170	1.61E-13
EVC	EvC ciliary complex subunit 1(EVC)	3.168	9.27E-21
LRP4	LDL receptor related protein 4(LRP4)	3.156	1.18E-15
SLC22A8	solute carrier family 22 member 8(SLC22A8)	3.154	6.01E-07
TMEM54	transmembrane protein 54(TMEM54)	3.146	1.73E-10
BLOC1S2	biogenesis of lysosomal organelles complex 1 subunit 2(BLOC1S2)	3.115	6.41E-07
C1QTNF6	C1q and TNF related 6(C1QTNF6)	3.110	4.06E-07
BEST4	bestrophin 4(BEST4)	3.085	1.45E-08

TM4SF5	transmembrane 4 L six family member 5(TM4SF5)	3.084	7.30E-16
ADGRB2	adhesion G protein-coupled receptor B2(ADGRB2)	3.062	1.23E-09
HTR5A	5-hydroxytryptamine receptor 5A(HTR5A)	3.061	4.46E-09
TDH	L-threonine dehydrogenase(TDH)	3.057	1.39E-08
LOC101696428	Locus 101696428	3.052	6.47E-07
UGT8A	UDP galactosyltransferase 8A	3.051	6.36E-14
1110017D15RIK	RIKEN cDNA 1110017D15 gene	3.045	1.36E-07
MYCN	MYCN proto-oncogene, bHLH transcription factor(MYCN)	3.028	7.53E-08
ACBD7	acyl-CoA binding domain containing 7(ACBD7)	3.006	2.08E-06
TDRD10	tudor domain containing 10(TDRD10)	2.994	1.10E-06
LIPC	lipase C, hepatic type(LIPC)	2.981	1.88E-06
MPO	myeloperoxidase(MPO)	2.978	7.77E-08
WNK4	WNK lysine deficient protein kinase 4(WNK4)	2.963	5.21E-08
SLC26A10	solute carrier family 26 member 10(SLC26A10)	2.943	6.14E-10
MMP2	matrix metallopeptidase 2(MMP2)	2.932	1.72E-24
VWDE	von Willebrand factor D and EGF domains(VWDE)	2.927	1.26E-06
TNNT3	troponin T3, fast skeletal type(TNNT3)	2.922	3.86E-10
APLNR	apelin receptor(APLNR)	2.910	7.06E-08
APLNRA	apelin receptor a	2.903	2.46E-10
SOX11	SRY-box transcription factor 11(SOX11)	2.897	3.23E-06
NOX3	NADPH oxidase 3(NOX3)	2.896	5.34E-06
MEGF6	multiple EGF like domains 6(MEGF6)	2.894	2.47E-08
XKR5	XK related 5(XKR5)	2.889	1.71E-06
CACNA2D1	calcium voltage-gated channel auxiliary subunit alpha2delta 1(CACNA2D1)	2.887	4.45E-17
EFNA3	ephrin A3(EFNA3)	2.876	2.36E-11
MYRF	myelin regulatory factor(MYRF)	2.868	2.47E-18
COL1A2	collagen type I alpha 2 chain(COL1A2)	2.864	6.94E-12
ADAMTSL1	ADAMTS like 1(ADAMTSL1)	2.851	1.96E-06
MFAP2	microfibril associated protein 2(MFAP2)	2.845	1.52E-06
NR4A1	nuclear receptor subfamily 4 group A member 1(NR4A1)	2.835	8.51E-09
DACT1	dishevelled binding antagonist of beta catenin 1(DACT1)	2.833	1.35E-08
PRDX2	peroxiredoxin 2(PRDX2)	2.826	1.39E-17
NCAM1	neural cell adhesion molecule 1(NCAM1)	2.820	9.55E-07
GDF15	growth differentiation factor 15(GDF15)	2.812	5.64E-09
HIF3A	hypoxia inducible factor 3 subunit alpha(HIF3A)	2.810	2.77E-07
TUBB1	tubulin beta 1 class VI(TUBB1)	2.800	4.00E-07
RASD1	ras related dexamethasone induced 1(RASD1)	2.799	1.50E-06
LOC101712013	Locus 101712013	2.798	2.80E-08
COL7A1	collagen type VII alpha 1 chain	2.786	9.81E-09
TSPAN5	tetraspanin 5(TSPAN5)	2.785	3.51E-07
KCP	kielin cysteine rich BMP regulator(KCP)	2.784	1.05E-11
TRIM10	tripartite motif containing 10(TRIM10)	2.781	4.16E-09
DGKI	diacylglycerol kinase iota(DGKI)	2.762	7.68E-06
SIGIRR	single Ig and TIR domain containing(SIGIRR)	2.762	1.81E-12
CDC20	cell division cycle 20(CDC20)	2.753	6.20E-08
KLC3	kinesin light chain 3(KLC3)	2.751	4.81E-08
AFP	alpha fetoprotein(AFP)	2.746	5.38E-09
IRS4	insulin receptor substrate 4(IRS4)	2.734	4.75E-06
CNIH2	cornichon family AMPA receptor auxiliary protein 2(CNIH2)	2.730	1.50E-07
EVC2	EvC ciliary complex subunit 2(EVC2)	2.718	3.10E-10
IGSF22	immunoglobulin superfamily member 22(IGSF22)	2.709	5.52E-06
SLC25A12	solute carrier family 25 member 12(SLC25A12)	2.707	2.11E-14
AMOT	angiomotin(AMOT)	2.705	2.30E-20
TEKT1	tektin 1(TEKT1)	2.691	8.28E-06
ACE2	angiotensin converting enzyme 2(ACE2)	2.680	1.73E-06
FAM131C	family with sequence similarity 131 member C(FAM131C)	2.677	6.23E-06
PCDH11X	protocadherin 11 X-linked(PCDH11X)	2.676	6.40E-06
CYGB	cytoglobin(CYGB)	2.655	2.03E-10
ALPI	alkaline phosphatase, intestinal(ALPI)	2.654	9.21E-06
ADAMTS6	ADAM metallopeptidase with thrombospondin type 1 motif 6(ADAMTS6)	2.646	2.21E-06

KIF4A	kinesin family member 4A(KIF4A)	2.644	7.60E-14
FRAS1	Fraser extracellular matrix complex subunit 1(FRAS1)	2.643	2.71E-06
RNASEH2A	ribonuclease H2 subunit A(RNASEH2A)	2.641	2.48E-10
FNDC4	fibronectin type III domain containing 4(FNDC4)	2.637	9.78E-07
SPARC	secreted protein acidic and cysteine rich(SPARC)	2.636	6.23E-15
ROBO1	roundabout guidance receptor 1(ROBO1)	2.635	4.75E-06
TSPO2	translocator protein 2(TSPO2)	2.627	2.44E-06
B4GALNT3	beta-1,4-N-acetyl-galactosaminyltransferase 3(B4GALNT3)	2.626	7.65E-07
KIF18A	kinesin family member 18A(KIF18A)	2.623	3.12E-15
MCCD1	mitochondrial coiled-coil domain 1(MCCD1)	2.623	1.61E-06
GPRC5C	G protein-coupled receptor class C group 5 member C(GPRC5C)	2.609	1.33E-07
ECRG4	ECRG4 augurin precursor	2.604	1.21E-07
KRT222	keratin 222(KRT222)	2.600	1.10E-06
COL13A1	collagen type XIII alpha 1 chain(COL13A1)	2.599	7.26E-06
HORMAD2	HORMA domain containing 2(HORMAD2)	2.593	1.80E-06
HS3ST5	heparan sulfate-glucosamine 3-sulfotransferase 5(HS3ST5)	2.579	2.17E-07
EFEMP2	EGF containing fibulin extracellular matrix protein 2(EFEMP2)	2.571	2.11E-12
IL17RB	interleukin 17 receptor B(IL17RB)	2.569	2.18E-13
TMC6	transmembrane channel like 6(TMC6)	2.567	4.44E-08
CRISPLD1	cysteine rich secretory protein LCCL domain containing 1(CRISPLD1)	2.562	3.76E-16
NSG1	neuronal vesicle trafficking associated 1(NSG1)	2.560	7.80E-07
DPEP2	dipeptidase 2(DPEP2)	2.553	2.29E-06
CENPP	centromere protein P(CENPP)	2.552	9.27E-07
LAMP2	lysosomal associated membrane protein 2(LAMP2)	2.545	5.92E-10
RGS1	regulator of G protein signaling 1(RGS1)	2.540	1.91E-08
BUB1	BUB1 mitotic checkpoint serine/threonine kinase(BUB1)	2.538	8.44E-11
AIPL1	aryl hydrocarbon receptor interacting protein like 1(AIPL1)	2.534	4.39E-07
MGAT4C	MGAT4 family member C(MGAT4C)	2.531	1.63E-13
STMN1	stathmin 1(STMN1)	2.526	7.07E-07
VSIG4	V-set and immunoglobulin domain containing 4(VSIG4)	2.525	1.69E-09
NCAPD2	non-SMC condensin I complex subunit D2(NCAPD2)	2.524	2.45E-11
COL14A1	collagen type XIV alpha 1 chain(COL14A1)	2.505	5.26E-12
UNC5C	unc-5 netrin receptor C(UNC5C)	2.505	7.00E-07
MIA	MIA SH3 domain containing(MIA)	2.499	1.70E-07
TNFRSF19	TNF receptor superfamily member 19(TNFRSF19)	2.491	1.67E-06
EGFLAM	EGF like, fibronectin type III and laminin G domains(EGFLAM)	2.487	3.14E-06
ENPEP	glutamyl aminopeptidase(ENPEP)	2.485	4.97E-07
MYCL	MYCL proto-oncogene, bHLH transcription factor(MYCL)	2.484	2.67E-09
MYL1	myosin light chain 1(MYL1)	2.483	7.40E-06
TGIF2	TGFB induced factor homeobox 2(TGIF2)	2.479	3.23E-09
SPC24	SPC24 component of NDC80 kinetochore complex(SPC24)	2.478	1.62E-07
EXO1	exonuclease 1(EXO1)	2.473	2.87E-06
SALL4	spalt like transcription factor 4(SALL4)	2.472	5.29E-06
GALK1	galactokinase 1(GALK1)	2.471	6.33E-11
ORC1	origin recognition complex subunit 1(ORC1)	2.469	5.22E-08
ITGA8	integrin subunit alpha 8(ITGA8)	2.466	6.89E-07
CIART	circadian associated repressor of transcription(CIART)	2.452	2.48E-06
RCN3	reticulocalbin 3(RCN3)	2.447	1.00E-08
SLC27A3	solute carrier family 27 member 3(SLC27A3)	2.438	2.13E-07
TONSL	tonsoku like, DNA repair protein(TONSL)	2.438	1.88E-07
ADRB2	adrenoceptor beta 2(ADRB2)	2.436	4.17E-09
NRXN2	neurexin 2(NRXN2)	2.432	1.26E-08
PTHLH	parathyroid hormone like hormone(PTHLH)	2.429	6.98E-07
LHFPL1	LHFPL tetraspan subfamily member 1(LHFPL1)	2.428	1.32E-07
HBA1	hemoglobin subunit alpha 1(HBA1)	2.427	3.45E-08
FSD1	fibronectin type III and SPRY domain containing 1(FSD1)	2.422	4.30E-10
SERPINI1	serpin family I member 1(SERPINI1)	2.417	6.86E-08
TST	thiosulfate sulfurtransferase(TST)	2.415	7.19E-12
MCM5	minichromosome maintenance complex component 5(MCM5)	2.413	7.80E-10

SYT8	synaptotagmin 8(SYT8)	2.412	9.32E-06
IGFBP3	insulin like growth factor binding protein 3(IGFBP3)	2.412	9.71E-09
GPR84	G protein-coupled receptor 84(GPR84)	2.406	4.33E-06
ENOX1	ecto-NOX disulfide-thiol exchanger 1(ENOX1)	2.399	7.04E-08
ZNF521	zinc finger protein 521(ZNF521)	2.399	4.74E-10
C2CD4B	C2 calcium dependent domain containing 4B(C2CD4B)	2.398	8.27E-07
COL9A3	collagen type IX alpha 3 chain(COL9A3)	2.398	2.11E-06
ADGRL3	adhesion G protein-coupled receptor L3(ADGRL3)	2.392	1.10E-07
EVPL	envoplakin(EVPL)	2.391	1.62E-08
SFRP4	secreted frizzled related protein 4(SFRP4)	2.388	1.10E-06
RECQL4	RecQ like helicase 4(RECQL4)	2.369	1.90E-09
FST	follistatin(FST)	2.368	2.35E-06
SPP2	secreted phosphoprotein 2(SPP2)	2.367	6.31E-09
CDC6	cell division cycle 6(CDC6)	2.363	1.78E-07
CDKN3	cyclin dependent kinase inhibitor 3(CDKN3)	2.362	3.22E-17
SPRY2	sprouty RTK signaling antagonist 2(SPRY2)	2.356	1.79E-11
CYFIP2	cytoplasmic FMR1 interacting protein 2(CYFIP2)	2.351	2.45E-07
RASL11A	RAS like family 11 member A(RASL11A)	2.349	7.30E-07
FKBP10	FKBP prolyl isomerase 10(FKBP10)	2.347	9.66E-12
BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B(BUB1B)	2.343	6.22E-10
SLITRK6	SLIT and NTRK like family member 6(SLITRK6)	2.341	1.62E-06
PHF19	PHD finger protein 19(PHF19)	2.341	3.95E-07
MNS1	meiosis specific nuclear structural 1(MNS1)	2.338	1.33E-06
MTBP	MDM2 binding protein(MTBP)	2.331	7.17E-07
SLC16A11	solute carrier family 16 member 11(SLC16A11)	2.327	1.74E-06
CCDC9B	coiled-coil domain containing 9B(CCDC9B)	2.323	1.02E-06
MEX3A	mex-3 RNA binding family member A(MEX3A)	2.321	1.21E-06
GPX3	glutathione peroxidase 3(GPX3)	2.320	1.45E-07
IFT140	intraflagellar transport 140(IFT140)	2.316	6.49E-18
MEX3B	mex-3 RNA binding family member B(MEX3B)	2.307	4.90E-07
HMGN3	high mobility group nucleosomal binding domain 3(HMGN3)	2.300	1.18E-07
GFPT2	glutamine-fructose-6-phosphate transaminase 2(GFPT2)	2.295	6.26E-08
KCNN2	potassium calcium-activated channel subfamily N member 2(KCNN2)	2.293	6.35E-08
ESPL1	extra spindle pole bodies like 1, separase(ESPL1)	2.291	5.06E-06
FOXM1	forkhead box M1(FOXM1)	2.289	3.81E-12
PCDH7	protocadherin 7(PCDH7)	2.288	2.20E-06
SUSD2	sushi domain containing 2(SUSD2)	2.283	5.05E-09
SI:DKEY-7E14.3	si:dkey-7e14.3	2.283	6.16E-06
ADCY4	adenylate cyclase 4(ADCY4)	2.279	2.35E-10
PDGFRL	platelet derived growth factor receptor like(PDGFRL)	2.277	9.40E-10
COL3A1	collagen type III alpha 1 chain(COL3A1)	2.274	3.56E-09
GNAL	G protein subunit alpha L(GNAL)	2.271	1.61E-06
MXRA5	matrix remodeling associated 5(MXRA5)	2.267	6.07E-07
DLGAP5	DLG associated protein 5(DLGAP5)	2.267	2.85E-06
GRM6	glutamate metabotropic receptor 6(GRM6)	2.261	5.32E-07
KANK4	KN motif and ankyrin repeat domains 4(KANK4)	2.259	2.50E-09
LNX1	ligand of numb-protein X 1(LNX1)	2.259	6.19E-06
PHETA2	PH domain containing endocytic trafficking adaptor 2(PHETA2)	2.250	2.26E-08
IGSF3	immunoglobulin superfamily member 3(IGSF3)	2.250	4.40E-10
DUSP1.S	dual specificity phosphatase 1 S homeolog	2.249	2.07E-06
GJC1	gap junction protein gamma 1(GJC1)	2.247	1.06E-08
FADS1	fatty acid desaturase 1(FADS1)	2.244	1.07E-08
PLK1	polo like kinase 1(PLK1)	2.235	3.64E-06
AHDC1	AT-hook DNA binding motif containing 1(AHDC1)	2.234	6.97E-12
COLEC12	collectin subfamily member 12(COLEC12)	2.232	1.05E-06
DLGAP1	DLG associated protein 1(DLGAP1)	2.231	5.47E-06
DIO3	iodothyronine deiodinase 3(DIO3)	2.228	3.06E-08
OGDHL	oxoglutarate dehydrogenase L(OGDHL)	2.223	4.30E-10
SMO	smoothened, frizzled class receptor(SMO)	2.217	8.38E-09
ANK2	ankyrin 2(ANK2)	2.216	5.35E-07

KCNH2	potassium voltage-gated channel subfamily H member 2(KCNH2)	2.216	1.36E-09
SMARCD3	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3(SMARCD3)	2.212	1.71E-09
NEK2	NIMA related kinase 2(NEK2)	2.211	1.86E-07
EMP1	epithelial membrane protein 1(EMP1)	2.209	6.71E-07
CDCA7L	cell division cycle associated 7 like(CDCA7L)	2.205	3.92E-06
P4HTM	prolyl 4-hydroxylase, transmembrane(P4HTM)	2.201	1.85E-11
NYAP1	neuronal tyrosine phosphorylated phosphoinositide-3-kinase adaptor 1(NYAP1)	2.201	2.57E-12
COL11A2	collagen type XI alpha 2 chain(COL11A2)	2.191	1.41E-09
APITD1	Centromere protein S	2.189	3.08E-07
MTMR7	myotubularin related protein 7(MTMR7)	2.189	1.95E-06
KIF15	kinesin family member 15(KIF15)	2.185	1.45E-06
CNMD	chondromodulin(CNMD)	2.184	7.94E-08
CDH3	cadherin 3(CDH3)	2.183	9.30E-07
ADAMTS14	ADAM metallopeptidase with thrombospondin type 1 motif 14(ADAMTS14)	2.175	5.82E-06
MYC	MYC proto-oncogene, bHLH transcription factor(MYC)	2.172	1.80E-06
SSPN	sarcospan(SSPN)	2.171	2.26E-08
MCM3	minichromosome maintenance complex component 3(MCM3)	2.167	9.67E-08
CAVIN3	caveolae associated protein 3(CAVIN3)	2.157	1.47E-10
COL20A1	collagen type XX alpha 1 chain(COL20A1)	2.157	8.25E-07
GPRIN2	G protein regulated inducer of neurite outgrowth 2(GPRIN2)	2.151	4.59E-06
CDKN2C	cyclin dependent kinase inhibitor 2C(CDKN2C)	2.151	1.55E-07
MLN	motilin(MLN)	2.146	9.77E-10
CLEC11A	C-type lectin domain containing 11A(CLEC11A)	2.146	5.87E-06
NPNT	nephronectin(NPNT)	2.146	8.49E-07
CCT6A	chaperonin containing TCP1 subunit 6A(CCT6A)	2.145	3.46E-06
EPHA4	EPH receptor A4(EPHA4)	2.145	2.51E-06
MYBL2	MYB proto-oncogene like 2(MYBL2)	2.139	1.94E-06
HDAC9	histone deacetylase 9(HDAC9)	2.138	7.26E-07
HIC2	HIC ZBTB transcriptional repressor 2(HIC2)	2.134	5.31E-13
RPS5	ribosomal protein S5(RPS5)	2.130	9.99E-08
NCAPH	non-SMC condensin I complex subunit H(NCAPH)	2.128	3.33E-06
S100A10	S100 calcium binding protein A10(S100A10)	2.127	3.20E-10
COL5A2	collagen type V alpha 2 chain(COL5A2)	2.123	1.59E-07
GMNN.S	geminin, DNA replication inhibitor S homeolog	2.123	1.69E-11
COL4A3	collagen type IV alpha 3 chain(COL4A3)	2.119	1.14E-10
DISC1	DISC1 scaffold protein(DISC1)	2.119	1.82E-06
BCL11A	BAF chromatin remodeling complex subunit BCL11A(BCL11A)	2.117	7.56E-07
PASK	PAS domain containing serine/threonine kinase(PASK)	2.115	4.88E-06
EPHB3	EPH receptor B3(EPHB3)	2.107	1.75E-10
SEMA6C	semaphorin 6C(SEMA6C)	2.103	6.94E-08
LOC100500766	Locus 100500766	2.099	2.84E-10
ARHGAP33	Rho GTPase activating protein 33(ARHGAP33)	2.097	1.94E-08
SP5	Sp5 transcription factor(SP5)	2.091	3.12E-08
PDE9A	phosphodiesterase 9A(PDE9A)	2.090	1.45E-08
MCAM	melanoma cell adhesion molecule(MCAM)	2.089	8.71E-06
CREB3L1	cAMP responsive element binding protein 3 like 1(CREB3L1)	2.088	9.20E-11
TRIM47	tripartite motif containing 47(TRIM47)	2.087	6.24E-11
RDH5	retinol dehydrogenase 5(RDH5)	2.084	2.93E-15
MCM4	minichromosome maintenance complex component 4(MCM4)	2.083	9.99E-07
CHAF1B	chromatin assembly factor 1 subunit B(CHAF1B)	2.080	2.35E-07
COL5A1	collagen type V alpha 1 chain(COL5A1)	2.078	2.64E-07
AMIGO3	adhesion molecule with Ig like domain 3(AMIGO3)	2.075	3.26E-07
AURKA	aurora kinase A(AURKA)	2.073	2.48E-06
FAT3	FAT atypical cadherin 3(FAT3)	2.072	3.03E-06
EEF1G.L	eukaryotic translation elongation factor 1 gamma L homeolog	2.065	6.86E-08
LTBP3	latent transforming growth factor beta binding protein 3(LTBP3)	2.061	4.10E-10

