# An investigation of two significant infectious diseases in populations of Victorian koalas (*Phascolarctos cinereus*)

Alistair Raymond Legione

ORCID: 0000-0003-2407-4496

Submitted in total fulfilment of the requirements of the degree of

Doctorate of Philosophy

August 2017

Melbourne Veterinary School

Faculty of Veterinary and Agricultural Sciences

The University of Melbourne

# Abstract

The koala (*Phascolarctos cinereus*), an iconic Australian marsupial, is considered a vulnerable species in parts of Australia due to recent rapid population declines. The role of infectious diseases in population declines in northern koalas (New South Wales (NSW) and Queensland populations) has been highly studied.

*Chlamydia pecorum* and koala retrovirus (KoRV) have both been given considerable attention. *C. pecorum* in koalas is associated with infertility and blindness through infection of the urogenital tract and conjunctiva, respectively. The prevalence of *C. pecorum* in northern koalas is as high as 87%. Different genotypes of KoRV have been identified in northern koalas. KoRV-A has been identified in all northern koalas tested to date. KoRV-B appears to be less prevalent but has been implicated as a cause of neoplasia.

Molecular testing of *C. pecorum* and KoRV prevalence in Victoria, a southern population of koalas, has been limited. No genotyping studies have been undertaken on either organism in Victorian koalas. This thesis conducted an extensive survey of Victorian koalas across seven separate regions to establish a prevalence estimate for both *C. pecorum* and KoRV. A genotyping study for each pathogen was also completed.

The estimated prevalence of *C. pecorum* in Victorian koalas was 15.2% (125/820, 95% confidence interval (CI) 12.9, 17.9%). Molecular evidence of *C. pecorum* infection in French Island koalas was detected for the first time. Only a single ocular *C. pecorum* infection was identified in Victorian koalas (1/459). A total of six *C. pecorum* genotypes were detected, the majority of which were genotype B, which has only been detected in southern koalas. Three of the genotypes were novel, each of which were found in distinct populations. Male koalas were more likely to be infected than females. *C. pecorum* infection was associated with ‘wet bottom’ (a sign of urinary incontinence and inflammation) in male koalas and reproductive tract disease in female koalas.

Not all koalas with ‘wet bottom’ had detectable *C. pecorum,* suggesting another potential cause. Analysis of the genetic diversity of the bacteria present in urogenital tract samples from ten koalas, of which only five displayed wet bottom, identified 13 operational taxonomic units that occurred at a higher abundance in wet bottom-affected koalas. These bacterial families are of interest for future studies.

The genomes of 57 *C. pecorum* samples from koalas across Australia were sequenced and assembled. The results showed that *C. pecorum* genomes from southern koalas were distinct from those of northern koalas.

KoRV prevalence in Victorian koalas was 24.7% (160/648, 95% CI 21.5, 28.2%). Only KoRV-A was detected. Koalas with ‘wet bottom’ were almost twice as likely to have KoRV detected. There was no association between KoRV and *C. pecorum* detection.

This research highlights that Victorian koalas are experiencing a reduced burden of infection compared to northern koalas, and this may be a factor in southern populations outgrowing the available habitat resources in Victoria, compared to northern populations. Victorian koalas free from *C. pecorum* and KoRV infection could be sourced from over-abundant populations to assist re-establishment of populations where koalas have become locally extinct. Overall, this research provides valuable information for both future research and koala population management.

## Declaration

This is to certify that

i) The thesis comprises only my original work towards the degree of doctorate of philosophy except where indicated in the Preface,

ii) Due acknowledgement has been made in the text to all other material used,

iii) The thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies, and appendices

**Alistair Legione**

# Preface

All experimental work, data analysis and writing carried out for this PhD was undertaken by Alistair Legione, aside from the following cases. No wildlife samples were collected by Alistair Legione, but provided to him by collaborators working in wildlife management roles. Additionally, all koala health data were provided by veterinarians working on these programs, or by pathologists at the University of Melbourne undertaking post-mortem assessment of koalas. Extraction of DNA, amplification, and next generation sequencing for microbiome samples was undertaken by the Australian Genomics Research Facility, Australia. *C. pecorum* genomic samples were hybridised and sequenced by collaborators at the Wellcome Trust Sanger Institute, Hinxton, United Kingdom.

## Written works

In total, three peer reviewed communications have been published as part of this degree, included in full in the following thesis. The citations, and contribution of each author are listed below. Alistair Legione contributed greater than 50% of the work in all cases.

**Legione AR**, Amery-Gale J, Lynch M, Haynes L, Gilkerson JR, Sansom FM, Devlin JM (2016). "*Chlamydia pecorum* infection in free-ranging koalas (*Phascolarctos cinereus*) on French Island, Victoria, Australia." *Journal of Wildlife Diseases* 52(2) 426-429.

ARL developed experimental design, processed samples, conducted the experiments, analysed the data, and wrote the manuscript, JAG collected the samples and reviewed the manuscript, ML reviewed the manuscript, LH collected the clinical data and reviewed the manuscript, JRG reviewed the manuscript, FMS developed experimental design, revised, and reviewed the manuscript, JMD developed experimental design, revised, and reviewed the manuscript.

**Legione AR**, Patterson JLP, Whiteley PL, Amery-Gale J, Lynch M, Haynes L, Gilkerson JR, Polkinghorne A, Devlin JM, Sansom FM (2016). "Identification of unusual *Chlamydia pecorum* genotypes in Victorian koalas (*Phascolarctos cinereus*) and clinical variables associated with infection." *Journal of Medical Microbiology* 65(5): 420-428.

ARL developed experimental design, processed samples, conducted the experiments, analysed data, and wrote the manuscript. JLSP developed experimental design, collected samples, processed samples, assessed clinical health of koalas, reviewed the manuscript. PW collected samples, assessed clinical health of koalas, and reviewed the manuscript. LH collected samples, assessed clinical health of koalas, and reviewed the manuscript. JAG collected samples, assessed clinical health of koalas