PLEKHG2	pleckstrin homology and RhoGEF domain containing G2(PLEKHG2)	2.057	7.27E-11
ZFHX4	zinc finger homeobox 4(ZFHX4)	2.057	1.42E-06
EFNB3	ephrin B3(EFNB3)	2.056	4.38E-06
RIMKLB	ribosomal modification protein rimK like family member B(RIMKLB)	2.055	6.93E-06
RPS6KA6	ribosomal protein S6 kinase A6(RPS6KA6)	2.042	5.68E-08
C1QTNF4	C1q and TNF related 4(C1QTNF4)	2.040	5.88E-07
MMP23	matrix metallopeptidase 23(MMP23)	2.033	2.02E-07
AVPR2	arginine vasopressin receptor 2(AVPR2)	2.032	4.69E-07
SESTD1	SEC14 and spectrin domain containing 1(SESTD1)	2.032	9.58E-09
TCTEX1D4	Tctex1 domain containing 4	2.025	1.49E-06
NR0B1	nuclear receptor subfamily 0 group B member 1(NR0B1)	2.012	8.20E-07
SOD3	superoxide dismutase 3(SOD3)	2.009	9.32E-07
PPIC	peptidylprolyl isomerase C(PPIC)	2.008	3.14E-06
SMC4	structural maintenance of chromosomes 4(SMC4)	2.007	6.26E-14
TTC39B	tetratricopeptide repeat domain 39B(TTC39B)	-2.012	4.74E-08
CCDC159	coiled-coil domain containing 159(CCDC159)	-2.015	9.42E-07
INSYN1	inhibitory synaptic factor 1(INSYN1)	-2.022	4.55E-06
IDI1	isopentenyl-diphosphate delta isomerase 1(IDI1)	-2.037	1.96E-06
AOX1	aldehyde oxidase 1(AOX1)	-2.044	1.99E-10
ARNTL	aryl hydrocarbon receptor nuclear translocator like(ARNTL)	-2.048	4.55E-15
SBK1	SH3 domain binding kinase 1(SBK1)	-2.049	8.57E-06
PLEKHD1	pleckstrin homology and coiled-coil domain containing D1(PLEKHD1)	-2.074	5.28E-06
SPR	sepiapterin reductase(SPR)	-2.092	9.04E-08
CTSH	cathepsin H(CTSH)	-2.097	1.03E-20
ADIPOR2	adiponectin receptor 2(ADIPOR2)	-2.107	2.99E-17
TMEM41B	transmembrane protein 41B(TMEM41B)	-2.108	7.42E-10
ACSS2	acyl-CoA synthetase short chain family member 2(ACSS2)	-2.110	2.65E-08
XDH	xanthine dehydrogenase(XDH)	-2.115	2.07E-11
AP1AR	adaptor related protein complex 1 associated regulatory protein(AP1AR)	-2.119	2.00E-07
SLC15A1	solute carrier family 15 member 1(SLC15A1)	-2.126	2.16E-06
CTF1	cardiotrophin 1(CTF1)	-2.127	2.82E-07
PSMB8	proteasome 20S subunit beta 8(PSMB8)	-2.130	1.76E-07
SLC22A1	solute carrier family 22 member 1(SLC22A1)	-2.135	1.17E-13
PDK4	pyruvate dehydrogenase kinase 4(PDK4)	-2.135	1.85E-11
KIF16B	kinesin family member 16B(KIF16B)	-2.139	1.80E-14
GFOD1	glucose-fructose oxidoreductase domain containing 1(GFOD1)	-2.142	5.86E-08
MME	membrane metalloendopeptidase(MME)	-2.144	4.92E-07
SMPDL3A	sphingomyelin phosphodiesterase acid like 3A(SMPDL3A)	-2.154	9.13E-09
CAND2	cullin associated and neddylation dissociated 2 (putative)(CAND2)	-2.155	2.86E-09
GBP2	guanylate binding protein 2(GBP2)	-2.163	9.54E-06
NPAS2	neuronal PAS domain protein 2(NPAS2)	-2.177	1.22E-20
TMEM205	transmembrane protein 205(TMEM205)	-2.181	1.97E-11
FAM83G	family with sequence similarity 83 member G(FAM83G)	-2.181	5.11E-13
CAMKK2	calcium/calmodulin dependent protein kinase kinase 2(CAMKK2)	-2.183	2.23E-14
OBSCN	obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF(OBSCN)	-2.183	3.31E-06
SLC25A47	solute carrier family 25 member 47(SLC25A47)	-2.184	4.84E-06
LACTB2	lactamase beta 2(LACTB2)	-2.185	1.37E-15
STEAP4	STEAP4 metalloreductase(STEAP4)	-2.186	8.27E-07
PKP2	plakophilin 2(PKP2)	-2.190	1.14E-09
KLF9	Kruppel like factor 9(KLF9)	-2.193	6.03E-08
SLC1A1	solute carrier family 1 member 1(SLC1A1)	-2.202	7.81E-10
KDSR	3-ketodihydrosphingosine reductase(KDSR)	-2.207	6.26E-11
HOMER2	homer scaffold protein 2(HOMER2)	-2.211	3.58E-10
LOC103177642	Locus 103177642	-2.212	2.25E-19
CYP2C49	cytochrome P450 2C49	-2.218	2.48E-11
THEM6	thioesterase superfamily member 6(THEM6)	-2.219	1.16E-08

GPT	glutamic--pyruvic transaminase(GPT)	-2.223	5.93E-07
AKR1C3	aldo-keto reductase family 1 member C3(AKR1C3)	-2.227	1.52E-12
KCNJ11	potassium inwardly rectifying channel subfamily J member 11(KCNJ11)	-2.239	4.40E-20
CCR10	C-C motif chemokine receptor 10(CCR10)	-2.270	7.37E-07
AQP11	aquaporin 11(AQP11)	-2.284	3.66E-08
ABHD3	abhydrolase domain containing 3, phospholipase(ABHD3)	-2.289	4.21E-16
KLC4	kinesin light chain 4(KLC4)	-2.298	3.13E-21
TUBB4A	tubulin beta 4A class IVa(TUBB4A)	-2.301	1.51E-10
RAB31	RAB31, member RAS oncogene family(RAB31)	-2.303	1.57E-08
CYP2A23	cytochrome P450 2A23	-2.317	1.35E-07
STAT3	signal transducer and activator of transcription 3(STAT3)	-2.318	2.45E-14
PITPNM3	PITPNM family member 3(PITPNM3)	-2.322	3.16E-15
SLC16A6	solute carrier family 16 member 6(SLC16A6)	-2.331	1.65E-12
ABCB11	ATP binding cassette subfamily B member 11(ABCB11)	-2.338	4.25E-08
GSKIP	GSK3B interacting protein(GSKIP)	-2.340	4.10E-14
SEBOX	SEBOX homeobox(SEBOX)	-2.341	3.64E-06
SAHAI-12	HLA class I histocompatibility antigen, A-69 alpha chain-like	-2.342	6.30E-08
TDO2	tryptophan 2,3-dioxygenase(TDO2)	-2.356	2.12E-14
ATP10A	ATPase phospholipid transporting 10A (putative)(ATP10A)	-2.357	1.97E-14
SLC51A	solute carrier family 51 subunit alpha(SLC51A)	-2.408	9.66E-07
DHRS1	dehydrogenase/reductase 1(DHRS1)	-2.428	5.52E-06
ABHD6	abhydrolase domain containing 6, acylglycerol lipase(ABHD6)	-2.431	2.42E-11
SLC16A5	solute carrier family 16 member 5(SLC16A5)	-2.443	3.46E-13
KLHL34	kelch like family member 34(KLHL34)	-2.445	2.05E-07
KIF1A	kinesin family member 1A(KIF1A)	-2.448	1.29E-06
PCK1	phosphoenolpyruvate carboxykinase 1(PCK1)	-2.455	3.08E-07
TMEM45A	transmembrane protein 45A(TMEM45A)	-2.463	1.15E-13
MGLL	monoglyceride lipase(MGLL)	-2.465	5.08E-14
CYP2E1	cytochrome P450 family 2 subfamily E member 1(CYP2E1)	-2.473	4.14E-06
SAHAI-01	class I histocompatibility antigen, Gogo-OKO alpha chain-like	-2.486	3.16E-08
B2M	beta-2-microglobulin(B2M)	-2.494	6.57E-13
TMEM229B	transmembrane protein 229B(TMEM229B)	-2.501	1.35E-06
SLC1A2	solute carrier family 1 member 2(SLC1A2)	-2.501	4.97E-07
SLC22A18	solute carrier family 22 member 18(SLC22A18)	-2.511	1.11E-13
MSMO1	methylsterol monooxygenase 1(MSMO1)	-2.517	8.64E-10
MCUB	mitochondrial calcium uniporter dominant negative subunit beta(MCUB)	-2.538	1.85E-15
LRIG3	leucine rich repeats and immunoglobulin like domains 3(LRIG3)	-2.541	1.37E-11
RCAN1	regulator of calcineurin 1(RCAN1)	-2.545	3.75E-14
DDC	dopa decarboxylase(DDC)	-2.546	3.26E-11
MODO-UE	MHC class I antigen	-2.561	2.20E-07
DDAH1	dimethylarginine dimethylaminohydrolase 1(DDAH1)	-2.565	2.47E-18
ADRA1A	adrenoceptor alpha 1A(ADRA1A)	-2.571	1.23E-09
NOTUM	notum, palmitoleoyl-protein carboxylesterase(NOTUM)	-2.587	3.86E-11
CFHR2	complement factor H related 2	-2.592	1.75E-20
TLR9	toll like receptor 9	-2.604	4.01E-09
SLC46A3	solute carrier family 46 member 3(SLC46A3)	-2.618	2.90E-23
CYP3A7-CYP3A51P	cytochrome P450 family 3 subfamily A member 7	-2.646	3.48E-14
LYSMD2	LysM domain containing 2(LYSMD2)	-2.656	3.94E-12
DHRS9	dehydrogenase/reductase 9(DHRS9)	-2.663	4.05E-07
RDH8	retinol dehydrogenase 8(RDH8)	-2.672	1.62E-15
SEL1L3	SEL1L family member 3(SEL1L3)	-2.677	4.27E-10
SLC13A2	solute carrier family 13 member 2(SLC13A2)	-2.684	4.89E-06
NRG4	neuregulin 4(NRG4)	-2.686	7.08E-14
TBC1D20.2.L	TBC1 domain family, member 20, gene 2 L homeolog	-2.697	3.32E-12
APOA5	apolipoprotein A5(APOA5)	-2.699	2.12E-14
FAM189A2	family with sequence similarity 189 member A2(FAM189A2)	-2.709	2.53E-16
KCNC3	potassium voltage-gated channel subfamily C member 3(KCNC3)	-2.748	2.36E-06

CA9	carbonic anhydrase 9(CA9)	-2.752	4.04E-14
CPT1B	carnitine palmitoyltransferase 1B(CPT1B)	-2.788	2.67E-11
SLC22A31	solute carrier family 22 member 31(SLC22A31)	-2.790	3.41E-06
PSMB9	proteasome 20S subunit beta 9(PSMB9)	-2.796	2.04E-10
SALL1	spalt like transcription factor 1(SALL1)	-2.798	6.17E-11
ANGPTL4	angiopoietin like 4(ANGPTL4)	-2.812	3.23E-08
CNGA3	cyclic nucleotide gated channel subunit alpha 3(CNGA3)	-2.812	9.06E-07
LOC100145789	Locus 100145789	-2.817	1.10E-16
SERINC2	serine incorporator 2(SERINC2)	-2.818	4.46E-13
TAFA3	TAFA chemokine like family member 3(TAFA3)	-2.840	1.79E-06
SLCO1C1	solute carrier organic anion transporter family member 1C1(SLCO1C1)	-2.890	2.70E-07
C2H2ORF88	chromosome 2 C2orf88 homolog	-2.894	2.52E-06
DIO2	iodothyronine deiodinase 2(DIO2)	-2.898	9.45E-10
TNFRSF6B	TNF receptor superfamily member 6b	-2.899	3.52E-08
LURAP1	leucine rich adaptor protein 1(LURAP1)	-2.906	2.23E-11
SULT5A1	sulfotransferase family 5A, member 1	-2.926	9.40E-10
STK32C	serine/threonine kinase 32C(STK32C)	-2.968	6.35E-10
FIBCD1	fibrinogen C domain containing 1(FIBCD1)	-3.008	5.25E-07
STS	steroid sulfatase(STS)	-3.027	9.14E-12
NGEF	neuronal guanine nucleotide exchange factor(NGEF)	-3.054	1.03E-07
MROH2B	maestro heat like repeat family member 2B(MROH2B)	-3.080	7.46E-08
NPC1L1	NPC1 like intracellular cholesterol transporter 1(NPC1L1)	-3.102	3.71E-19
FZD10	frizzled class receptor 10(FZD10)	-3.107	6.29E-24
ACADS	acyl-CoA dehydrogenase short chain(ACADS)	-3.113	1.54E-21
PLA2R1	phospholipase A2 receptor 1(PLA2R1)	-3.118	8.62E-16
TINAG	tubulointerstitial nephritis antigen(TINAG)	-3.128	4.80E-11
DECR2	2,4-dienoyl-CoA reductase 2(DECR2)	-3.130	3.92E-21
TUBA8	tubulin alpha 8(TUBA8)	-3.147	2.35E-14
SFTPA2	surfactant protein A2(SFTPA2)	-3.189	6.02E-13
PTGR2	prostaglandin reductase 2(PTGR2)	-3.205	2.89E-20
NQO1	NAD(P)H quinone dehydrogenase 1(NQO1)	-3.209	1.22E-20
CST6	cystatin E/M(CST6)	-3.214	1.61E-07
KLHDC7A	kelch domain containing 7A(KLHDC7A)	-3.226	1.84E-13
GRTP1	growth hormone regulated TBC protein 1(GRTP1)	-3.232	7.69E-08
GGT1	gamma-glutamyltransferase 1(GGT1)	-3.233	2.05E-12
ACY3	aminoacylase 3(ACY3)	-3.234	4.54E-18
DYNLL1	dynein light chain LC8-type 1(DYNLL1)	-3.262	3.73E-08
SEC14L3	SEC14 like lipid binding 3(SEC14L3)	-3.276	3.56E-10
CYP8B1	cytochrome P450 family 8 subfamily B member 1(CYP8B1)	-3.280	4.74E-28
FMO3	flavin containing dimethylaniline monoxygenase 3(FMO3)	-3.315	9.20E-11
TAP1	transporter 1, ATP binding cassette subfamily B member(TAP1)	-3.342	4.47E-14
AADAC	arylacetamide deacetylase(AADAC)	-3.354	7.88E-16
CBLC	Cbl proto-oncogene C(CBLC)	-3.363	1.67E-12
MB	myoglobin(MB)	-3.364	1.51E-10
MSRB3	methionine sulfoxide reductase B3(MSRB3)	-3.372	1.86E-12
SCUBE2	signal peptide, CUB domain and EGF like domain containing 2(SCUBE2)	-3.420	3.23E-14
CYP26A1	cytochrome P450 family 26 subfamily A member 1(CYP26A1)	-3.422	8.22E-12
ALDH1	Aldehyde dehydrogenase, cytosolic	-3.434	4.25E-32
CRYL1	crystallin lambda 1(CRYL1)	-3.444	2.47E-18
NNMT	nicotinamide N-methyltransferase(NNMT)	-3.505	4.08E-17
PID1	phosphotyrosine interaction domain containing 1(PID1)	-3.550	3.74E-25
CHIC1	cysteine rich hydrophobic domain 1(CHIC1)	-3.633	1.62E-12
TREH	trehalase(TREH)	-3.651	4.37E-13
PPM1J	protein phosphatase, Mg2+/Mn2+ dependent 1J(PPM1J)	-3.672	2.31E-14
FCGBP	Fc gamma binding protein(FCGBP)	-3.678	1.24E-09
NIPAL1	NIPA like domain containing 1(NIPAL1)	-3.844	2.00E-16
ISG20	interferon stimulated exonuclease gene 20(ISG20)	-3.920	4.89E-24
SPON1	spondin 1(SPON1)	-3.928	2.84E-17
LOC106108060	Locus 106108060	-3.945	9.59E-17

SLC41A2	solute carrier family 41 member 2(SLC41A2)	-3.953	2.93E-19
UNC5A	unc-5 netrin receptor A(UNC5A)	-4.004	2.04E-16
ZGC:103559	Allantoinase, mitochondrial	-4.182	2.76E-20
CBLN3	cerebellin 3 precursor(CBLN3)	-4.189	2.32E-16
LOC396781	Locus 396781	-4.308	3.99E-07
CABP7	calcium binding protein 7(CABP7)	-4.424	2.38E-13
DAO	D-amino acid oxidase(DAO)	-5.036	5.57E-37
GABBR2	gamma-aminobutyric acid type B receptor subunit 2(GABBR2)	-5.226	2.88E-21
CYP2C33	cytochrome P450 family 2 subfamily C member 33	-5.331	4.02E-33
NMRAL1	NmrA like redox sensor 1(NMRAL1)	-5.508	6.56E-39
CP	ceruloplasmin(CP)	-5.659	1.09E-36
MGC152281	uncharacterized LOC507942	-5.915	7.66E-32
SRD5A2	steroid 5 alpha-reductase 2(SRD5A2)	-6.394	1.06E-44
LOC511498	Locus 511498	-6.688	1.35E-60
GBA3	glucosylceramidase beta 3 (gene/pseudogene)	-6.907	2.37E-32

Appendix C3 Table 2: Gene ontology (GO) enrichment analysis result on **upregulated** differentially expressed genes between adults and juveniles identified using **DESeq2**. The analysis approximates the true distribution of numbers of members of a category amongst differentially expressed genes by the Wallenius non-central hypergeometric distribution. False Discovery Rate (FDR) is then calculated using Benjamini and Hochberg method.

GO category	Number of genes	Number in category	Term	Ontology	P-value	FDR
GO:0030312	33	239	external encapsulating structure	CC	6.81E-15	6.37E-11
GO:0031012	33	239	extracellular matrix	CC	6.81E-15	6.37E-11
GO:0005576	65	909	extracellular region	CC	1.15E-14	7.18E-11
GO:0071944	114	2604	cell periphery	CC	1.26E-09	5.91E-06
GO:0005615	43	623	extracellular space	CC	1.58E-09	5.91E-06
GO:0005581	11	36	collagen trimer	CC	1.90E-09	5.92E-06
GO:0030198	20	159	extracellular matrix organization	BP	7.70E-09	1.79E-05
GO:0045229	20	159	external encapsulating structure organization	BP	7.70E-09	1.79E-05
GO:0043062	20	160	extracellular structure organization	BP	8.63E-09	1.79E-05
GO:1903047	31	379	mitotic cell cycle process	BP	2.01E-08	3.77E-05
GO:0000819	16	111	sister chromatid segregation	BP	3.54E-08	6.01E-05
GO:0000070	14	89	mitotic sister chromatid segregation	BP	8.73E-08	0.000136
GO:0032501	125	3189	multicellular organismal process	BP	9.92E-08	0.0001427
GO:0062023	17	132	collagen-containing extracellular matrix	CC	1.29E-07	0.00017276
GO:0048856	109	2662	anatomical structure development	BP	1.40E-07	0.00017421
GO:0000278	33	469	mitotic cell cycle	BP	2.46E-07	0.00028698
GO:0007275	99	2398	multicellular organism development	BP	4.94E-07	0.00054291
GO:0140014	17	157	mitotic nuclear division	BP	8.12E-07	0.00074457
GO:0098644	6	13	complex of collagen trimers	CC	8.19E-07	0.00074457
GO:0098813	17	157	nuclear chromosome segregation	BP	8.21E-07	0.00074457
GO:0030199	8	30	collagen fibril organization	BP	8.36E-07	0.00074457
GO:0048731	92	2187	system development	BP	1.04E-06	0.00088468
GO:0007059	18	179	chromosome segregation	BP	1.13E-06	0.00091635
GO:0051301	21	240	cell division	BP	1.34E-06	0.0010413
GO:0048513	70	1537	animal organ development	BP	1.75E-06	0.00131009
GO:0022402	38	641	cell cycle process	BP	2.08E-06	0.00149825
GO:0032502	111	2858	developmental process	BP	2.19E-06	0.0015156
GO:0099080	36	586	supramolecular complex	CC	2.40E-06	0.00160473
GO:0005583	5	10	fibrillar collagen trimer	CC	4.04E-06	0.00251655
GO:0098643	5	10	banded collagen fibril	CC	4.04E-06	0.00251655
GO:0099081	28	417	supramolecular polymer	CC	7.28E-06	0.0043907
GO:0000280	19	230	nuclear division	BP	1.02E-05	0.00596084
GO:0007049	43	828	cell cycle	BP	1.22E-05	0.00689235

GO:0005201	8	41	extracellular matrix structural constituent	MF	1.59E-05	0.00873618
GO:0099512	27	414	supramolecular fiber	CC	1.77E-05	0.00944579
GO:0051233	6	26	spindle midzone	CC	4.71E-05	0.02448127
GO:0022610	36	670	biological adhesion	BP	4.89E-05	0.0247172
GO:0003688	5	17	DNA replication origin binding	MF	5.34E-05	0.02628734
GO:0032133	3	4	chromosome passenger complex	CC	6.22E-05	0.02982573
GO:0000775	11	100	chromosome, centromeric region	CC	6.77E-05	0.03165825
GO:0048407	4	9	platelet-derived growth factor binding	MF	7.15E-05	0.03258987
GO:0009653	58	1320	anatomical structure morphogenesis	BP	7.44E-05	0.03312373
GO:0048285	19	266	organelle fission	BP	7.67E-05	0.03334774
GO:0008237	12	123	metallopeptidase activity	MF	8.92E-05	0.03790504
GO:0007155	35	666	cell adhesion	BP	9.80E-05	0.04071322

Appendix C3 Table 3: Gene ontology (GO) enrichment analysis result on **downregulated** differentially expressed genes between adults and juveniles identified using **DESeq2**. The analysis approximates the true distribution of numbers of members of a category amongst differentially expressed genes by the Wallenius non-central hypergeometric distribution. False Discovery Rate (FDR) is then calculated using Benjamini and Hochberg method.

GO category	Number of genes	Number in category	Term	Ontology	P-value	FDR
GO:0016491	24	493	oxidoreductase activity	MF	4.94E-09	9.24E-05
GO:1901615	15	239	organic hydroxy compound metabolic process	BP	1.73E-07	0.00162064
GO:0044242	11	125	cellular lipid catabolic process	BP	2.77E-07	0.00172732
GO:0019752	20	459	carboxylic acid metabolic process	BP	5.97E-07	0.00232332
GO:0008028	7	43	monocarboxylic acid transmembrane transporter activity	MF	6.64E-07	0.00232332
GO:0043436	20	466	oxoacid metabolic process	BP	7.46E-07	0.00232332
GO:0006082	20	479	organic acid metabolic process	BP	1.16E-06	0.00266738
GO:0032787	15	279	monocarboxylic acid metabolic process	BP	1.27E-06	0.00266738
GO:0044281	29	923	small molecule metabolic process	BP	1.28E-06	0.00266738
GO:0044282	12	179	small molecule catabolic process	BP	1.62E-06	0.00291454
GO:0006631	12	181	fatty acid metabolic process	BP	1.71E-06	0.00291454
GO:0006629	24	705	lipid metabolic process	BP	2.91E-06	0.00453042
GO:0044255	20	548	cellular lipid metabolic process	BP	7.86E-06	0.01130177
GO:0016042	11	177	lipid catabolic process	BP	8.52E-06	0.01137337
GO:0015718	7	66	monocarboxylic acid transport	BP	1.18E-05	0.01467123
GO:0046943	8	93	carboxylic acid transmembrane transporter activity	MF	1.35E-05	0.01572416
GO:0005342	8	94	organic acid transmembrane transporter activity	MF	1.44E-05	0.01588416
GO:1901617	9	123	organic hydroxy compound biosynthetic process	BP	1.66E-05	0.01665888
GO:0046395	9	123	carboxylic acid catabolic process	BP	1.69E-05	0.01665888
GO:0008202	9	126	steroid metabolic process	BP	2.13E-05	0.01960994
GO:0016054	9	127	organic acid catabolic process	BP	2.20E-05	0.01960994
GO:0004497	6	55	monooxygenase activity	MF	4.96E-05	0.04216214
GO:0016628	4	18	oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor	MF	5.66E-05	0.04603712

Appendix C3 Table 4: Gene ontology (GO) enrichment analysis result on **upregulated** differentially expressed genes between adults and juveniles identified using **Limma**. The analysis approximates the true distribution of numbers of members of a category amongst differentially expressed genes by the Wallenius non-central hypergeometric distribution. False Discovery Rate (FDR) is then calculated using Benjamini and Hochberg method.

GO category	Number of genes	Number in category	Term	Ontology	P-value	FDR
GO:0005576	103	980	extracellular region	CC	5.79E-21	1.09E-16
GO:0030312	44	244	external encapsulating structure	CC	1.99E-16	1.25E-12
GO:0031012	44	244	extracellular matrix	CC	1.99E-16	1.25E-12
GO:0071944	189	2762	cell periphery	CC	2.09E-14	9.86E-11
GO:0005615	64	673	extracellular space	CC	1.99E-11	7.52E-08
GO:1903047	44	385	mitotic cell cycle process	BP	7.66E-10	2.41E-06
GO:0000819	22	114	sister chromatid segregation	BP	1.97E-09	5.30E-06
GO:0007059	28	185	chromosome segregation	BP	3.67E-09	8.66E-06
GO:0000070	19	90	mitotic sister chromatid segregation	BP	6.05E-09	1.27E-05
GO:0098813	25	160	nuclear chromosome segregation	BP	1.44E-08	2.71E-05
GO:0022402	58	657	cell cycle process	BP	2.47E-08	3.97E-05
GO:0000278	47	480	mitotic cell cycle	BP	2.68E-08	3.97E-05
GO:0005581	12	37	collagen trimer	CC	2.73E-08	3.97E-05
GO:0051301	31	247	cell division	BP	3.65E-08	4.92E-05
GO:0007049	68	852	cell cycle	BP	6.58E-08	8.28E-05
GO:0062023	22	136	collagen-containing extracellular matrix	CC	9.31E-08	0.00010979
GO:0030198	24	164	extracellular matrix organization	BP	1.06E-07	0.00011111
GO:0045229	24	164	external encapsulating structure organization	BP	1.06E-07	0.00011111
GO:0043062	24	165	extracellular structure organization	BP	1.20E-07	0.00011877
GO:0048856	167	2854	anatomical structure development	BP	1.67E-07	0.00015792
GO:0140014	23	160	mitotic nuclear division	BP	2.25E-07	0.00020165
GO:0032502	174	3061	developmental process	BP	5.43E-07	0.00046515
GO:0009888	72	968	tissue development	BP	6.90E-07	0.00056593
GO:0032501	189	3425	multicellular organismal process	BP	1.28E-06	0.00100422
GO:0007275	150	2574	multicellular organism development	BP	2.45E-06	0.00185067
GO:0030199	9	30	collagen fibril organization	BP	3.01E-06	0.0021848
GO:0005886	145	2504	plasma membrane	CC	6.23E-06	0.00421784
GO:0000775	16	103	chromosome, centromeric region	CC	6.26E-06	0.00421784
GO:0000280	26	238	nuclear division	BP	7.39E-06	0.00480666
GO:0045839	8	28	negative regulation of mitotic nuclear division	BP	1.35E-05	0.00849346
GO:0048731	136	2346	system development	BP	1.52E-05	0.00926667
GO:1902850	13	79	microtubule cytoskeleton organization involved in mitosis	BP	2.50E-05	0.01475437
GO:0005583	5	10	fibrillar collagen trimer	CC	3.75E-05	0.0207898
GO:0098643	5	10	banded collagen fibril	CC	3.75E-05	0.0207898
GO:0006270	7	24	DNA replication initiation	BP	4.40E-05	0.02370706
GO:0030261	6	18	chromosome condensation	BP	4.96E-05	0.0259983
GO:0048513	100	1656	animal organ development	BP	5.15E-05	0.02624454
GO:0009653	89	1418	anatomical structure morphogenesis	BP	5.29E-05	0.02626758
GO:0007052	11	63	mitotic spindle organization	BP	5.61E-05	0.0271511

GO:0031262	3	3	Ndc80 complex	CC	5.93E-05	0.02727999
GO:0031617	3	3	NMS complex	CC	5.93E-05	0.02727999
GO:0051784	8	34	negative regulation of nuclear division	BP	6.18E-05	0.02776284
GO:0008237	16	125	metallopeptidase activity	MF	6.84E-05	0.02946915
GO:0010564	30	336	regulation of cell cycle process	BP	6.88E-05	0.02946915
GO:0005509	36	441	calcium ion binding	MF	7.39E-05	0.03098659
GO:0048285	26	274	organelle fission	BP	8.49E-05	0.03463962
GO:0031226	53	740	intrinsic component of plasma membrane	CC	8.63E-05	0.03463962
GO:0005887	51	705	integral component of plasma membrane	CC	9.51E-05	0.03736668
GO:0051726	41	535	regulation of cell cycle	BP	0.00010516	0.04047441
GO:0000796	4	7	condensin complex	CC	0.00010947	0.04129012

Appendix C3 Table 5: Gene ontology (GO) enrichment analysis result on **downregulated** differentially expressed genes between adults and juveniles identified using **Limma**. The analysis approximates the true distribution of numbers of members of a category amongst differentially expressed genes by the Wallenius non-central hypergeometric distribution. False Discovery Rate (FDR) is then calculated using Benjamini and Hochberg method.

GO category	Number of genes	Number in category	Term	Ontology	P-value	FDR
GO:0006629	39	720	lipid metabolic process	BP	2.11E-11	2.70E-07
GO:0044281	45	945	small molecule metabolic process	BP	2.86E-11	2.70E-07
GO:0032787	23	284	monocarboxylic acid metabolic process	BP	1.86E-10	1.17E-06
GO:0019752	29	467	carboxylic acid metabolic process	BP	3.93E-10	1.60E-06
GO:0044255	32	557	cellular lipid metabolic process	BP	4.25E-10	1.60E-06
GO:0043436	29	474	oxoacid metabolic process	BP	5.63E-10	1.77E-06
GO:0006082	29	487	organic acid metabolic process	BP	1.05E-09	2.83E-06
GO:0008202	15	131	steroid metabolic process	BP	2.72E-09	6.41E-06
GO:1901615	20	250	organic hydroxy compound metabolic process	BP	3.97E-09	8.31E-06
GO:0016491	28	505	oxidoreductase activity	MF	4.61E-09	8.70E-06
GO:0044282	17	182	small molecule catabolic process	BP	6.99E-09	1.20E-05
GO:0006631	17	184	fatty acid metabolic process	BP	8.44E-09	1.33E-05
GO:0008610	21	331	lipid biosynthetic process	BP	1.05E-07	0.00015299
GO:0046395	13	124	carboxylic acid catabolic process	BP	1.22E-07	0.00016495
GO:0044242	13	127	cellular lipid catabolic process	BP	1.71E-07	0.00021061
GO:0016054	13	128	organic acid catabolic process	BP	1.79E-07	0.00021061
GO:1901617	12	127	organic hydroxy compound biosynthetic process	BP	1.04E-06	0.0010913
GO:0031224	86	3352	intrinsic component of membrane	CC	1.04E-06	0.0010913
GO:0006638	9	66	neutral lipid metabolic process	BP	1.27E-06	0.00116613
GO:0006639	9	66	acylglycerol metabolic process	BP	1.27E-06	0.00116613
GO:0006641	8	50	triglyceride metabolic process	BP	1.30E-06	0.00116613
GO:0016021	85	3303	integral component of membrane	CC	1.67E-06	0.00143052
GO:0016042	14	183	lipid catabolic process	BP	1.77E-06	0.00144754
GO:0019216	13	159	regulation of lipid metabolic process	BP	2.30E-06	0.00180764
GO:0006694	9	75	steroid biosynthetic process	BP	2.82E-06	0.00212582
GO:0004497	8	56	monooxygenase activity	MF	3.65E-06	0.00264629
GO:0016627	7	42	oxidoreductase activity, acting on the CH-CH group of donors	MF	4.67E-06	0.00326202
GO:0090207	5	18	regulation of triglyceride metabolic process	BP	5.53E-06	0.00372248
GO:0046890	9	81	regulation of lipid biosynthetic process	BP	6.03E-06	0.0039222
GO:0016126	6	30	sterol biosynthetic process	BP	6.63E-06	0.00416821
GO:0005215	32	832	transporter activity	MF	8.27E-06	0.00503119

GO:0072329	8	63	monocarboxylic acid catabolic process	BP	9.47E-06	0.00558413
GO:0016125	8	67	sterol metabolic process	BP	1.03E-05	0.00586203
GO:1901575	39	1156	organic substance catabolic process	BP	1.06E-05	0.00590554
GO:0010873	3	5	positive regulation of cholesterol esterification	BP	2.74E-05	0.01479078
GO:0005319	9	93	lipid transporter activity	MF	2.97E-05	0.01547297
GO:0006084	5	23	acetyl-CoA metabolic process	BP	3.12E-05	0.01547297
GO:0140042	3	5	lipid droplet formation	BP	3.12E-05	0.01547297
GO:0050660	7	56	flavin adenine dinucleotide binding	MF	4.27E-05	0.02013661
GO:0016705	9	99	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	MF	4.27E-05	0.02013661
GO:0042180	9	103	cellular ketone metabolic process	BP	4.45E-05	0.02049296
GO:0009056	42	1375	catabolic process	BP	5.41E-05	0.02430693
GO:0044248	38	1191	cellular catabolic process	BP	5.62E-05	0.02466963
GO:0010872	3	6	regulation of cholesterol esterification	BP	5.96E-05	0.02514217
GO:0009062	7	60	fatty acid catabolic process	BP	6.07E-05	0.02514217
GO:0016709	5	26	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen	MF	6.13E-05	0.02514217
GO:1902652	7	66	secondary alcohol metabolic process	BP	8.04E-05	0.03226096
GO:0019915	6	44	lipid storage	BP	8.40E-05	0.03299106
GO:0019218	6	47	regulation of steroid metabolic process	BP	9.70E-05	0.03733003
GO:0019432	4	16	triglyceride biosynthetic process	BP	0.00010581	0.03991083
GO:0046486	13	228	glycerolipid metabolic process	BP	0.00012097	0.04473624
GO:0062012	10	143	regulation of small molecule metabolic process	BP	0.00012898	0.0467805
GO:0010876	13	232	lipid localization	BP	0.00013312	0.04737107

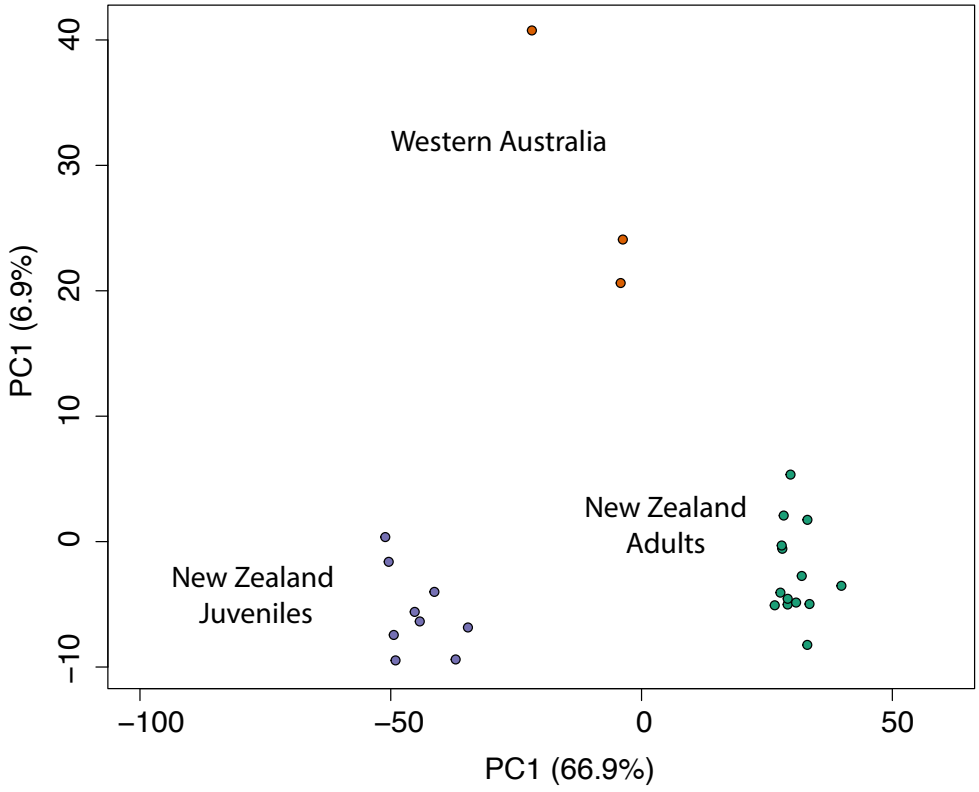
Appendix C3 Table 6: Gene ontology (GO) enrichment analysis result on **downregulated** differentially expressed genes between adults and juveniles identified using **WGCNA**. The analysis approximates the true distribution of numbers of members of a category amongst differentially expressed genes by the Wallenius non-central hypergeometric distribution. False Discovery Rate (FDR) is then calculated using Benjamini and Hochberg method.

GO category	Number of genes	Number in category	Term	Ontology	P-value	FDR
GO:0000278	107	469	mitotic cell cycle	BP	7.26E-12	6.79E-08
GO:0022402	132	641	cell cycle process	BP	4.81E-11	3.00E-07
GO:0007049	156	828	cell cycle	BP	6.55E-10	3.06E-06
GO:0000070	33	89	mitotic sister chromatid segregation	BP	1.34E-09	5.01E-06
GO:0000819	37	111	sister chromatid segregation	BP	4.09E-09	1.20E-05
GO:0140014	46	157	mitotic nuclear division	BP	4.48E-09	1.20E-05
GO:0098813	45	157	nuclear chromosome segregation	BP	2.02E-08	4.72E-05
GO:0007059	49	179	chromosome segregation	BP	2.31E-08	4.81E-05
GO:0005581	18	36	collagen trimer	CC	3.30E-08	6.18E-05
GO:0048285	62	266	organelle fission	BP	9.79E-08	0.00016643
GO:0051301	57	240	cell division	BP	1.17E-07	0.00018158
GO:0000280	55	230	nuclear division	BP	2.17E-07	0.00030197
GO:0006261	30	92	DNA-dependent DNA replication	BP	2.26E-07	0.00030197
GO:0005576	151	909	extracellular region	CC	2.88E-07	0.0003592
GO:0000775	31	100	chromosome, centromeric region	CC	3.86E-07	0.00045093
GO:0000796	7	7	condensin complex	CC	4.28E-07	0.0004708
GO:0007052	23	63	mitotic spindle organization	BP	4.86E-07	0.00050435
GO:1902850	26	79	microtubule cytoskeleton organization involved in mitosis	BP	8.93E-07	0.00087856
GO:0044786	13	26	cell cycle DNA replication	BP	3.37E-06	0.00314777
GO:0030312	53	239	external encapsulating structure	CC	6.42E-06	0.00545192

GO:0031012	53	239	extracellular matrix	CC	6.42E-06	0.00545192
GO:0006270	12	24	DNA replication initiation	BP	8.07E-06	0.00656377
GO:0007076	7	9	mitotic chromosome condensation	BP	1.24E-05	0.00955865
GO:0006260	38	153	DNA replication	BP	1.28E-05	0.00955865
GO:0007051	28	102	spindle organization	BP	1.70E-05	0.01221994
GO:0098687	38	159	chromosomal region	CC	2.08E-05	0.01441785
GO:0005819	45	203	spindle	CC	2.47E-05	0.01602005
GO:0005615	104	623	extracellular space	CC	2.54E-05	0.01602005
GO:0045839	12	27	negative regulation of mitotic nuclear division	BP	2.57E-05	0.01602005
GO:0006635	16	44	fatty acid beta-oxidation	BP	2.82E-05	0.01700882
GO:0005777	22	75	peroxisome	CC	3.29E-05	0.0183011
GO:0042579	22	75	microbody	CC	3.29E-05	0.0183011
GO:0033260	11	23	nuclear DNA replication	BP	3.33E-05	0.0183011
GO:0030261	9	17	chromosome condensation	BP	3.77E-05	0.02011583
GO:0098644	8	13	complex of collagen trimers	CC	4.01E-05	0.02082386
GO:0000217	9	17	DNA secondary structure binding	MF	5.93E-05	0.02994784
GO:0003688	9	17	DNA replication origin binding	MF	6.55E-05	0.03223995
GO:0061134	23	88	peptidase regulator activity	MF	7.21E-05	0.03456014
GO:0006267	6	8	pre-replicative complex assembly involved in nuclear cell cycle DNA replication	BP	8.17E-05	0.03635284
GO:0036388	6	8	pre-replicative complex assembly	BP	8.17E-05	0.03635284
GO:1902299	6	8	pre-replicative complex assembly involved in cell cycle DNA replication	BP	8.17E-05	0.03635284
GO:0033046	10	22	negative regulation of sister chromatid segregation	BP	0.00010023	0.03933457
GO:0033048	10	22	negative regulation of mitotic sister chromatid segregation	BP	0.00010023	0.03933457
GO:0045841	10	22	negative regulation of mitotic metaphase/anaphase transition	BP	0.00010023	0.03933457
GO:2000816	10	22	negative regulation of mitotic sister chromatid separation	BP	0.00010023	0.03933457
GO:0003777	16	47	microtubule motor activity	MF	0.00010148	0.03933457
GO:0019395	18	58	fatty acid oxidation	BP	0.00010293	0.03933457
GO:0033047	11	26	regulation of mitotic sister chromatid segregation	BP	0.00010472	0.03933457
GO:0032465	17	53	regulation of cytokinesis	BP	0.00010718	0.03933457
GO:0090329	11	26	regulation of DNA-dependent DNA replication	BP	0.00010729	0.03933457
GO:0000910	24	90	cytokinesis	BP	0.0001233	0.04418743
GO:0044770	47	232	cell cycle phase transition	BP	0.00012596	0.04418743
GO:0044772	43	207	mitotic cell cycle phase transition	BP	0.00012762	0.04418743
GO:0061952	5	6	midbody abscission	BP	0.00013321	0.04512536
GO:0030198	36	159	extracellular matrix organization	BP	0.00013757	0.04512536
GO:0045229	36	159	external encapsulating structure organization	BP	0.00013757	0.04512536
GO:0010032	5	6	meiotic chromosome condensation	BP	0.0001516	0.04756997
GO:0000922	22	80	spindle pole	CC	0.00015623	0.04756997
GO:0043062	36	160	extracellular structure organization	BP	0.00015769	0.04756997
GO:0051985	10	23	negative regulation of chromosome segregation	BP	0.00016029	0.04756997
GO:1902100	10	23	negative regulation of metaphase/anaphase transition of cell cycle	BP	0.00016029	0.04756997
GO:1905819	10	23	negative regulation of chromosome separation	BP	0.00016029	0.04756997
GO:0005871	13	35	kinesin complex	CC	0.00016322	0.04768237

Chapter 4

Appendix C4 Figure 1: Principal component analysis of the DeSeq2 normalized counts on the 475 developmental genes comparing the Western Australian samples to the New Zealand juvenile and adult samples.



Appendix C4 Table 1: Gene ontology (GO) enrichment analysis result on **downregulated** differentially expressed genes between New Zealand and western Australian brushtail possums identified using **DESeq2**. The analysis approximates the true distribution of numbers of members of a category amongst differentially expressed genes by the Wallenius non-central hypergeometric distribution. False Discovery Rate (FDR) is then calculated using Benjamini and Hochberg method.

GO category	Number of genes	Number in category	Term	Ontology	P-value
GO:0016477	98	696	cell migration	BP	1.56E-17
GO:0048870	100	761	cell motility	BP	4.76E-16
GO:0051674	100	761	localization of cell	BP	4.76E-16
GO:0040011	106	840	locomotion	BP	5.24E-16
GO:0048856	216	2518	anatomical structure development	BP	8.28E-15
GO:0032502	226	2710	developmental process	BP	2.42E-14
GO:0005886	195	2221	plasma membrane	CC	7.15E-14
GO:0006928	110	973	movement of cell or subcellular component	BP	1.82E-13
GO:0048646	74	522	anatomical structure formation involved in morphogenesis	BP	2.98E-13
GO:0032501	240	3015	multicellular organismal process	BP	3.22E-13
GO:0009653	127	1249	anatomical structure morphogenesis	BP	1.72E-12
GO:0001525	44	228	angiogenesis	BP	8.78E-12
GO:0035239	60	399	tube morphogenesis	BP	1.12E-11
GO:0007275	190	2266	multicellular organism development	BP	1.12E-11
GO:0001944	54	333	vasculature development	BP	1.12E-11
GO:0072359	68	495	circulatory system development	BP	1.43E-11
GO:0001568	52	316	blood vessel development	BP	1.74E-11
GO:0050896	279	3830	response to stimulus	BP	2.47E-11
GO:0048731	175	2064	system development	BP	5.76E-11
GO:0048514	47	277	blood vessel morphogenesis	BP	9.11E-11
GO:0030334	61	435	regulation of cell migration	BP	1.25E-10

GO:0040012	63	475	regulation of locomotion	BP	6.07E-10
GO:2000145	61	457	regulation of cell motility	BP	1.03E-09
GO:0050793	107	1075	regulation of developmental process	BP	1.06E-09
GO:0001667	39	216	ameboidal-type cell migration	BP	1.40E-09
GO:0005576	90	830	extracellular region	CC	1.57E-09
GO:0051270	62	477	regulation of cellular component movement	BP	1.84E-09
GO:0035295	62	482	tube development	BP	2.50E-09
GO:0048869	151	1783	cellular developmental process	BP	4.61E-09
GO:0030154	148	1741	cell differentiation	BP	5.61E-09
GO:0030198	30	141	extracellular matrix organization	BP	5.61E-09
GO:0045229	30	141	external encapsulating structure organization	BP	5.61E-09
GO:0043062	30	142	extracellular structure organization	BP	6.55E-09
GO:0030312	36	203	external encapsulating structure	CC	9.27E-09
GO:0031012	36	203	extracellular matrix	CC	9.27E-09
GO:0007155	72	631	cell adhesion	BP	9.77E-09
GO:0022610	72	633	biological adhesion	BP	1.10E-08
GO:2000026	69	591	regulation of multicellular organismal development	BP	1.10E-08
GO:0009888	87	848	tissue development	BP	2.18E-08
GO:0061061	42	272	muscle structure development	BP	2.18E-08
GO:0048513	127	1447	animal organ development	BP	2.52E-08
GO:0023052	205	2758	signaling	BP	4.78E-08
GO:0007154	206	2777	cell communication	BP	4.94E-08
GO:0022603	52	398	regulation of anatomical structure morphogenesis	BP	5.49E-08
GO:0042221	137	1620	response to chemical	BP	5.49E-08
GO:0051239	110	1200	regulation of multicellular organismal process	BP	5.49E-08
GO:0006935	42	282	chemotaxis	BP	6.70E-08
GO:0007165	190	2512	signal transduction	BP	7.00E-08
GO:0042330	42	283	taxis	BP	7.20E-08
GO:0005615	65	566	extracellular space	CC	1.22E-07
GO:0051716	228	3267	cellular response to stimulus	BP	9.28E-07
GO:0010631	28	154	epithelial cell migration	BP	1.01E-06
GO:0090132	28	155	epithelium migration	BP	1.15E-06
GO:0090130	28	157	tissue migration	BP	1.52E-06
GO:0062023	23	114	collagen-containing extracellular matrix	CC	2.13E-06
GO:0045765	23	114	regulation of angiogenesis	BP	3.04E-06
GO:0051241	51	435	negative regulation of multicellular organismal process	BP	3.53E-06
GO:0009605	92	1017	response to external stimulus	BP	3.58E-06
GO:1901342	23	116	regulation of vasculature development	BP	4.11E-06
GO:0050789	340	5455	regulation of biological process	BP	4.11E-06
GO:0035556	105	1228	intracellular signal transduction	BP	4.11E-06
GO:0007166	102	1186	cell surface receptor signaling pathway	BP	5.22E-06
GO:0070887	105	1239	cellular response to chemical stimulus	BP	7.49E-06
GO:0050794	323	5154	regulation of cellular process	BP	8.25E-06
GO:0065007	356	5817	biological regulation	BP	1.02E-05
GO:0040013	26	153	negative regulation of locomotion	BP	1.15E-05
GO:0048583	134	1726	regulation of response to stimulus	BP	1.30E-05
GO:0048523	159	2154	negative regulation of cellular process	BP	1.43E-05
GO:0043542	22	115	endothelial cell migration	BP	1.56E-05
GO:0043534	16	63	blood vessel endothelial cell migration	BP	1.57E-05
GO:0045595	66	672	regulation of cell differentiation	BP	1.68E-05
GO:0065008	136	1774	regulation of biological quality	BP	1.87E-05
GO:0045766	16	64	positive regulation of angiogenesis	BP	1.87E-05
GO:1904018	16	64	positive regulation of vasculature development	BP	1.87E-05
GO:0008283	75	809	cell population proliferation	BP	2.38E-05
GO:0051179	206	3004	localization	BP	2.38E-05
GO:0031226	63	639	intrinsic component of plasma membrane	CC	2.77E-05
GO:0043535	14	52	regulation of blood vessel endothelial cell migration	BP	4.30E-05
GO:0030155	41	343	regulation of cell adhesion	BP	5.15E-05
GO:0042127	65	680	regulation of cell population proliferation	BP	5.69E-05
GO:0002376	85	982	immune system process	BP	7.02E-05
GO:0030335	33	251	positive regulation of cell migration	BP	9.54E-05
GO:0010596	11	34	negative regulation of endothelial cell migration	BP	0.00011187
GO:0040017	34	265	positive regulation of locomotion	BP	0.00011187
GO:0071363	36	292	cellular response to growth factor stimulus	BP	0.00012228
GO:0051094	56	566	positive regulation of developmental process	BP	0.00016078
GO:0010632	21	123	regulation of epithelial cell migration	BP	0.00017795
GO:0043537	9	23	negative regulation of blood vessel endothelial cell migration	BP	0.00017959
GO:0005887	59	612	integral component of plasma membrane	CC	0.00017959
GO:0010633	12	43	negative regulation of epithelial cell migration	BP	0.00018837
GO:0070848	36	298	response to growth factor	BP	0.00018837

GO:2000147	33	260	positive regulation of cell motility	BP	0.00019393
GO:0051272	33	262	positive regulation of cellular component movement	BP	0.00022909
GO:0051240	60	631	positive regulation of multicellular organismal process	BP	0.00023915
GO:0030336	21	129	negative regulation of cell migration	BP	0.000334
GO:0048519	166	2379	negative regulation of biological process	BP	0.00037693
GO:0042692	24	163	muscle cell differentiation	BP	0.00039294
GO:0000165	39	346	MAPK cascade	BP	0.00041807
GO:0048018	28	205	receptor ligand activity	MF	0.00042275
GO:0048518	183	2688	positive regulation of biological process	BP	0.00043945
GO:0065009	92	1136	regulation of molecular function	BP	0.00044686
GO:0030546	28	207	signaling receptor activator activity	MF	0.00049642
GO:0010594	17	92	regulation of endothelial cell migration	BP	0.00055104
GO:0051093	42	392	negative regulation of developmental process	BP	0.00057854
GO:0032879	96	1215	regulation of localization	BP	0.00067074
GO:1902531	67	758	regulation of intracellular signal transduction	BP	0.00067628
GO:0060326	22	144	cell chemotaxis	BP	0.00068176
GO:0030029	44	425	actin filament-based process	BP	0.00068437
GO:2000146	21	136	negative regulation of cell motility	BP	0.00070849
GO:0051271	21	138	negative regulation of cellular component movement	BP	0.00089457
GO:0060429	49	500	epithelium development	BP	0.000926
GO:0048872	20	127	homeostasis of number of cells	BP	0.000926
GO:0030545	28	215	receptor regulator activity	MF	0.00094255
GO:0005102	62	688	signaling receptor binding	MF	0.00103114
GO:0090066	32	271	regulation of anatomical structure size	BP	0.00111158
GO:0009966	103	1344	regulation of signal transduction	BP	0.00117446
GO:0045785	26	198	positive regulation of cell adhesion	BP	0.00125152
GO:0002682	52	544	regulation of immune system process	BP	0.00125267
GO:0098772	94	1196	molecular function regulator	MF	0.00126914
GO:0009986	34	299	cell surface	CC	0.00129965
GO:0000902	47	485	cell morphogenesis	BP	0.00150825
GO:0098609	39	371	cell-cell adhesion	BP	0.00152788
GO:0048468	75	910	cell development	BP	0.00173933
GO:0030036	40	388	actin cytoskeleton organization	BP	0.00188452
GO:0030215	7	18	semaphorin receptor binding	MF	0.00207726
GO:0043408	35	320	regulation of MAPK cascade	BP	0.00213631
GO:0001775	41	401	cell activation	BP	0.0022095
GO:0048585	66	775	negative regulation of response to stimulus	BP	0.00256007
GO:0050900	22	158	leukocyte migration	BP	0.00257737
GO:0005515	224	3515	protein binding	MF	0.00257737
GO:0044093	58	654	positive regulation of molecular function	BP	0.00257808
GO:0019838	14	76	growth factor binding	MF	0.00292249
GO:0045597	38	368	positive regulation of cell differentiation	BP	0.00305089
GO:0071310	78	971	cellular response to organic substance	BP	0.00329392
GO:0051146	19	130	striated muscle cell differentiation	BP	0.00360865
GO:0005539	15	87	glycosaminoglycan binding	MF	0.00380895
GO:0006809	8	26	nitric oxide biosynthetic process	BP	0.00398674
GO:0045499	7	20	chemorepellent activity	MF	0.00411016
GO:0010646	110	1511	regulation of cell communication	BP	0.00465748
GO:0023051	110	1513	regulation of signaling	BP	0.00494754
GO:0046209	8	27	nitric oxide metabolic process	BP	0.00516867
GO:2001057	8	27	reactive nitrogen species metabolic process	BP	0.00516867
GO:0010033	90	1180	response to organic substance	BP	0.00533949
GO:0071526	8	28	semaphorin-plexin signaling pathway	BP	0.00536852
GO:0010562	40	407	positive regulation of phosphorus metabolic process	BP	0.00536852
GO:0045937	40	407	positive regulation of phosphate metabolic process	BP	0.00536852
GO:0048729	31	286	tissue morphogenesis	BP	0.00555142
GO:0007507	28	248	heart development	BP	0.00586857
GO:0038023	49	538	signaling receptor activity	MF	0.00588954
GO:0060089	49	538	molecular transducer activity	MF	0.00588954
GO:0050839	19	136	cell adhesion molecule binding	MF	0.00593274
GO:0061138	16	103	morphogenesis of a branching epithelium	BP	0.00607165
GO:0048584	74	926	positive regulation of response to stimulus	BP	0.00607301
GO:0072593	17	114	reactive oxygen species metabolic process	BP	0.00749689
GO:0048754	14	84	branching morphogenesis of an epithelial tube	BP	0.00783708
GO:0005604	10	47	basement membrane	CC	0.00783708
GO:0009719	52	594	response to endogenous stimulus	BP	0.00783708
GO:0003179	7	22	heart valve morphogenesis	BP	0.00783708
GO:0043491	14	84	protein kinase B signaling	BP	0.0081716
GO:0045428	7	22	regulation of nitric oxide biosynthetic process	BP	0.00872716
GO:0080164	7	22	regulation of nitric oxide metabolic process	BP	0.00872716
GO:0043405	18	129	regulation of MAP kinase activity	BP	0.00957751

GO:0042325	53	613	regulation of phosphorylation	BP	0.00964846
GO:0030595	16	105	leukocyte chemotaxis	BP	0.00972493
GO:0048522	161	2439	positive regulation of cellular process	BP	0.01037156
GO:0006955	47	519	immune response	BP	0.01037914
GO:0042327	37	379	positive regulation of phosphorylation	BP	0.01048031
GO:0051128	84	1112	regulation of cellular component organization	BP	0.01078057
GO:0002009	27	245	morphogenesis of an epithelium	BP	0.0111636
GO:0032535	24	205	regulation of cellular component size	BP	0.0115273
GO:0048843	6	17	negative regulation of axon extension involved in axon guidance	BP	0.01213417
GO:0009887	43	471	animal organ morphogenesis	BP	0.01213417
GO:0001935	12	67	endothelial cell proliferation	BP	0.01283485
GO:1902895	7	24	positive regulation of pri-miRNA transcription by RNA polymerase II	BP	0.01395865
GO:0071495	48	549	cellular response to endogenous stimulus	BP	0.01398394
GO:0110110	6	17	positive regulation of animal organ morphogenesis	BP	0.01428741
GO:0001763	16	112	morphogenesis of a branching structure	BP	0.01430388
GO:0050920	16	110	regulation of chemotaxis	BP	0.01440601
GO:0001706	8	32	endoderm formation	BP	0.01445071
GO:0022407	23	195	regulation of cell-cell adhesion	BP	0.01445071
GO:0030855	28	262	epithelial cell differentiation	BP	0.01482496
GO:0003674	497	9138	molecular_function	MF	0.01482496
GO:0003158	11	59	endothelium development	BP	0.01513296
GO:0045321	34	343	leukocyte activation	BP	0.01553349
GO:0048841	6	18	regulation of axon extension involved in axon guidance	BP	0.01581163
GO:0003013	24	210	circulatory system process	BP	0.01602124
GO:0003170	7	25	heart valve development	BP	0.01602124
GO:0007517	18	136	muscle organ development	BP	0.01602124
GO:0071711	6	18	basement membrane organization	BP	0.01687663
GO:0031674	11	60	I band	CC	0.01728816
GO:0001570	9	42	vasculogenesis	BP	0.01973115
GO:0046649	30	293	lymphocyte activation	BP	0.02006406
GO:0015629	27	256	actin cytoskeleton	CC	0.02077303
GO:0007167	41	457	enzyme linked receptor protein signaling pathway	BP	0.02084748
GO:0030097	36	380	hemopoiesis	BP	0.02103808
GO:0019220	56	686	regulation of phosphate metabolic process	BP	0.02313238
GO:0051174	56	686	regulation of phosphorus metabolic process	BP	0.02313238
GO:0048598	30	299	embryonic morphogenesis	BP	0.02369539
GO:0007399	77	1034	nervous system development	BP	0.02389813
GO:0005178	11	63	integrin binding	MF	0.02540618
GO:0071560	15	107	cellular response to transforming growth factor beta stimulus	BP	0.02558062
GO:0045661	7	27	regulation of myoblast differentiation	BP	0.02579097
GO:1902532	27	259	negative regulation of intracellular signal transduction	BP	0.02579097
GO:0007492	9	44	endoderm development	BP	0.02579097
GO:0001503	20	167	ossification	BP	0.02619858
GO:0001891	5	13	phagocytic cup	CC	0.02648329
GO:0009611	22	192	response to wounding	BP	0.02648329
GO:0048846	6	20	axon extension involved in axon guidance	BP	0.02648329
GO:1902284	6	20	neuron projection extension involved in neuron projection guidance	BP	0.02648329
GO:0071559	15	108	response to transforming growth factor beta	BP	0.02715755
GO:0045446	10	54	endothelial cell differentiation	BP	0.02753532
GO:0030864	8	36	cortical actin cytoskeleton	CC	0.02782299
GO:1902533	40	446	positive regulation of intracellular signal transduction	BP	0.02782299
GO:0001228	23	207	DNA-binding transcription activator activity, RNA polymerase II-specific	MF	0.02782299
GO:0070161	30	304	anchoring junction	CC	0.02782299
GO:0001932	45	523	regulation of protein phosphorylation	BP	0.02807497
GO:0045596	29	289	negative regulation of cell differentiation	BP	0.02807497
GO:0038024	8	36	cargo receptor activity	MF	0.02851097
GO:0048534	37	403	hematopoietic or lymphoid organ development	BP	0.02879858
GO:0005520	6	20	insulin-like growth factor binding	MF	0.02889099
GO:0000904	32	336	cell morphogenesis involved in differentiation	BP	0.02889099
GO:1903037	18	143	regulation of leukocyte cell-cell adhesion	BP	0.02915543
GO:0031224	187	2968	intrinsic component of membrane	CC	0.02948448
GO:1902893	7	28	regulation of pri-miRNA transcription RNA polymerase II	BP	0.03061266
GO:0051090	22	195	regulation of DNA-binding transcription factor activity	BP	0.03071052
GO:0045019	4	8	negative regulation of nitric oxide biosynthetic process	BP	0.03073426
GO:1904406	4	8	negative regulation of nitric oxide metabolic process	BP	0.03073426
GO:0007411	16	123	axon guidance	BP	0.03083164

GO:0097485	16	123	neuron projection guidance	BP	0.03083164
GO:0042110	23	208	T cell activation	BP	0.03085798
GO:0030016	14	98	myofibril	CC	0.03125949
GO:0043410	25	236	positive regulation of MAPK cascade	BP	0.03219497
GO:0009967	55	684	positive regulation of signal transduction	BP	0.03219923
GO:0009897	15	109	external side of plasma membrane	CC	0.03219923
GO:0060562	19	160	epithelial tube morphogenesis	BP	0.03230314
GO:0001216	23	211	DNA-binding transcription activator activity	MF	0.03338346
GO:0007159	19	158	leukocyte cell-cell adhesion	BP	0.03355483
GO:0004697	5	14	protein kinase C activity	MF	0.03369132
GO:0004698	5	14	calcium-dependent protein kinase C activity	MF	0.03369132
GO:0030863	9	47	cortical cytoskeleton	CC	0.03402227
GO:0008284	34	365	positive regulation of cell population proliferation	BP	0.03402227
GO:0060348	13	89	bone development	BP	0.03557812
GO:0061614	7	29	pri-miRNA transcription by RNA polymerase II	BP	0.03557812
GO:0050790	64	832	regulation of catalytic activity	BP	0.03557812
GO:0016310	68	898	phosphorylation	BP	0.03644726
GO:0043292	14	100	contractile fiber	CC	0.03644726
GO:0009968	50	611	negative regulation of signal transduction	BP	0.03665778
GO:0003018	13	88	vascular process in circulatory system	BP	0.03665778
GO:0032970	22	200	regulation of actin filament-based process	BP	0.03665778
GO:0015803	3	4	branched-chain amino acid transport	BP	0.03665778
GO:0098713	3	4	leucine import across plasma membrane	BP	0.03665778
GO:1903801	3	4	L-leucine import across plasma membrane	BP	0.03665778
GO:0030017	13	89	sarcomere	CC	0.03665778
GO:0009628	36	397	response to abiotic stimulus	BP	0.03686273
GO:0008201	10	57	heparin binding	MF	0.03686273
GO:0030030	55	695	cell projection organization	BP	0.03686273
GO:0002694	23	212	regulation of leukocyte activation	BP	0.03693878
GO:0048017	13	90	inositol lipid-mediated signaling	BP	0.03693878
GO:0042482	3	4	positive regulation of odontogenesis	BP	0.0372017
GO:0007162	17	137	negative regulation of cell adhesion	BP	0.0373632
GO:0008361	12	79	regulation of cell size	BP	0.0373632
GO:0048866	3	4	stem cell fate specification	BP	0.03807676
GO:0050919	6	22	negative chemotaxis	BP	0.03896419
GO:0051058	7	30	negative regulation of small GTPase mediated signal transduction	BP	0.03980222
GO:0001501	23	216	skeletal system development	BP	0.03980222
GO:0005518	7	30	collagen binding	MF	0.03980222
GO:0006690	8	38	icosanoid metabolic process	BP	0.04001512
GO:0001885	7	30	endothelial cell development	BP	0.04100132
GO:2000377	11	68	regulation of reactive oxygen species metabolic process	BP	0.04193915
GO:0030234	53	665	enzyme regulator activity	MF	0.043315
GO:0002262	11	69	myeloid cell homeostasis	BP	0.04342921
GO:0016021	183	2927	integral component of membrane	CC	0.04364779
GO:1903039	14	102	positive regulation of leukocyte cell-cell adhesion	BP	0.04524111
GO:0005488	389	6942	binding	MF	0.04568986
GO:0001936	10	59	regulation of endothelial cell proliferation	BP	0.0463656
GO:0002684	33	357	positive regulation of immune system process	BP	0.04688176
GO:0044092	36	403	negative regulation of molecular function	BP	0.04688176
GO:0006044	4	9	N-acetylglucosamine metabolic process	BP	0.04688176
GO:0019955	10	59	cytokine binding	MF	0.04709407
GO:0048762	16	128	mesenchymal cell differentiation	BP	0.04717841
GO:0008277	9	49	regulation of G protein-coupled receptor signaling pathway	BP	0.0474284
GO:0050865	24	231	regulation of cell activation	BP	0.0474284
GO:0007169	28	291	transmembrane receptor protein tyrosine kinase signaling pathway	BP	0.0480754
GO:0031589	19	168	cell-substrate adhesion	BP	0.04864514
GO:0014902	10	60	myotube differentiation	BP	0.04943591
GO:0030695	29	308	GTPase regulator activity	MF	0.04989088
GO:0085029	6	23	extracellular matrix assembly	BP	0.04993619

Appendix C4 Table 2: Gene ontology (GO) enrichment analysis result on **upregulated** differentially expressed genes between New Zealand and western Australian brushtail possums identified using **DESeq2**. The analysis approximates the true distribution of numbers of members of a category amongst

differentially expressed genes by the Wallenius non-central hypergeometric distribution. False Discovery Rate (FDR) is then calculated using Benjamini and Hochberg method.

GO category	Number of genes	Number in category	Term	Ontology	P-value
GO:0044281	42	883	small molecule metabolic process	BP	5.59E-10
GO:0006082	29	452	organic acid metabolic process	BP	4.25E-09
GO:0019752	28	433	carboxylic acid metabolic process	BP	6.30E-09
GO:0043436	28	439	oxoacid metabolic process	BP	6.45E-09
GO:0046395	15	114	carboxylic acid catabolic process	BP	5.41E-08
GO:0016054	15	118	organic acid catabolic process	BP	7.55E-08
GO:0044282	17	167	small molecule catabolic process	BP	1.05E-07
GO:0006629	31	669	lipid metabolic process	BP	6.39E-07
GO:0003824	86	3924	catalytic activity	MF	1.67E-06
GO:0006520	14	147	cellular amino acid metabolic process	BP	1.15E-05
GO:0016491	23	457	oxidoreductase activity	MF	3.49E-05
GO:0009063	9	53	cellular amino acid catabolic process	BP	3.49E-05
GO:0044255	23	521	cellular lipid metabolic process	BP	0.000275976
GO:0016627	7	36	oxidoreductase activity, acting on the CH-CH group of donors	MF	0.000425603
GO:1901605	10	105	alpha-amino acid metabolic process	BP	0.001240989
GO:0016829	11	138	lyase activity	MF	0.002027212
GO:0050660	7	50	flavin adenine dinucleotide binding	MF	0.00316724
GO:0032787	14	262	monocarboxylic acid metabolic process	BP	0.007331718
GO:0019842	8	84	vitamin binding	MF	0.012260175

Appendix C4 Table 3: Gene ontology (GO) enrichment analysis result on **downregulated** differentially expressed genes between New Zealand and western Australian brushtail possums identified using **Limma**. The analysis approximates the true distribution of numbers of members of a category amongst differentially expressed genes by the Wallenius non-central hypergeometric distribution. False Discovery Rate (FDR) is then calculated using Benjamini and Hochberg method.

GO category	Number of genes	Number in category	Term	Ontology	P-value
GO:0071944	278	2639	cell periphery	CC	8.22E-25
GO:0005886	240	2410	plasma membrane	CC	8.03E-17
GO:0016477	108	735	cell migration	BP	2.26E-16
GO:0040011	118	896	locomotion	BP	1.48E-14
GO:0048870	110	806	cell motility	BP	1.48E-14
GO:0051674	110	806	localization of cell	BP	1.48E-14
GO:0003735	70	257	structural constituent of ribosome	MF	3.51E-14
GO:0006928	125	1035	movement of cell or subcellular component	BP	7.85E-13
GO:0005840	71	303	ribosome	CC	1.64E-11
GO:0022626	48	149	cytosolic ribosome	CC	2.08E-11
GO:0005198	80	414	structural molecule activity	MF	5.88E-11
GO:0006935	53	299	chemotaxis	BP	5.75E-10
GO:0042330	53	300	taxis	BP	6.00E-10
GO:0035239	63	422	tube morphogenesis	BP	9.69E-10
GO:0001525	44	241	angiogenesis	BP	3.37E-09
GO:0030334	66	454	regulation of cell migration	BP	3.80E-09
GO:0009653	137	1360	anatomical structure morphogenesis	BP	7.26E-09
GO:0007155	80	655	cell adhesion	BP	1.21E-08
GO:0022610	80	658	biological adhesion	BP	1.47E-08
GO:0048856	233	2746	anatomical structure development	BP	1.90E-08
GO:0048646	72	559	anatomical structure formation involved in morphogenesis	BP	2.31E-08
GO:0023052	247	2964	signaling	BP	2.48E-08
GO:0040012	68	500	regulation of locomotion	BP	2.59E-08
GO:2000145	66	479	regulation of cell motility	BP	2.61E-08
GO:0044391	53	227	ribosomal subunit	CC	3.16E-08
GO:0032502	245	2950	developmental process	BP	3.44E-08
GO:0007154	247	2984	cell communication	BP	4.22E-08
GO:0001568	51	334	blood vessel development	BP	4.38E-08
GO:0048514	47	294	blood vessel morphogenesis	BP	5.07E-08
GO:0048731	197	2253	system development	BP	5.18E-08
GO:0051270	67	501	regulation of cellular component movement	BP	5.18E-08
GO:0009986	49	315	cell surface	CC	6.03E-08
GO:0001944	52	351	vasculature development	BP	7.23E-08
GO:0072359	67	526	circulatory system development	BP	8.06E-08
GO:0035295	66	514	tube development	BP	8.10E-08
GO:0032501	267	3300	multicellular organismal process	BP	8.10E-08
GO:0007165	226	2692	signal transduction	BP	9.30E-08
GO:0022625	30	89	cytosolic large ribosomal subunit	CC	1.26E-07
GO:0001667	39	225	ameboidal-type cell migration	BP	2.50E-07
GO:0007275	209	2477	multicellular organism development	BP	2.50E-07
GO:0005576	104	916	extracellular region	CC	4.02E-07
GO:0042221	157	1716	response to chemical	BP	5.61E-07
GO:0030198	29	146	extracellular matrix organization	BP	6.47E-07
GO:0045229	29	146	external encapsulating structure organization	BP	6.47E-07
GO:0031226	79	707	intrinsic component of plasma membrane	CC	6.47E-07
GO:0043062	29	147	extracellular structure organization	BP	7.32E-07
GO:0050896	312	4090	response to stimulus	BP	7.32E-07
GO:0005887	75	674	integral component of plasma membrane	CC	1.71E-06
GO:0005615	78	627	extracellular space	CC	1.71E-06
GO:0043043	72	482	peptide biosynthetic process	BP	2.13E-06
GO:0022603	54	416	regulation of anatomical structure morphogenesis	BP	2.37E-06

GO:0006412	71	472	translation	BP	2.37E-06
GO:0009605	109	1074	response to external stimulus	BP	2.37E-06
GO:0007166	119	1250	cell surface receptor signaling pathway	BP	3.24E-06
GO:0030312	35	214	external encapsulating structure	CC	3.26E-06
GO:0031012	35	214	extracellular matrix	CC	3.26E-06
GO:0051239	121	1268	regulation of multicellular organismal process	BP	3.86E-06
GO:0043604	77	555	amide biosynthetic process	BP	3.86E-06
GO:0040013	29	157	negative regulation of locomotion	BP	4.88E-06
GO:0030154	163	1899	cell differentiation	BP	5.03E-06
GO:2000026	70	626	regulation of multicellular organismal development	BP	7.70E-06
GO:0048869	165	1941	cellular developmental process	BP	7.98E-06
GO:0009888	93	926	tissue development	BP	8.95E-06
GO:0030335	39	254	positive regulation of cell migration	BP	9.06E-06
GO:0048513	141	1587	animal organ development	BP	1.12E-05
GO:0003018	21	92	vascular process in circulatory system	BP	1.18E-05
GO:0040017	40	271	positive regulation of locomotion	BP	1.80E-05
GO:0051716	263	3470	cellular response to stimulus	BP	1.84E-05
GO:0050878	27	149	regulation of body fluid levels	BP	2.01E-05
GO:0050793	107	1140	regulation of developmental process	BP	2.19E-05
GO:0070887	120	1301	cellular response to chemical stimulus	BP	2.36E-05
GO:0010631	28	157	epithelial cell migration	BP	2.64E-05
GO:0060326	28	147	cell chemotaxis	BP	2.67E-05
GO:0002376	100	1020	immune system process	BP	2.67E-05
GO:2000147	39	265	positive regulation of cell motility	BP	2.70E-05
GO:0031224	248	3223	intrinsic component of membrane	CC	2.92E-05
GO:0051272	39	267	positive regulation of cellular component movement	BP	3.16E-05
GO:0090132	28	159	epithelium migration	BP	3.40E-05
GO:0050900	29	162	leukocyte migration	BP	3.42E-05
GO:0061061	41	296	muscle structure development	BP	3.73E-05
GO:0051179	242	3177	localization	BP	3.97E-05
GO:0098552	33	211	side of membrane	CC	3.98E-05
GO:0030155	45	351	regulation of cell adhesion	BP	4.05E-05
GO:0090130	28	162	tissue migration	BP	4.56E-05
GO:0003013	34	224	circulatory system process	BP	4.64E-05
GO:0051240	71	666	positive regulation of multicellular organismal process	BP	5.99E-05
GO:0016021	243	3178	integral component of membrane	CC	6.00E-05
GO:0016020	351	4947	membrane	CC	7.62E-05
GO:0006518	74	570	peptide metabolic process	BP	7.62E-05
GO:0030336	24	132	negative regulation of cell migration	BP	7.84E-05
GO:0045766	16	65	positive regulation of angiogenesis	BP	0.000102199
GO:1904018	16	65	positive regulation of vasculature development	BP	0.000102199
GO:0009897	22	115	external side of plasma membrane	CC	0.000120821
GO:0098609	46	385	cell-cell adhesion	BP	0.000141967
GO:0043534	16	64	blood vessel endothelial cell migration	BP	0.000144044
GO:0045765	22	119	regulation of angiogenesis	BP	0.000149837
GO:0015934	31	138	large ribosomal subunit	CC	0.000185291
GO:0051241	53	454	negative regulation of multicellular organismal process	BP	0.000185438
GO:2000146	24	139	negative regulation of cell motility	BP	0.000190075
GO:1901342	22	121	regulation of vasculature development	BP	0.000190075
GO:0003674	640	9972	molecular_function	MF	0.00019809
GO:0051271	24	141	negative regulation of cellular component movement	BP	0.00023512
GO:0008015	31	209	blood circulation	BP	0.00023512
GO:0035150	16	68	regulation of tube size	BP	0.00023512
GO:0035296	16	68	regulation of tube diameter	BP	0.00023512
GO:0097746	16	68	blood vessel diameter maintenance	BP	0.00023512
GO:0001775	50	419	cell activation	BP	0.000241569
GO:0042310	12	41	vasoconstriction	BP	0.000263301
GO:0050920	22	114	regulation of chemotaxis	BP	0.000294919
GO:0060429	59	547	epithelium development	BP	0.000311263
GO:0001763	21	118	morphogenesis of a branching structure	BP	0.000311342
GO:0005102	78	746	signaling receptor binding	MF	0.000318467

GO:0061138	20	109	morphogenesis of a branching epithelium	BP	0.000344267
GO:0043535	14	53	regulation of blood vessel endothelial cell migration	BP	0.000351623
GO:0090066	37	280	regulation of anatomical structure size	BP	0.000399166
GO:0030595	21	107	leukocyte chemotaxis	BP	0.000433286
GO:0048583	148	1801	regulation of response to stimulus	BP	0.000473003
GO:0030029	49	433	actin filament-based process	BP	0.000498151
GO:0003158	14	59	endothelium development	BP	0.000524587
GO:0030215	8	19	semaphorin receptor binding	MF	0.000545616
GO:0045499	8	20	chemorepellent activity	MF	0.000669109
GO:0043542	21	118	endothelial cell migration	BP	0.000785778
GO:0043603	82	716	cellular amide metabolic process	BP	0.000797012
GO:0045446	13	54	endothelial cell differentiation	BP	0.000947416
GO:0065008	152	1882	regulation of biological quality	BP	0.001199488
GO:0051094	61	602	positive regulation of developmental process	BP	0.001199488
GO:0042482	4	4	positive regulation of odontogenesis	BP	0.001218374
GO:0062023	20	120	collagen-containing extracellular matrix	CC	0.001392316
GO:0048585	75	795	negative regulation of response to stimulus	BP	0.00152925
GO:0022627	18	60	cytosolic small ribosomal subunit	CC	0.001720068
GO:0010632	21	125	regulation of epithelial cell migration	BP	0.001743482
GO:0006939	12	47	smooth muscle contraction	BP	0.001848242
GO:0010596	10	34	negative regulation of endothelial cell migration	BP	0.001862424
GO:0008283	81	868	cell population proliferation	BP	0.001902542
GO:0032879	111	1298	regulation of localization	BP	0.001950634
GO:0010646	131	1596	regulation of cell communication	BP	0.002129738
GO:0071310	92	1022	cellular response to organic substance	BP	0.002152834
GO:0002682	58	560	regulation of immune system process	BP	0.002187934
GO:0015935	22	89	small ribosomal subunit	CC	0.002209141
GO:0023051	131	1600	regulation of signaling	BP	0.002318927
GO:0009966	118	1411	regulation of signal transduction	BP	0.002330372
GO:0050839	21	139	cell adhesion molecule binding	MF	0.002401261
GO:0010633	11	43	negative regulation of epithelial cell migration	BP	0.002505908
GO:0071363	38	317	cellular response to growth factor stimulus	BP	0.002586612
GO:0038023	64	634	signaling receptor activity	MF	0.002635653
GO:0060089	64	634	molecular transducer activity	MF	0.002635653
GO:0007399	98	1138	nervous system development	BP	0.002651113
GO:0048843	7	18	negative regulation of axon extension involved in axon guidance	BP	0.002783539
GO:0035556	109	1291	intracellular signal transduction	BP	0.002789557
GO:0007411	21	137	axon guidance	BP	0.002789557
GO:0097485	21	137	neuron projection guidance	BP	0.002789557
GO:0001935	14	68	endothelial cell proliferation	BP	0.002822018
GO:0048523	175	2275	negative regulation of cellular process	BP	0.002888679
GO:0043537	8	23	negative regulation of blood vessel endothelial cell migration	BP	0.003009257
GO:0001885	9	30	endothelial cell development	BP	0.003009257
GO:0019229	9	30	regulation of vasoconstriction	BP	0.00310113
GO:0001936	13	60	regulation of endothelial cell proliferation	BP	0.003161409
GO:0009611	27	199	response to wounding	BP	0.003322079
GO:0006936	21	131	muscle contraction	BP	0.003397848
GO:0007162	21	139	negative regulation of cell adhesion	BP	0.003411065
GO:0030545	36	254	receptor regulator activity	MF	0.003468237
GO:0070848	38	323	response to growth factor	BP	0.003468237
GO:0030036	43	394	actin cytoskeleton organization	BP	0.003608271
GO:0065009	103	1195	regulation of molecular function	BP	0.003739966
GO:0048841	7	19	regulation of axon extension involved in axon guidance	BP	0.003767873
GO:0048018	35	244	receptor ligand activity	MF	0.003794803
GO:0097529	18	97	myeloid leukocyte migration	BP	0.003802932
GO:0048872	21	135	homeostasis of number of cells	BP	0.003830963
GO:1903522	17	98	regulation of blood circulation	BP	0.00398124
GO:0019955	13	61	cytokine binding	MF	0.003981822
GO:0030546	35	246	signaling receptor activator activity	MF	0.004264713
GO:0045785	27	203	positive regulation of cell adhesion	BP	0.004713484

GO:0042692	24	173	muscle cell differentiation	BP	0.00478557
GO:0003170	8	26	heart valve development	BP	0.004909644
GO:0032101	44	395	regulation of response to external stimulus	BP	0.005048783
GO:0009887	53	528	animal organ morphogenesis	BP	0.005251412
GO:0010033	105	1238	response to organic substance	BP	0.005512382
GO:0000165	41	363	MAPK cascade	BP	0.005696712
GO:0042127	69	734	regulation of cell population proliferation	BP	0.006329298
GO:0048846	7	21	axon extension involved in axon guidance	BP	0.007204482
GO:1902284	7	21	neuron projection extension involved in neuron projection guidance	BP	0.007204482
GO:0030855	34	289	epithelial cell differentiation	BP	0.007421761
GO:0071526	8	28	semaphorin-plexin signaling pathway	BP	0.007421761
GO:0007589	9	33	body fluid secretion	BP	0.007716315
GO:0002685	17	95	regulation of leukocyte migration	BP	0.007727876
GO:0048865	4	5	stem cell fate commitment	BP	0.007867922
GO:0045595	66	715	regulation of cell differentiation	BP	0.007867922
GO:0006809	8	26	nitric oxide biosynthetic process	BP	0.0081341
GO:0048468	85	993	cell development	BP	0.008284823
GO:0000902	51	523	cell morphogenesis	BP	0.009307343
GO:0048754	15	88	branching morphogenesis of an epithelial tube	BP	0.009449618
GO:0110110	7	20	positive regulation of animal organ morphogenesis	BP	0.009449618
GO:0050919	7	22	negative chemotaxis	BP	0.009449618
GO:0022407	26	202	regulation of cell-cell adhesion	BP	0.009461228
GO:0046209	8	27	nitric oxide metabolic process	BP	0.010259388
GO:2001057	8	27	reactive nitrogen species metabolic process	BP	0.010259388
GO:1901566	100	1058	organonitrogen compound biosynthetic process	BP	0.011332752
GO:0006955	54	534	immune response	BP	0.011577659
GO:0010594	16	94	regulation of endothelial cell migration	BP	0.012102706
GO:0061028	7	22	establishment of endothelial barrier	BP	0.012480781
GO:0002675	4	6	positive regulation of acute inflammatory response	BP	0.01255206
GO:0048017	15	93	inositol lipid-mediated signaling	BP	0.01294669
GO:0042325	60	641	regulation of phosphorylation	BP	0.013569799
GO:0032103	24	169	positive regulation of response to external stimulus	BP	0.013727805
GO:0005178	12	63	integrin binding	MF	0.013727805
GO:0003179	7	23	heart valve morphogenesis	BP	0.013778419
GO:0071560	18	117	cellular response to transforming growth factor beta stimulus	BP	0.013798753
GO:0048729	34	311	tissue morphogenesis	BP	0.013996671
GO:0007167	48	489	enzyme linked receptor protein signaling pathway	BP	0.014048034
GO:0002009	30	262	morphogenesis of an epithelium	BP	0.014048034
GO:0050699	5	11	WW domain binding	MF	0.014048034
GO:0030308	15	88	negative regulation of cell growth	BP	0.01407343
GO:0002064	17	111	epithelial cell development	BP	0.014357709
GO:0048762	19	132	mesenchymal cell differentiation	BP	0.014357709
GO:0071559	18	118	response to transforming growth factor beta	BP	0.014685336
GO:0050865	29	239	regulation of cell activation	BP	0.015050999
GO:0050729	10	42	positive regulation of inflammatory response	BP	0.015438586
GO:0030097	42	403	hemopoiesis	BP	0.0156619
GO:0019220	65	715	regulation of phosphate metabolic process	BP	0.0156619
GO:0051174	65	715	regulation of phosphorus metabolic process	BP	0.0156619
GO:0042110	27	219	T cell activation	BP	0.015886278
GO:0045428	7	22	regulation of nitric oxide biosynthetic process	BP	0.015937276
GO:0080164	7	22	regulation of nitric oxide metabolic process	BP	0.015937276
GO:0050921	14	71	positive regulation of chemotaxis	BP	0.016081629
GO:0045321	39	358	leukocyte activation	BP	0.016406637
GO:0007507	30	268	heart development	BP	0.016979947
GO:0044092	43	415	negative regulation of molecular function	BP	0.017280796
GO:0043408	37	336	regulation of MAPK cascade	BP	0.017985807
GO:0030516	9	40	regulation of axon extension	BP	0.018514035
GO:2000027	11	55	regulation of animal organ morphogenesis	BP	0.018734958
GO:0002694	27	219	regulation of leukocyte activation	BP	0.019212781
GO:0050431	5	12	transforming growth factor beta binding	MF	0.01923078

GO:0098801	5	12	regulation of renal system process	BP	0.019238881
GO:0030517	7	25	negative regulation of axon extension	BP	0.019368475
GO:0001527	4	7	microfibril	CC	0.019368475
GO:0048519	184	2509	negative regulation of biological process	BP	0.0193922
GO:0071674	15	83	mononuclear cell migration	BP	0.019586867
GO:0045597	40	390	positive regulation of cell differentiation	BP	0.019586867
GO:0019932	17	113	second-messenger-mediated signaling	BP	0.019586867
GO:0050673	24	193	epithelial cell proliferation	BP	0.019793978
GO:0007156	12	69	homophilic cell adhesion via plasma membrane adhesion molecules	BP	0.019793978
GO:0007596	15	94	blood coagulation	BP	0.019913252
GO:0007599	15	94	hemostasis	BP	0.019913252
GO:0042060	21	159	wound healing	BP	0.020178881
GO:1902531	70	792	regulation of intracellular signal transduction	BP	0.020312372
GO:0050678	21	159	regulation of epithelial cell proliferation	BP	0.021203152
GO:0015179	8	32	L-amino acid transmembrane transporter activity	MF	0.022053033
GO:0071711	6	18	basement membrane organization	BP	0.02289995
GO:0008201	12	64	heparin binding	MF	0.023160264
GO:0051146	19	140	striated muscle cell differentiation	BP	0.02365143
GO:0050679	14	86	positive regulation of epithelial cell proliferation	BP	0.024072412
GO:0006954	31	261	inflammatory response	BP	0.024702243
GO:0043536	8	29	positive regulation of blood vessel endothelial cell migration	BP	0.024853065
GO:0008016	13	76	regulation of heart contraction	BP	0.025801427
GO:0048534	43	428	hematopoietic or lymphoid organ development	BP	0.025912981
GO:1903037	20	147	regulation of leukocyte cell-cell adhesion	BP	0.025976403
GO:0005925	14	91	focal adhesion	CC	0.02700688
GO:0014065	12	70	phosphatidylinositol 3-kinase signaling	BP	0.027148169
GO:0050817	15	97	coagulation	BP	0.027722983
GO:0048015	14	91	phosphatidylinositol-mediated signaling	BP	0.027722983
GO:0005539	15	95	glycosaminoglycan binding	MF	0.028005787
GO:0010648	59	662	negative regulation of cell communication	BP	0.028288913
GO:0023057	59	663	negative regulation of signaling	BP	0.029210082
GO:0051128	94	1164	regulation of cellular component organization	BP	0.029649712
GO:0050790	74	859	regulation of catalytic activity	BP	0.030178326
GO:0032637	10	45	interleukin-8 production	BP	0.031170427
GO:0032677	10	45	regulation of interleukin-8 production	BP	0.031170427
GO:1903039	16	105	positive regulation of leukocyte cell-cell adhesion	BP	0.0316989
GO:0050870	15	95	positive regulation of T cell activation	BP	0.0316989
GO:0007159	21	162	leukocyte cell-cell adhesion	BP	0.0316989
GO:0030055	14	93	cell-substrate junction	CC	0.032000283
GO:0015629	29	262	actin cytoskeleton	CC	0.032798521
GO:0009968	56	625	negative regulation of signal transduction	BP	0.032995193
GO:0051249	24	195	regulation of lymphocyte activation	BP	0.033630822
GO:0010562	43	432	positive regulation of phosphorus metabolic process	BP	0.033630822
GO:0045937	43	432	positive regulation of phosphate metabolic process	BP	0.033630822
GO:0055078	6	20	sodium ion homeostasis	BP	0.033957667
GO:0010975	22	181	regulation of neuron projection development	BP	0.033957667
GO:0050794	366	5545	regulation of cellular process	BP	0.03491304
GO:0031589	21	172	cell-substrate adhesion	BP	0.035437758
GO:0002520	45	464	immune system development	BP	0.036344882
GO:0071621	11	53	granulocyte chemotaxis	BP	0.037144925
GO:0015803	3	4	branched-chain amino acid transport	BP	0.037886539
GO:0098713	3	4	leucine import across plasma membrane	BP	0.037886539
GO:1903801	3	4	L-leucine import across plasma membrane	BP	0.037886539
GO:0060485	20	157	mesenchyme development	BP	0.038040146
GO:0048589	32	303	developmental growth	BP	0.038040146
GO:0006940	7	25	regulation of smooth muscle contraction	BP	0.038148012
GO:0051093	41	416	negative regulation of developmental process	BP	0.038148012
GO:0046649	33	307	lymphocyte activation	BP	0.038148012
GO:0004888	53	561	transmembrane signaling receptor activity	MF	0.038156131
GO:0001932	51	550	regulation of protein phosphorylation	BP	0.038334047

GO:0050863	19	143	regulation of T cell activation	BP	0.038867485
GO:0070254	4	8	mucus secretion	BP	0.038867485
GO:0003012	21	163	muscle system process	BP	0.039544296
GO:0008361	13	84	regulation of cell size	BP	0.039544296
GO:0098772	106	1280	molecular function regulator	MF	0.039608279
GO:0060562	21	171	epithelial tube morphogenesis	BP	0.039616417
GO:0045019	4	8	negative regulation of nitric oxide biosynthetic process	BP	0.039616417
GO:1904406	4	8	negative regulation of nitric oxide metabolic process	BP	0.039616417
GO:0001938	8	36	positive regulation of endothelial cell proliferation	BP	0.040951815
GO:1904036	6	20	negative regulation of epithelial cell apoptotic process	BP	0.040951815
GO:0001816	35	330	cytokine production	BP	0.041163058
GO:0060560	16	115	developmental growth involved in morphogenesis	BP	0.041873227
GO:0022409	17	121	positive regulation of cell-cell adhesion	BP	0.042666027
GO:0097530	12	64	granulocyte migration	BP	0.044874751
GO:0022408	14	93	negative regulation of cell-cell adhesion	BP	0.045113566
GO:1904062	20	153	regulation of cation transmembrane transport	BP	0.045226009
GO:0003014	8	37	renal system process	BP	0.046014537
GO:0071345	33	311	cellular response to cytokine stimulus	BP	0.047032977
GO:0002042	7	28	cell migration involved in sprouting angiogenesis	BP	0.047576242
GO:0014033	11	63	neural crest cell differentiation	BP	0.048588686
GO:0034097	36	349	response to cytokine	BP	0.048774884
GO:0014829	4	9	vascular associated smooth muscle contraction	BP	0.048774884
GO:0098858	14	96	actin-based cell projection	CC	0.049844877
GO:0031532	11	62	actin cytoskeleton reorganization	BP	0.049844877

Appendix C4 Table 4: Gene ontology (GO) enrichment analysis result on **upregulated** differentially expressed genes between New Zealand and western Australian brushtail possums identified using **Limma**. The analysis approximates the true distribution of numbers of members of a category amongst differentially expressed genes by the Wallenius non-central hypergeometric distribution. False Discovery Rate (FDR) is then calculated using Benjamini and Hochberg method.

GO category	Number of genes	Number in category	Term	Ontology	P-value
GO:0044281	22	910	small molecule metabolic process	BP	0.003069542
GO:0016829	9	142	lyase activity	MF	0.00715456
GO:0019752	14	445	carboxylic acid metabolic process	BP	0.012505174
GO:0043436	14	451	oxoacid metabolic process	BP	0.012505174
GO:0006082	14	464	organic acid metabolic process	BP	0.01414524
GO:0003824	46	4043	catalytic activity	MF	0.036975684
GO:0046395	7	115	carboxylic acid catabolic process	BP	0.03892264
GO:0016054	7	119	organic acid catabolic process	BP	0.043202279
GO:0044282	8	169	small molecule catabolic process	BP	0.048727616

Appendix C4 Table 5: Gene ontology (GO) enrichment analysis result on **downregulated** differentially expressed genes between New Zealand and western Australian brushtail possums identified using **WGCNA**. The analysis approximates the true distribution of numbers of members of a category amongst differentially expressed genes by the Wallenius non-central hypergeometric distribution. False Discovery Rate (FDR) is then calculated using Benjamini and Hochberg method.

GO category	Number of genes	Number in category	Term	Ontology	P-value
GO:0071944	250	2558	cell periphery	CC	1.24E-15
GO:0005886	222	2313	plasma membrane	CC	3.01E-12
GO:0031226	88	671	intrinsic component of plasma membrane	CC	4.39E-10
GO:0005887	84	642	integral component of plasma membrane	CC	1.29E-09
GO:0040011	103	865	locomotion	BP	5.77E-09
GO:0003735	61	241	structural constituent of ribosome	MF	2.08E-08

GO:0016477	87	710	cell migration	BP	5.00E-08
GO:0022626	43	139	cytosolic ribosome	CC	9.71E-08
GO:0005198	75	402	structural molecule activity	MF	9.71E-08
GO:0001525	42	239	angiogenesis	BP	9.71E-08
GO:0032501	261	3153	multicellular organismal process	BP	3.06E-07
GO:0005840	63	285	ribosome	CC	4.39E-07
GO:0048870	89	777	cell motility	BP	5.35E-07
GO:0051674	89	777	localization of cell	BP	5.35E-07
GO:0048514	45	293	blood vessel morphogenesis	BP	1.46E-06
GO:0001568	49	337	blood vessel development	BP	1.46E-06
GO:0048646	67	545	anatomical structure formation involved in morphogenesis	BP	1.47E-06
GO:0035239	56	418	tube morphogenesis	BP	1.47E-06
GO:0001944	50	354	vasculature development	BP	2.23E-06
GO:0050896	311	3993	response to stimulus	BP	2.78E-06
GO:0006928	104	1007	movement of cell or subcellular component	BP	2.81E-06
GO:0009653	125	1311	anatomical structure morphogenesis	BP	2.94E-06
GO:0031224	252	3094	intrinsic component of membrane	CC	4.16E-06
GO:0072359	63	522	circulatory system development	BP	4.58E-06
GO:0016021	248	3049	integral component of membrane	CC	5.97E-06
GO:0030198	29	159	extracellular matrix organization	BP	1.26E-05
GO:0045229	29	159	external encapsulating structure organization	BP	1.26E-05
GO:0043062	29	160	extracellular structure organization	BP	1.40E-05
GO:0007186	49	327	G protein-coupled receptor signaling pathway	BP	1.46E-05
GO:0044391	48	214	ribosomal subunit	CC	2.44E-05
GO:0042221	152	1693	response to chemical	BP	2.55E-05
GO:0048856	213	2638	anatomical structure development	BP	3.61E-05
GO:0035295	59	507	tube development	BP	4.33E-05
GO:0042330	43	296	taxis	BP	5.59E-05
GO:0007275	194	2375	multicellular organism development	BP	5.59E-05
GO:0043043	70	464	peptide biosynthetic process	BP	5.84E-05
GO:0032502	224	2834	developmental process	BP	7.98E-05
GO:0022625	26	82	cytosolic large ribosomal subunit	CC	8.70E-05
GO:0006412	68	452	translation	BP	0.00011794
GO:0006935	42	295	chemotaxis	BP	0.00012879
GO:0007154	227	2894	cell communication	BP	0.00012879
GO:0023052	225	2874	signaling	BP	0.00016264
GO:0045765	23	121	regulation of angiogenesis	BP	0.00016264
GO:0043604	74	534	amide biosynthetic process	BP	0.00016352
GO:0005576	95	893	extracellular region	CC	0.0001699
GO:0009888	87	890	tissue development	BP	0.00017318
GO:0003018	20	93	vascular process in circulatory system	BP	0.00017707
GO:1901342	23	123	regulation of vasculature development	BP	0.00019499
GO:0048731	176	2165	system development	BP	0.00020586
GO:0003013	33	221	circulatory system process	BP	0.00028639
GO:0030334	53	445	regulation of cell migration	BP	0.00030194
GO:0007165	207	2616	signal transduction	BP	0.00041373
GO:0040013	26	157	negative regulation of locomotion	BP	0.0004222
GO:0040012	56	486	regulation of locomotion	BP	0.00042654
GO:0038023	62	562	signaling receptor activity	MF	0.00042654
GO:0060089	62	562	molecular transducer activity	MF	0.00042654
GO:0007155	67	659	cell adhesion	BP	0.00053844
GO:0022610	67	662	biological adhesion	BP	0.00063094
GO:0001667	32	222	ameboidal-type cell migration	BP	0.00094135
GO:0007411	22	133	axon guidance	BP	0.00113893
GO:0097485	22	133	neuron projection guidance	BP	0.00113893
GO:0008015	30	206	blood circulation	BP	0.00132718
GO:2000145	53	468	regulation of cell motility	BP	0.00132718
GO:0045766	15	66	positive regulation of angiogenesis	BP	0.00139168
GO:1904018	15	66	positive regulation of vasculature development	BP	0.00139168
GO:0090066	37	281	regulation of anatomical structure size	BP	0.0013923
GO:0003158	14	60	endothelium development	BP	0.00147078
GO:0007188	20	106	adenylate cyclase-modulating G protein-coupled receptor signaling pathway	BP	0.0016557
GO:0051716	254	3401	cellular response to stimulus	BP	0.00203384
GO:0004888	54	490	transmembrane signaling receptor activity	MF	0.002249
GO:0016020	344	4804	membrane	CC	0.00243365
GO:0098801	6	11	regulation of renal system process	BP	0.00243365
GO:0045446	13	55	endothelial cell differentiation	BP	0.00247937
GO:0043542	21	119	endothelial cell migration	BP	0.00249285
GO:0035150	15	68	regulation of tube size	BP	0.0025616
GO:0035296	15	68	regulation of tube diameter	BP	0.0025616

GO:0097746	15	68	blood vessel diameter maintenance	BP	0.0025616
GO:0051270	53	489	regulation of cellular component movement	BP	0.00386213
GO:0051239	110	1246	regulation of multicellular organismal process	BP	0.00400874
GO:0022603	46	414	regulation of anatomical structure morphogenesis	BP	0.00400874
GO:0006518	71	552	peptide metabolic process	BP	0.00433511
GO:0030336	21	131	negative regulation of cell migration	BP	0.00494627
GO:0010631	24	159	epithelial cell migration	BP	0.00616914
GO:0002064	18	107	epithelial cell development	BP	0.00626944
GO:0048872	22	138	homeostasis of number of cells	BP	0.00626944
GO:0003014	10	37	renal system process	BP	0.00626944
GO:0090132	24	160	epithelium migration	BP	0.00645519
GO:0043534	14	65	blood vessel endothelial cell migration	BP	0.00647379
GO:0044057	30	218	regulation of system process	BP	0.00666814
GO:0072015	4	5	glomerular visceral epithelial cell development	BP	0.00745805
GO:0090130	24	162	tissue migration	BP	0.00745805
GO:0030154	145	1815	cell differentiation	BP	0.00765424
GO:0015820	4	5	leucine transport	BP	0.0081708
GO:0098552	28	206	side of membrane	CC	0.00834102
GO:0009986	37	313	cell surface	CC	0.00875898
GO:0015934	28	129	large ribosomal subunit	CC	0.00875898
GO:2000146	21	139	negative regulation of cell motility	BP	0.01057933
GO:0048869	147	1857	cellular developmental process	BP	0.01079585
GO:0050793	96	1117	regulation of developmental process	BP	0.0110582
GO:2000026	60	617	regulation of multicellular organismal development	BP	0.01106072
GO:0003008	68	708	system process	BP	0.01176496
GO:0051271	21	141	negative regulation of cellular component movement	BP	0.01231986
GO:0003674	633	9677	molecular_function	MF	0.0126542
GO:0005615	65	612	extracellular space	CC	0.01461791
GO:0065008	148	1855	regulation of biological quality	BP	0.01560146
GO:0050900	24	159	leukocyte migration	BP	0.01560146
GO:0022627	17	57	cytosolic small ribosomal subunit	CC	0.01560146
GO:0008217	14	72	regulation of blood pressure	BP	0.0157623
GO:0060429	52	523	epithelium development	BP	0.01622488
GO:0050878	21	144	regulation of body fluid levels	BP	0.01622488
GO:0051093	44	414	negative regulation of developmental process	BP	0.01681283
GO:0051241	48	453	negative regulation of multicellular organismal process	BP	0.01707408
GO:0072310	4	6	glomerular epithelial cell development	BP	0.01768619
GO:0048588	14	81	developmental cell growth	BP	0.01781051
GO:0010596	9	34	negative regulation of endothelial cell migration	BP	0.01819867
GO:0030312	28	232	external encapsulating structure	CC	0.01819867
GO:0031012	28	232	extracellular matrix	CC	0.01819867
GO:0042310	10	42	vasoconstriction	BP	0.01906457
GO:0007193	9	34	adenylate cyclase-inhibiting G protein-coupled receptor signaling pathway	BP	0.02021302
GO:0010633	10	43	negative regulation of epithelial cell migration	BP	0.021656
GO:1903053	7	21	regulation of extracellular matrix organization	BP	0.021656
GO:0009605	94	1056	response to external stimulus	BP	0.021656
GO:0035850	7	23	epithelial cell differentiation involved in kidney development	BP	0.021656
GO:0006939	10	43	smooth muscle contraction	BP	0.02380716
GO:0000902	49	506	cell morphogenesis	BP	0.02697824
GO:0030855	32	276	epithelial cell differentiation	BP	0.02697824
GO:0048589	32	289	developmental growth	BP	0.03618568
GO:0048513	122	1524	animal organ development	BP	0.03774117
GO:0098609	39	383	cell-cell adhesion	BP	0.04070954
GO:0043603	77	695	cellular amide metabolic process	BP	0.04190722
GO:0061564	27	234	axon development	BP	0.04190722
GO:0070887	107	1291	cellular response to chemical stimulus	BP	0.04264025
GO:1901566	100	1026	organonitrogen compound biosynthetic process	BP	0.04421146
GO:0030215	6	19	semaphorin receptor binding	MF	0.04481037
GO:0019229	8	31	regulation of vasoconstriction	BP	0.04528714
GO:0015935	20	85	small ribosomal subunit	CC	0.0484273
GO:0046456	7	20	icosanoid biosynthetic process	BP	0.04962842

Chapter 5

Appendix C5 Table 1: List of positively selected genes among western Australian brushtail possums discovered using **CodeML**.

Gene	Gene Name
TSEN34	tRNA splicing endonuclease subunit 34(TSEN34)
DDX49	DEAD-box helicase 49(DDX49)
BMPR2	bone morphogenetic protein receptor type 2(BMPR2)
UNC93B1	unc-93 homolog B1, TLR signaling regulator(UNC93B1)
GDA	guanine deaminase(GDA)
JPT2	Jupiter microtubule associated homolog 2(JPT2)
MLLT1	MLLT1 super elongation complex subunit(MLLT1)
HSCB	HscB mitochondrial iron-sulfur cluster cochaperone(HSCB)
VPS26B	VPS26, retromer complex component B(VPS26B)
WDR24	WD repeat domain 24(WDR24)
RHOBTB1	Rho related BTB domain containing 1(RHOBTB1)
CENPB	centromere protein B(CENPB)
FAM117A	family with sequence similarity 117 member A(FAM117A)
MTDH	metadherin(MTDH)
DCAF1	DDB1 and CUL4 associated factor 1(DCAF1)
MNT	MAX network transcriptional repressor(MNT)
PSMD5	proteasome 26S subunit, non-ATPase 5(PSMD5)
TRAPPC11	trafficking protein particle complex subunit 11(TRAPPC11)
ZNF821	zinc finger protein 821(ZNF821)
GBP2	guanylate binding protein 2(GBP2)
ABCF3	ATP binding cassette subfamily F member 3(ABCF3)
SLC25A44	solute carrier family 25 member 44(SLC25A44)
FADS1	fatty acid desaturase 1(FADS1)
PDHX	pyruvate dehydrogenase complex component X(PDHX)
ZYG11B	zyg-11 family member B, cell cycle regulator(ZYG11B)
MAP1S	microtubule associated protein 1S(MAP1S)
UTP3	UTP3 small subunit processome component(UTP3)
TTC33	tetratricopeptide repeat domain 33(TTC33)
NCBP2	nuclear cap binding protein subunit 2(NCBP2)
VPS37C	VPS37C subunit of ESCRT-I(VPS37C)
SAMD4B	sterile alpha motif domain containing 4B(SAMD4B)
ALG3	ALG3 alpha-1,3- mannosyltransferase(ALG3)
KLHL22	kelch like family member 22(KLHL22)
ACSL3	acyl-CoA synthetase long chain family member 3(ACSL3)
SMC1A	structural maintenance of chromosomes 1A(SMC1A)
OXR1	oxidation resistance 1(OXR1)
WDR77	WD repeat domain 77(WDR77)
THOC6	THO complex subunit 6(THOC6)
MYBBP1A	MYB binding protein 1a(MYBBP1A)
SGSM3	small G protein signaling modulator 3(SGSM3)
COPG2	COPI coat complex subunit gamma 2(COPG2)
PNPLA6	patatin like phospholipase domain containing 6(PNPLA6)
CACTIN	cactin, spliceosome C complex subunit(CACTIN)
SSBP3	single stranded DNA binding protein 3(SSBP3)
MBTPS2	membrane bound transcription factor peptidase, site 2(MBTPS2)
PIGH	phosphatidylinositol glycan anchor biosynthesis class H(PIGH)

Appendix C5 Table 2: List of positively selected genes among western Australian brushtail possums discovered using **HYPHY Fubar**.

Gene	Gene Name
CLPB	caseinolytic mitochondrial matrix peptidase chaperone subunit B(CLPB)
SERPINE2	serpin family E member 2(SERPINE2)
TESK1	testis-specific kinase 1(TESK1)
IPP	intracisternal A particle-promoted polypeptide(IPP)
PHAX	phosphorylated adaptor for RNA export(PHAX)
RRP8	ribosomal RNA processing 8(RRP8)
ZFYVE27	zinc finger FYVE-type containing 27(ZFYVE27)
RNF114	ring finger protein 114(RNF114)
PSMD5	proteasome 26S subunit, non-ATPase 5(PSMD5)
ZMIZ2	zinc finger MIZ-type containing 2(ZMIZ2)
GJA4	gap junction protein alpha 4(GJA4)
ZC3H12A	zinc finger CCCH-type containing 12A(ZC3H12A)
PSMD1	proteasome 26S subunit, non-ATPase 1(PSMD1)
IER5	immediate early response 5(IER5)
EARS2	glutamyl-tRNA synthetase 2, mitochondrial(EARS2)
WLS	Wnt ligand secretion mediator(WLS)
IAH1	isoamyl acetate hydrolyzing esterase 1 (putative)(IAH1)
TYW5	tRNA-yW synthesizing protein 5(TYW5)
NHLRC3	NHL repeat containing 3(NHLRC3)
MAP1S	microtubule associated protein 1S(MAP1S)
PIH1D1	PIH1 domain containing 1(PIH1D1)
HDHD3	haloacid dehalogenase like hydrolase domain containing 3(HDHD3)
WDR77	WD repeat domain 77(WDR77)
EWSR1	EWS RNA binding protein 1(EWSR1)
RFX5	regulatory factor X5(RFX5)
ACOT2	acyl-CoA thioesterase 2(ACOT2)
TARS2	threonyl-tRNA synthetase 2, mitochondrial(TARS2)
PRR12	proline rich 12(PRR12)
RNF121	ring finger protein 121(RNF121)
ZNF276	zinc finger protein 276(ZNF276)
POM121	POM121 transmembrane nucleoporin(POM121)
PRKRIP1	PRKR interacting protein 1(PRKRIP1)
NOLC1	nucleolar and coiled-body phosphoprotein 1(NOLC1)
BLOC1S5	biogenesis of lysosomal organelles complex 1 subunit 5(BLOC1S5)
WDR70	WD repeat domain 70(WDR70)
PNPO	pyridoxamine 5'-phosphate oxidase(PNPO)
RRS1	ribosome biogenesis regulator 1 homolog(RRS1)
MCMBP	minichromosome maintenance complex binding protein(MCMBP)
LTA4H	leukotriene A4 hydrolase(LTA4H)
SLC19A1	solute carrier family 19 member 1(SLC19A1)
LHPP	phospholysine phosphohistidine inorganic pyrophosphate phosphatase(LHPP)
CHKB	choline kinase beta(CHKB)
MTERF3	mitochondrial transcription termination factor 3(MTERF3)
FUCA1	alpha-L-fucosidase 1(FUCA1)
SETD1A	SET domain containing 1A, histone lysine methyltransferase(SETD1A)
TRIP4	thyroid hormone receptor interactor 4(TRIP4)
TYSND1	trypsin like peroxisomal matrix peptidase 1(TYSND1)
MTHFS	methenyltetrahydrofolate synthetase(MTHFS)
PAFAH2	platelet activating factor acetylhydrolase 2(PAFAH2)
SNUPN	snurportin 1(SNUPN)
ABTB2	ankyrin repeat and BTB domain containing 2(ABTB2)
ABTB1	ankyrin repeat and BTB domain containing 1(ABTB1)
GNPTG	N-acetylglucosamine-1-phosphate transferase subunit gamma(GNPTG)
TBL3	transducin beta like 3(TBL3)
UCK1	uridine-cytidine kinase 1(UCK1)
RNF103	ring finger protein 103(RNF103)

TDP2	tyrosyl-DNA phosphodiesterase 2(TDP2)
HBP1	HMG-box transcription factor 1(HBP1)
PPT2	palmitoyl-protein thioesterase 2(PPT2)
DUS1L	dihydrouridine synthase 1 like(DUS1L)
MOCS2	molybdenum cofactor synthesis 2(MOCS2)
CDKN1B	cyclin dependent kinase inhibitor 1B(CDKN1B)
PIGO	phosphatidylinositol glycan anchor biosynthesis class O(PIGO)
HEXB	hexosaminidase subunit beta(HEXB)
KDM8	lysine demethylase 8(KDM8)
WDR46	WD repeat domain 46(WDR46)
TMEM186	transmembrane protein 186(TMEM186)
SPOUT1	SPOUT domain containing methyltransferase 1(SPOUT1)
FARS2	phenylalanyl-tRNA synthetase 2, mitochondrial(FARS2)
MED17	mediator complex subunit 17(MED17)
SMPD5	sphingomyelin phosphodiesterase 5(SMPD5)
SMPD1	sphingomyelin phosphodiesterase 1(SMPD1)
CIC	capicua transcriptional repressor(CIC)
WDR36	WD repeat domain 36(WDR36)
TIMM29	translocase of inner mitochondrial membrane 29(TIMM29)
TIGD7	tigger transposable element derived 7(TIGD7)
NSUN4	NOP2/Sun RNA methyltransferase 4(NSUN4)
NSUN5	NOP2/Sun RNA methyltransferase 5(NSUN5)
SPHK2	sphingosine kinase 2(SPHK2)
NSUN3	NOP2/Sun RNA methyltransferase 3(NSUN3)
KIAA1143	KIAA1143(KIAA1143)
MRPL48	mitochondrial ribosomal protein L48(MRPL48)
SIRT3	sirtuin 3(SIRT3)
LRRC42	leucine rich repeat containing 42(LRRC42)
PIGC	phosphatidylinositol glycan anchor biosynthesis class C(PIGC)
FAM222B	family with sequence similarity 222 member B(FAM222B)
KAT2A	lysine acetyltransferase 2A(KAT2A)
ZNF598	zinc finger protein 598, E3 ubiquitin ligase(ZNF598)
MOCS1	molybdenum cofactor synthesis 1(MOCS1)
PIGH	phosphatidylinositol glycan anchor biosynthesis class H(PIGH)
TP53I13	tumor protein p53 inducible protein 13(TP53I13)
TSR3	TSR3 ribosome maturation factor(TSR3)
MIEF2	mitochondrial elongation factor 2(MIEF2)
OVCA2	OVCA2 serine hydrolase domain containing(OVCA2)
UNG	uracil DNA glycosylase(UNG)
FAM117A	family with sequence similarity 117 member A(FAM117A)
GGA2	golgi associated, gamma adaptin ear containing, ARF binding protein 2(GGA2)
CCL19	C-C motif chemokine ligand 19(CCL19)
LDLR	low density lipoprotein receptor(LDLR)
DTD2	D-aminoacyl-tRNA deacylase 2(DTD2)
EGLN2	egl-9 family hypoxia inducible factor 2(EGLN2)
UTP4	UTP4 small subunit processome component(UTP4)
MGMT	O-6-methylguanine-DNA methyltransferase(MGMT)
RIOK3	RIO kinase 3(RIOK3)
UTP3	UTP3 small subunit processome component(UTP3)
GSS	glutathione synthetase(GSS)
GATC	glutamyl-tRNA amidotransferase subunit C(GATC)
WWP1	WW domain containing E3 ubiquitin protein ligase 1(WWP1)
SAMD4B	sterile alpha motif domain containing 4B(SAMD4B)
NMNAT3	nicotinamide nucleotide adenylyltransferase 3(NMNAT3)
ELP5	elongator acetyltransferase complex subunit 5(ELP5)
SLC8B1	solute carrier family 8 member B1(SLC8B1)
PSMB8	proteasome 20S subunit beta 8(PSMB8)
LRPPRC	leucine rich pentatricopeptide repeat containing(LRPPRC)
DHRS7	dehydrogenase/reductase 7(DHRS7)
LYRM4	LYR motif containing 4(LYRM4)
UBAP1	ubiquitin associated protein 1(UBAP1)

TSR1	TSR1 ribosome maturation factor(TSR1)
COQ8B	coenzyme Q8B(COQ8B)
GZF1	GDNF inducible zinc finger protein 1(GZF1)
MTFMT	mitochondrial methionyl-tRNA formyltransferase(MTFMT)
POGLUT2	protein O-glucosyltransferase 2(POGLUT2)
SLC27A1	solute carrier family 27 member 1(SLC27A1)
TXNL4B	thioredoxin like 4B(TXNL4B)
CEP19	centrosomal protein 19(CEP19)
STK19	serine/threonine kinase 19(STK19)
PLOD1	procollagen-lysine,2-oxoglutarate 5-dioxygenase 1(PLOD1)
TM7SF3	transmembrane 7 superfamily member 3(TM7SF3)
FAM204A	family with sequence similarity 204 member A(FAM204A)
PYCARD	PYD and CARD domain containing(PYCARD)
MTRF1	mitochondrial translation release factor 1(MTRF1)
POFUT2	protein O-fucosyltransferase 2(POFUT2)
RUVBL2	RuvB like AAA ATPase 2(RUVBL2)
GRB10	growth factor receptor bound protein 10(GRB10)
NEPRO	nucleolus and neural progenitor protein(NEPRO)
LZTS3	leucine zipper tumor suppressor family member 3(LZTS3)
PNISR	PNN interacting serine and arginine rich protein(PNISR)
HARBI1	harbinger transposase derived 1(HARBI1)
VPS33A	VPS33A core subunit of CORVET and HOPS complexes(VPS33A)
THOC3	THO complex subunit 3(THOC3)
TEX2	testis expressed 2(TEX2)
LEMD2	LEM domain nuclear envelope protein 2(LEMD2)
THNSL2	threonine synthase like 2(THNSL2)
FTSJ3	FtsJ RNA 2'-O-methyltransferase 3(FTSJ3)
FAM118B	family with sequence similarity 118 member B(FAM118B)
THOC6	THO complex subunit 6(THOC6)
ACLY	ATP citrate lyase(ACLY)
FRMD8	FERM domain containing 8(FRMD8)
NAAA	N-acylethanolamine acid amidase(NAAA)
FKBP8	FKBP prolyl isomerase 8(FKBP8)
MAPKAPK5	MAPK activated protein kinase 5(MAPKAPK5)
MINDY3	MINDY lysine 48 deubiquitinase 3(MINDY3)
RNASEH2A	ribonuclease H2 subunit A(RNASEH2A)
LY96	lymphocyte antigen 96(LY96)
MRI1	methylthioribose-1-phosphate isomerase 1(MRI1)
RCC1L	RCC1 like(RCC1L)
AMDHD2	amidohydrolase domain containing 2(AMDHD2)
KDSR	3-ketodihydrosphingosine reductase(KDSR)
NADK	NAD kinase(NADK)
OSBPL5	oxysterol binding protein like 5(OSBPL5)
TCEANC	transcription elongation factor A N-terminal and central domain containing(TCEANC)
GLMN	glomulin, FKBP associated protein(GLMN)
GLMP	glycosylated lysosomal membrane protein(GLMP)
FUZ	fuzzy planar cell polarity protein(FUZ)
ZC3HC1	zinc finger C3HC-type containing 1(ZC3HC1)
STK25	serine/threonine kinase 25(STK25)
CPD	carboxypeptidase D(CPD)
CNOT3	CCR4-NOT transcription complex subunit 3(CNOT3)
RIOX2	ribosomal oxygenase 2(RIOX2)
PIR	pirin(PIR)
ZNF777	zinc finger protein 777(ZNF777)
CLPTM1L	CLPTM1 like(CLPTM1L)
LZTR1	leucine zipper like transcription regulator 1(LZTR1)
MCM7	minichromosome maintenance complex component 7(MCM7)
WDR1	WD repeat domain 1(WDR1)
POMT1	protein O-mannosyltransferase 1(POMT1)
ECE1	endothelin converting enzyme 1(ECE1)
DUSP18	dual specificity phosphatase 18(DUSP18)

NAGPA	N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase(NAGPA)
DUSP16	dual specificity phosphatase 16(DUSP16)
MTO1	mitochondrial tRNA translation optimization 1(MTO1)
DUSP12	dual specificity phosphatase 12(DUSP12)
SPR	sepiapterin reductase(SPR)
OPA3	outer mitochondrial membrane lipid metabolism regulator OPA3(OPA3)
SCRN2	secernin 2(SCRN2)
LBR	lamin B receptor(LBR)
SNAP29	synaptosome associated protein 29(SNAP29)
ITGA3	integrin subunit alpha 3(ITGA3)
PLEKHG5	pleckstrin homology and RhoGEF domain containing G5(PLEKHG5)
SMIM19	small integral membrane protein 19(SMIM19)
PLPBP	pyridoxal phosphate binding protein(PLPBP)
DNAJC3	DnaJ heat shock protein family (Hsp40) member C3(DNAJC3)
WRAP73	WD repeat containing, antisense to TP73(WRAP73)
METTL22	methyltransferase 22, Kin17 lysine(METTL22)
SARS2	seryl-tRNA synthetase 2, mitochondrial(SARS2)
PECAM1	platelet and endothelial cell adhesion molecule 1(PECAM1)
MTRF1L	mitochondrial translation release factor 1 like(MTRF1L)
SMPDL3A	sphingomyelin phosphodiesterase acid like 3A(SMPDL3A)
BCAT2	branched chain amino acid transaminase 2(BCAT2)
NDUFB8	NADH:ubiquinone oxidoreductase subunit B8(NDUFB8)
FRRS1	ferric chelate reductase 1(FRRS1)
CARNMT1	carnosine N-methyltransferase 1(CARNMT1)
ATP10D	ATPase phospholipid transporting 10D (putative)(ATP10D)
MAK16	MAK16 homolog(MAK16)
LRP5	LDL receptor related protein 5(LRP5)
LENG8	leukocyte receptor cluster member 8(LENG8)
SEMA3F	semaphorin 3F(SEMA3F)
MALT1	MALT1 paracaspase(MALT1)
TMA16	translation machinery associated 16 homolog(TMA16)
NCLN	nicalin(NCLN)
LRRC8D	leucine rich repeat containing 8 VRAC subunit D(LRRC8D)
LRRC8A	leucine rich repeat containing 8 VRAC subunit A(LRRC8A)
RFFL	ring finger and FYVE like domain containing E3 ubiquitin protein ligase(RFFL)
CNNM3	cyclin and CBS domain divalent metal cation transport mediator 3(CNNM3)
SH2B1	SH2B adaptor protein 1(SH2B1)
SLC35A1	solute carrier family 35 member A1(SLC35A1)
SLC35A4	solute carrier family 35 member A4(SLC35A4)
FEM1A	fem-1 homolog A(FEM1A)
KRR1	KRR1 small subunit processome component homolog(KRR1)
RNF26	ring finger protein 26(RNF26)
ERAP1	endoplasmic reticulum aminopeptidase 1(ERAP1)
NAGA	alpha-N-acetylgalactosaminidase(NAGA)
IMMT	inner membrane mitochondrial protein(IMMT)
MCOLN1	mucolipin TRP cation channel 1(MCOLN1)
GFER	growth factor, augmenter of liver regeneration(GFER)
CPPED1	calcineurin like phosphoesterase domain containing 1(CPPED1)
MYO1E	myosin IE(MYO1E)
NECAP1	NECAP endocytosis associated 1(NECAP1)
MTRES1	mitochondrial transcription rescue factor 1(MTRES1)
VAPB	VAMP associated protein B and C(VAPB)
POLR3D	RNA polymerase III subunit D(POLR3D)
SLC46A3	solute carrier family 46 member 3(SLC46A3)
SLC35B2	solute carrier family 35 member B2(SLC35B2)
TP53RK	TP53 regulating kinase(TP53RK)
CCNH	cyclin H(CCNH)
ZDHHC6	zinc finger DHHC-type palmitoyltransferase 6(ZDHHC6)
NADSYN1	NAD synthetase 1(NADSYN1)
NUBP1	NUBP iron-sulfur cluster assembly factor 1, cytosolic(NUBP1)
NUBP2	NUBP iron-sulfur cluster assembly factor 2, cytosolic(NUBP2)

PNP	purine nucleoside phosphorylase(PNP)
HYKK	hydroxylysine kinase(HYKK)
ADGRA2	adhesion G protein-coupled receptor A2(ADGRA2)
ANPEP	alanyl aminopeptidase, membrane(ANPEP)
DAG1	dystroglycan 1(DAG1)
EIF2D	eukaryotic translation initiation factor 2D(EIF2D)
ACAD11	acyl-CoA dehydrogenase family member 11(ACAD11)
ACAD10	acyl-CoA dehydrogenase family member 10(ACAD10)
CCDC93	coiled-coil domain containing 93(CCDC93)
ARL6IP6	ADP ribosylation factor like GTPase 6 interacting protein 6(ARL6IP6)
TRIM23	tripartite motif containing 23(TRIM23)
COG7	component of oligomeric golgi complex 7(COG7)
TANGO2	transport and golgi organization 2 homolog(TANGO2)
NFKBIL1	NFKB inhibitor like 1(NFKBIL1)
COG4	component of oligomeric golgi complex 4(COG4)
ALG2	ALG2 alpha-1,3/1,6-mannosyltransferase(ALG2)
COG2	component of oligomeric golgi complex 2(COG2)
ALG3	ALG3 alpha-1,3- mannosyltransferase(ALG3)
COG1	component of oligomeric golgi complex 1(COG1)
ACSL3	acyl-CoA synthetase long chain family member 3(ACSL3)
JMJD8	jumonji domain containing 8(JMJD8)
AFG3L2	AFG3 like matrix AAA peptidase subunit 2(AFG3L2)
CFDP1	craniofacial development protein 1(CFDP1)
BTD	biotinidase(BTD)
PPA2	inorganic pyrophosphatase 2(PPA2)
PDE12	phosphodiesterase 12(PDE12)
TRIM14	tripartite motif containing 14(TRIM14)
PREB	prolactin regulatory element binding(PREB)
NEIL1	nei like DNA glycosylase 1(NEIL1)
PIK3R4	phosphoinositide-3-kinase regulatory subunit 4(PIK3R4)
RTF2	replication termination factor 2(RTF2)
TWF1	twinfilin actin binding protein 1(TWF1)
CD99L2	CD99 molecule like 2(CD99L2)
TRMT2A	tRNA methyltransferase 2 homolog A(TRMT2A)
LIAS	lipoic acid synthetase(LIAS)
ZMPSTE24	zinc metallopeptidase STE24(ZMPSTE24)
PGS1	phosphatidylglycerophosphate synthase 1(PGS1)
INPP5B	inositol polyphosphate-5-phosphatase B(INPP5B)
PRRC1	proline rich coiled-coil 1(PRRC1)
AP5M1	adaptor related protein complex 5 subunit mu 1(AP5M1)
DHX33	DEAH-box helicase 33(DHX33)
DPH2	diphthamide biosynthesis 2(DPH2)
ATP6V0A2	ATPase H+ transporting V0 subunit a2(ATP6V0A2)
DPH5	diphthamide biosynthesis 5(DPH5)
SYMPK	symplekin scaffold protein(SYMPK)
AAR2	AAR2 splicing factor(AAR2)
BBS2	Bardet-Biedl syndrome 2(BBS2)
BBS1	Bardet-Biedl syndrome 1(BBS1)
IQSEC1	IQ motif and Sec7 domain ArfGEF 1(IQSEC1)
SIAH2	siah E3 ubiquitin protein ligase 2(SIAH2)
REX1BD	required for excision 1-B domain containing(REX1BD)
PARP10	poly(ADP-ribose) polymerase family member 10(PARP10)
PARVG	parvin gamma(PARVG)
NDUFAF4	NADH:ubiquinone oxidoreductase complex assembly factor 4(NDUFAF4)
NDUFAF5	NADH:ubiquinone oxidoreductase complex assembly factor 5(NDUFAF5)
INTS7	integrator complex subunit 7(INTS7)
TMCC1	transmembrane and coiled-coil domain family 1(TMCC1)
NOL11	nucleolar protein 11(NOL11)
BAP1	BRCA1 associated protein 1(BAP1)
CC2D1B	coiled-coil and C2 domain containing 1B(CC2D1B)
CSRNP2	cysteine and serine rich nuclear protein 2(CSRNP2)

HSCB	HscB mitochondrial iron-sulfur cluster cochaperone(HSCB)
AP2A2	adaptor related protein complex 2 subunit alpha 2(AP2A2)
PTPRF	protein tyrosine phosphatase receptor type F(PTPRF)
THTPA	thiamine triphosphatase(THTPA)
IMMP2L	inner mitochondrial membrane peptidase subunit 2(IMMP2L)
DHX16	DEAH-box helicase 16(DHX16)
NANS	N-acetylneuraminate synthase(NANS)
TSC22D4	TSC22 domain family member 4(TSC22D4)
ZHX3	zinc fingers and homeoboxes 3(ZHX3)
PARP2	poly(ADP-ribose) polymerase 2(PARP2)
GAA	alpha glucosidase(GAA)
AGTRAP	angiotensin II receptor associated protein(AGTRAP)
MRPS5	mitochondrial ribosomal protein S5(MRPS5)
RPUSD4	RNA pseudouridine synthase D4(RPUSD4)
WAC	WW domain containing adaptor with coiled-coil(WAC)
CRELD1	cysteine rich with EGF like domains 1(CRELD1)
KPTN	kaptin, actin binding protein(KPTN)
MRRF	mitochondrial ribosome recycling factor(MRRF)
MRPS34	mitochondrial ribosomal protein S34(MRPS34)
GDPD3	glycerophosphodiester phosphodiesterase domain containing 3(GDPD3)
ARVCF	ARVCF delta catenin family member(ARVCF)
EXOSC4	exosome component 4(EXOSC4)
METTL8	methyltransferase 8, methylcytidine(METTL8)
EXOSC3	exosome component 3(EXOSC3)
RFTN1	raftlin, lipid raft linker 1(RFTN1)
RPA2	replication protein A2(RPA2)
VEGFC	vascular endothelial growth factor C(VEGFC)
PEX10	peroxisomal biogenesis factor 10(PEX10)
BOP1	BOP1 ribosomal biogenesis factor(BOP1)
NUDT8	nudix hydrolase 8(NUDT8)
TTC4	tetratricopeptide repeat domain 4(TTC4)
SGSM3	small G protein signaling modulator 3(SGSM3)
R3HCC1	R3H domain and coiled-coil containing 1(R3HCC1)
F2RL2	coagulation factor II thrombin receptor like 2(F2RL2)
WASHC5	WASH complex subunit 5(WASHC5)
GDA	guanine deaminase(GDA)
TATDN1	TatD DNase domain containing 1(TATDN1)
TBC1D31	TBC1 domain family member 31(TBC1D31)
CAPN7	calpain 7(CAPN7)
RGS1	regulator of G protein signaling 1(RGS1)
TRAPPC12	trafficking protein particle complex subunit 12(TRAPPC12)
TRAPPC11	trafficking protein particle complex subunit 11(TRAPPC11)
SIRPA	signal regulatory protein alpha(SIRPA)
TMOD3	tropomodulin 3(TMOD3)
PPP4R3A	protein phosphatase 4 regulatory subunit 3A(PPP4R3A)
ACYP2	acylphosphatase 2(ACYP2)
LIG3	DNA ligase 3(LIG3)
TMX2	thioredoxin related transmembrane protein 2(TMX2)
SELENOS	selenoprotein S(SELENOS)
HNRNPUL2	heterogeneous nuclear ribonucleoprotein U like 2(HNRNPUL2)
ILVBL	ilvB acetolactate synthase like(ILVBL)
FAM185A	family with sequence similarity 185 member A(FAM185A)
SPNS1	SPNS lysolipid transporter 1, lysophospholipid(SPNS1)
PPIE	peptidylprolyl isomerase E(PPIE)
GEMIN6	gem nuclear organelle associated protein 6(GEMIN6)
GEMIN7	gem nuclear organelle associated protein 7(GEMIN7)
PXK	PX domain containing serine/threonine kinase like(PXK)
UBA7	ubiquitin like modifier activating enzyme 7(UBA7)
ZCCHC9	zinc finger CCHC-type containing 9(ZCCHC9)
PRPF8	pre-mRNA processing factor 8(PRPF8)
ACACB	acetyl-CoA carboxylase beta(ACACB)

NOP9	NOP9 nucleolar protein(NOP9)
MYG1	MYG1 exonuclease(MYG1)
ZC3H18	zinc finger CCCH-type containing 18(ZC3H18)
MORC2	MORC family CW-type zinc finger 2(MORC2)
ANKZF1	ankyrin repeat and zinc finger peptidyl tRNA hydrolase 1(ANKZF1)
RHBDD3	rhomboid domain containing 3(RHBDD3)
UQCC1	ubiquinol-cytochrome c reductase complex assembly factor 1(UQCC1)
DNAJC11	DnaJ heat shock protein family (Hsp40) member C11(DNAJC11)
KLHL22	kelch like family member 22(KLHL22)
TJAP1	tight junction associated protein 1(TJAP1)
NMRAL1	NmrA like redox sensor 1(NMRAL1)
TMEM115	transmembrane protein 115(TMEM115)
TMEM237	transmembrane protein 237(TMEM237)
MAN2B2	mannosidase alpha class 2B member 2(MAN2B2)
RSRP1	arginine and serine rich protein 1(RSRP1)
PXMP4	peroxisomal membrane protein 4(PXMP4)
GCLM	glutamate-cysteine ligase modifier subunit(GCLM)
GALK2	galactokinase 2(GALK2)
GALK1	galactokinase 1(GALK1)
NGLY1	N-glycanase 1(NGLY1)
CHD3	chromodomain helicase DNA binding protein 3(CHD3)
USE1	unconventional SNARE in the ER 1(USE1)
GHR	growth hormone receptor(GHR)
TCHP	trichoplein keratin filament binding(TCHP)
ABHD16A	abhydrolase domain containing 16A, phospholipase(ABHD16A)
CPT2	carnitine palmitoyltransferase 2(CPT2)
MFSD5	major facilitator superfamily domain containing 5(MFSD5)
TOP1MT	DNA topoisomerase I mitochondrial(TOP1MT)
MFSD3	major facilitator superfamily domain containing 3(MFSD3)
POLI	DNA polymerase iota(POLI)
VCPKMT	valosin containing protein lysine methyltransferase(VCPKMT)
SUN2	Sad1 and UNC84 domain containing 2(SUN2)
RBM19	RNA binding motif protein 19(RBM19)
SPATS2	spermatogenesis associated serine rich 2(SPATS2)
TBC1D8	TBC1 domain family member 8(TBC1D8)
ABHD14B	abhydrolase domain containing 14B(ABHD14B)
QRSL1	glutaminyl-tRNA amidotransferase subunit QRSL1(QRSL1)
POMGNT2	protein O-linked mannose N-acetylglucosaminyltransferase 2 (beta 1,4-)(POMGNT2)
TTI2	TELO2 interacting protein 2(TTI2)
OXLD1	oxidoreductase like domain containing 1(OXLD1)
ELF1	E74 like ETS transcription factor 1(ELF1)
ELF2	E74 like ETS transcription factor 2(ELF2)
ATG4B	autophagy related 4B cysteine peptidase(ATG4B)
USP40	ubiquitin specific peptidase 40(USP40)
TSEN34	tRNA splicing endonuclease subunit 34(TSEN34)
DCTN5	dynactin subunit 5(DCTN5)
DDX27	DEAD-box helicase 27(DDX27)
DDX24	DEAD-box helicase 24(DDX24)
POLDIP3	DNA polymerase delta interacting protein 3(POLDIP3)
ABRAXAS1	abraxas 1, BRCA1 A complex subunit(ABRAXAS1)
ANKS3	ankyrin repeat and sterile alpha motif domain containing 3(ANKS3)
PON2	paraoxonase 2(PON2)
PNKP	polynucleotide kinase 3'-phosphatase(PNKP)
RNF8	ring finger protein 8(RNF8)
RRBP1	ribosome binding protein 1(RRBP1)
PPOX	protoporphyrinogen oxidase(PPOX)
TBRG4	transforming growth factor beta regulator 4(TBRG4)
TBRG1	transforming growth factor beta regulator 1(TBRG1)
TAF1D	TATA-box binding protein associated factor, RNA polymerase I subunit D(TAF1D)
FAM168A	family with sequence similarity 168 member A(FAM168A)
SH3TC1	SH3 domain and tetratricopeptide repeats 1(SH3TC1)

HES1	hes family bHLH transcription factor 1(HES1)
GBP2	guanylate binding protein 2(GBP2)
LURAP1	leucine rich adaptor protein 1(LURAP1)
ABCF1	ATP binding cassette subfamily F member 1(ABCF1)
ATP13A1	ATPase 13A1(ATP13A1)
LIMK2	LIM domain kinase 2(LIMK2)
IRGQ	immunity related GTPase Q(IRGQ)
NEURL4	neuralized E3 ubiquitin protein ligase 4(NEURL4)
ERCC3	ERCC excision repair 3, TFIIH core complex helicase subunit(ERCC3)
TEF	TEF transcription factor, PAR bZIP family member(TEF)
KANSL2	KAT8 regulatory NSL complex subunit 2(KANSL2)
ASB8	ankyrin repeat and SOCS box containing 8(ASB8)
SIMC1	SUMO interacting motifs containing 1(SIMC1)
SUGT1	SGT1 homolog, MIS12 kinetochore complex assembly cochaperone(SUGT1)
NRDE2	NRDE-2, necessary for RNA interference, domain containing(NRDE2)
CPVL	carboxypeptidase vitellogenic like(CPVL)
MBTPS2	membrane bound transcription factor peptidase, site 2(MBTPS2)
SNTB1	syntrophin beta 1(SNTB1)
ZPR1	ZPR1 zinc finger(ZPR1)

Appendix C4 Table 3: Gene ontology (GO) enrichment analysis result on the under selection within western Australian population discovered using **HYPHY Fubar**. The analysis approximates the true distribution of numbers of members of a category amongst selected genes by the Wallenius non-central hypergeometric distribution. False Discovery Rate (FDR) is then calculated using Benjamini and Hochberg method.

GO category	Number of genes	Number in category	Term	Ontology	P-value
GO:0046395	16	125	carboxylic acid catabolic process	BP	3.83E-08
GO:0016054	16	127	organic acid catabolic process	BP	3.83E-08
GO:0044282	18	183	small molecule catabolic process	BP	8.45E-08
GO:0019752	27	457	carboxylic acid metabolic process	BP	8.45E-08
GO:0006082	27	463	organic acid metabolic process	BP	8.45E-08
GO:0043436	27	463	oxoacid metabolic process	BP	8.45E-08
GO:0016491	25	433	oxidoreductase activity	MF	6.22E-07
GO:0003824	85	3822	catalytic activity	MF	1.31E-06
GO:0006629	30	703	lipid metabolic process	BP	7.91E-06
GO:0009063	10	64	cellular amino acid catabolic process	BP	1.49E-05
GO:0006520	14	161	cellular amino acid metabolic process	BP	3.19E-05
GO:0050660	9	54	flavin adenine dinucleotide binding	MF	3.19E-05
GO:0016627	7	34	oxidoreductase activity, acting on the CH-CH group of donors	MF	0.0002845
GO:0044255	23	542	cellular lipid metabolic process	BP	0.000493786
GO:1901605	11	120	alpha-amino acid metabolic process	BP	0.000493786
GO:0016829	11	142	lyase activity	MF	0.002610915
GO:1901606	7	56	alpha-amino acid catabolic process	BP	0.007951468
GO:0019842	8	87	vitamin binding	MF	0.016176918

Appendix C5 Table 4: List of positively selected genes among western Australian brushtail possums discovered using **HYPHY Meme**.

Gene	Gene Name
CSRNP2	cysteine and serine rich nuclear protein 2(CSRNP2)
MOCS2	molybdenum cofactor synthesis 2(MOCS2)
TXNL4B	thioredoxin like 4B(TXNL4B)
PIGO	phosphatidylinositol glycan anchor biosynthesis class O(PIGO)
CCNH	cyclin H(CCNH)
POMT1	protein O-mannosyltransferase 1(POMT1)
ECE1	endothelin converting enzyme 1(ECE1)
PLOD1	procollagen-lysine,2-oxoglutarate 5-dioxygenase 1(PLOD1)

CHD3	chromodomain helicase DNA binding protein 3(CHD3)
WDR46	WD repeat domain 46(WDR46)
TMEM186	transmembrane protein 186(TMEM186)
MED17	mediator complex subunit 17(MED17)
SMPD5	sphingomyelin phosphodiesterase 5(SMPD5)
IMMP2L	inner mitochondrial membrane peptidase subunit 2(IMMP2L)
TATDN1	TatD DNase domain containing 1(TATDN1)
RNF114	ring finger protein 114(RNF114)
PSMD5	proteasome 26S subunit, non-ATPase 5(PSMD5)
MFSD5	major facilitator superfamily domain containing 5(MFSD5)
RUVBL2	RuvB like AAA ATPase 2(RUVBL2)
ANPEP	alanyl aminopeptidase, membrane(ANPEP)
ZC3H12A	zinc finger CCCH-type containing 12A(ZC3H12A)
GRB10	growth factor receptor bound protein 10(GRB10)
DAG1	dystroglycan 1(DAG1)
DHX16	DEAH-box helicase 16(DHX16)
SIRPA	signal regulatory protein alpha(SIRPA)
TRAPPC11	trafficking protein particle complex subunit 11(TRAPPC11)
EIF2D	eukaryotic translation initiation factor 2D(EIF2D)
IER5	immediate early response 5(IER5)
SNAP29	synaptosome associated protein 29(SNAP29)
TSC22D4	TSC22 domain family member 4(TSC22D4)
IAH1	isoamyl acetate hydrolyzing esterase 1 (putative)(IAH1)
MAP1S	microtubule associated protein 1S(MAP1S)
NFKBIL1	NFKB inhibitor like 1(NFKBIL1)
COG2	component of oligomeric golgi complex 2(COG2)
TBC1D8	TBC1 domain family member 8(TBC1D8)
AGTRAP	angiotensin II receptor associated protein(AGTRAP)
COG1	component of oligomeric golgi complex 1(COG1)
HDHD3	haloacid dehalogenase like hydrolase domain containing 3(HDHD3)
MRPS5	mitochondrial ribosomal protein S5(MRPS5)
THOC6	THO complex subunit 6(THOC6)
SIRT3	sirtuin 3(SIRT3)
METTL22	methyltransferase 22, Kin17 lysine(METTL22)
CFDP1	craniofacial development protein 1(CFDP1)
SELENOS	selenoprotein S(SELENOS)
PPA2	inorganic pyrophosphatase 2(PPA2)
SPNS1	SPNS lysolipid transporter 1, lysophospholipid(SPNS1)
FKBP8	FKBP prolyl isomerase 8(FKBP8)
PPIE	peptidylprolyl isomerase E(PPIE)
GEMIN6	gem nuclear organelle associated protein 6(GEMIN6)
PREB	prolactin regulatory element binding(PREB)
MOCS1	molybdenum cofactor synthesis 1(MOCS1)
PIGH	phosphatidylinositol glycan anchor biosynthesis class H(PIGH)
UBA7	ubiquitin like modifier activating enzyme 7(UBA7)
MAK16	MAK16 homolog(MAK16)
LY96	lymphocyte antigen 96(LY96)
NOLC1	nucleolar and coiled-body phosphoprotein 1(NOLC1)
PPOX	protoporphyrinogen oxidase(PPOX)
UNG	uracil DNA glycosylase(UNG)
TBRG4	transforming growth factor beta regulator 4(TBRG4)
RCC1L	RCC1 like(RCC1L)
PGS1	phosphatidylglycerophosphate synthase 1(PGS1)
MYG1	MYG1 exonuclease(MYG1)
NCLN	nicalin(NCLN)
PRRC1	proline rich coiled-coil 1(PRRC1)
DHX33	DEAH-box helicase 33(DHX33)
SH3TC1	SH3 domain and tetratricopeptide repeats 1(SH3TC1)
LRRC8D	leucine rich repeat containing 8 VRAC subunit D(LRRC8D)
RFFL	ring finger and FYVE like domain containing E3 ubiquitin protein ligase(RFFL)
ANKZF1	ankyrin repeat and zinc finger peptidyl tRNA hydrolase 1(ANKZF1)

FEM1A	fem-1 homolog A(FEM1A)
GLMP	glycosylated lysosomal membrane protein(GLMP)
TRIP4	thyroid hormone receptor interactor 4(TRIP4)
SAMD4B	sterile alpha motif domain containing 4B(SAMD4B)
RPA2	replication protein A2(RPA2)
MTHFS	methenyltetrahydrofolate synthetase(MTHFS)
NMNAT3	nicotinamide nucleotide adenylyltransferase 3(NMNAT3)
MCOLN1	mucolipin TRP cation channel 1(MCOLN1)
NMRAL1	NmrA like redox sensor 1(NMRAL1)
TJAP1	tight junction associated protein 1(TJAP1)
ZC3HC1	zinc finger C3HC-type containing 1(ZC3HC1)
BOP1	BOP1 ribosomal biogenesis factor(BOP1)
NEURL4	neuralized E3 ubiquitin protein ligase 4(NEURL4)
LYRM4	LYR motif containing 4(LYRM4)
PARVG	parvin gamma(PARVG)
MAN2B2	mannosidase alpha class 2B member 2(MAN2B2)
NDUFAF4	NADH:ubiquinone oxidoreductase complex assembly factor 4(NDUFAF4)
NDUFAF5	NADH:ubiquinone oxidoreductase complex assembly factor 5(NDUFAF5)
RSRP1	arginine and serine rich protein 1(RSRP1)
PNPLA6	patatin like phospholipase domain containing 6(PNPLA6)
GCLM	glutamate-cysteine ligase modifier subunit(GCLM)

Appendix C5 Table 5: List of genes showing significantly high Gst and Fst value between western Australian and New Zealand brushtail possums.

Gene	Gene Name
MUS81	MUS81 structure-specific endonuclease subunit(MUS81)
STARD3	StAR related lipid transfer domain containing 3(STARD3)
RHBDD3	rhomboid domain containing 3(RHBDD3)
BAD	BCL2 associated agonist of cell death(BAD)
STK11IP	serine/threonine kinase 11 interacting protein(STK11IP)
DDX55	DEAD-box helicase 55(DDX55)
GTF2H3	general transcription factor IIH subunit 3(GTF2H3)
PARP12	poly(ADP-ribose) polymerase family member 12(PARP12)
UNG	uracil DNA glycosylase(UNG)
DPP3	dipeptidyl peptidase 3(DPP3)
NUDT22	nudix hydrolase 22(NUDT22)
PLCB3	phospholipase C beta 3(PLCB3)
AFTPH	aftiphilin(AFTPH)
DPP7	dipeptidyl peptidase 7(DPP7)
VPS72	vacuolar protein sorting 72 homolog(VPS72)
LDHD	lactate dehydrogenase D(LDHD)
ASCC2	activating signal cointegrator 1 complex subunit 2(ASCC2)
CSK	C-terminal Src kinase(CSK)
NELFCD	negative elongation factor complex member C/D(NELFCD)
EIF2B1	eukaryotic translation initiation factor 2B subunit alpha(EIF2B1)
RBM45	RNA binding motif protein 45(RBM45)
MTPAP	mitochondrial poly(A) polymerase(MTPAP)
UTP14A	UTP14A small subunit processome component(UTP14A)
MON1B	MON1 homolog B, secretory trafficking associated(MON1B)

Appendix C5 Table 6: List of genes carrying SNPs associated with western Australian brushtail possums discovered using **rTassel**.

Gene	Gene Name
PIGS	phosphatidylinositol glycan anchor biosynthesis class S(PIGS)
TOP2B	DNA topoisomerase II beta(TOP2B)

MOCS2	molybdenum cofactor synthesis 2(MOCS2)
TES	testin LIM domain protein(TES)
ITGB5	integrin subunit beta 5(ITGB5)
HP1BP3	heterochromatin protein 1 binding protein 3(HP1BP3)
LRRC3	leucine rich repeat containing 3(LRRC3)
MRPL39	mitochondrial ribosomal protein L39(MRPL39)
PCSK7	proprotein convertase subtilisin/kexin type 7(PCSK7)
MTO1	mitochondrial tRNA translation optimization 1(MTO1)
CCDC28A	coiled-coil domain containing 28A(CCDC28A)
HEBP2	heme binding protein 2(HEBP2)
NUDCD3	NudC domain containing 3(NUDCD3)
HERC1	HECT and RLD domain containing E3 ubiquitin protein ligase family member 1(HERC1)
VPS72	vacuolar protein sorting 72 homolog(VPS72)
ARHGDIB	Rho GDP dissociation inhibitor beta(ARHGDIB)
NFKBIZ	NFKB inhibitor zeta(NFKBIZ)
LONP2	lon peptidase 2, peroxisomal(LONP2)
HMG20B	high mobility group 20B(HMG20B)
GUSB	glucuronidase beta(GUSB)
SPG11	SPG11 vesicle trafficking associated, spatacsin(SPG11)
VPS18	VPS18 core subunit of CORVET and HOPS complexes(VPS18)
C1H1ORF50	chromosome 1 C1orf50 homolog(C1H1orf50)
PLPBP	pyridoxal phosphate binding protein(PLPBP)
CASD1	CAS1 domain containing 1(CASD1)
FBXO11	F-box protein 11(FBXO11)
MED28	mediator complex subunit 28(MED28)
MSH6	mutS homolog 6(MSH6)
TMEM135	transmembrane protein 135(TMEM135)
PSPC1	paraspeckle component 1(PSPC1)
ADORA2B	adenosine A2b receptor(ADORA2B)
GRWD1	glutamate rich WD repeat containing 1(GRWD1)
B3GNT2	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 2(B3GNT2)
KDM7A	lysine demethylase 7A(KDM7A)
AHNAK	AHNAK nucleoprotein(AHNAK)
TUBGCP2	tubulin gamma complex associated protein 2(TUBGCP2)
POLDIP3	DNA polymerase delta interacting protein 3(POLDIP3)
FLII	FLII actin remodeling protein(FLII)
NMD3	NMD3 ribosome export adaptor(NMD3)
MIEF1	mitochondrial elongation factor 1(MIEF1)
SPG7	SPG7 matrix AAA peptidase subunit, paraplegin(SPG7)
PPM1D	protein phosphatase, Mg2+/Mn2+ dependent 1D(PPM1D)
USP19	ubiquitin specific peptidase 19(USP19)
UNG	uracil DNA glycosylase(UNG)
CAMKK2	calcium/calmodulin dependent protein kinase kinase 2(CAMKK2)
RASGRP3	RAS guanyl releasing protein 3(RASGRP3)
CYB5R1	cytochrome b5 reductase 1(CYB5R1)
MAN2A1	mannosidase alpha class 2A member 1(MAN2A1)
HSF2	heat shock transcription factor 2(HSF2)
CHST15	carbohydrate sulfotransferase 15(CHST15)
LEO1	LEO1 homolog, Paf1/RNA polymerase II complex component(LEO1)
RRS1	ribosome biogenesis regulator 1 homolog(RRS1)
TNRC18	trinucleotide repeat containing 18(TNRC18)
ASCC2	activating signal cointegrator 1 complex subunit 2(ASCC2)
CCL19	C-C motif chemokine ligand 19(CCL19)
ADSS	adenylosuccinate synthase 2(ADSS2)
DEDD	death effector domain containing(DEDD)
ATP6V0A1	ATPase H+ transporting V0 subunit a1(ATP6V0A1)
EDEM3	ER degradation enhancing alpha-mannosidase like protein 3(EDEM3)
DYNC1H1	dynein cytoplasmic 1 heavy chain 1(DYNC1H1)
MBTPS1	membrane bound transcription factor peptidase, site 1(MBTPS1)
CPSF7	cleavage and polyadenylation specific factor 7(CPSF7)
MGMT	O-6-methylguanine-DNA methyltransferase(MGMT)

LAMB2	laminin subunit beta 2(LAMB2)
GALNT1	polypeptide N-acetylgalactosaminyltransferase 1(GALNT1)
FBXO38	F-box protein 38(FBXO38)
OGA	O-GlcNAcase(OGA)
CYBRD1	cytochrome b reductase 1(CYBRD1)
KIAA0232	KIAA0232(KIAA0232)
PARP12	poly(ADP-ribose) polymerase family member 12(PARP12)
TM9SF1	transmembrane 9 superfamily member 1(TM9SF1)
XPNPEP1	X-prolyl aminopeptidase 1(XPNPEP1)
GIPC1	GIPC PDZ domain containing family member 1(GIPC1)
DNAJC10	DnaJ heat shock protein family (Hsp40) member C10(DNAJC10)
FAT1	FAT atypical cadherin 1(FAT1)
STT3B	STT3 oligosaccharyltransferase complex catalytic subunit B(STT3B)
FAM98A	family with sequence similarity 98 member A(FAM98A)
TGFBI	transforming growth factor beta induced(TGFBI)
PPIL3	peptidylprolyl isomerase like 3(PPIL3)
RBM45	RNA binding motif protein 45(RBM45)
CCDC174	coiled-coil domain containing 174(CCDC174)

Appendix C4 Table 7: Gene ontology (GO) enrichment analysis result on the genes showing significantly associated SNPs with western Australian population found using **rTassel**. The analysis approximates the true distribution of numbers of members of a category amongst selected genes by the Wallenius non-central hypergeometric distribution. False Discovery Rate (FDR) is then calculated using Benjamini and Hochberg method.

GO category	Number of genes	Number in category	Term	Ontology	P-value
GO:0043231	345	7027	intracellular membrane-bounded organelle	CC	0.00089911
GO:0006396	54	659	RNA processing	BP	0.00099312
GO:0003824	221	4061	catalytic activity	MF	0.00364468
GO:0043227	348	7258	membrane-bounded organelle	CC	0.00364468
GO:0004553	12	57	hydrolase activity, hydrolyzing O-glycosyl compounds	MF	0.0137096
GO:0005622	409	8964	intracellular anatomical structure	CC	0.0137096
GO:0043229	367	7871	intracellular organelle	CC	0.01620235
GO:0030684	11	57	preribosome	CC	0.02188199
GO:0008152	335	7093	metabolic process	BP	0.02188199
GO:0005730	46	604	nucleolus	CC	0.02892671
GO:0032991	185	3622	protein-containing complex	CC	0.02892671
GO:0006378	7	20	mRNA polyadenylation	BP	0.02892671
GO:0043226	372	8061	organelle	CC	0.03079639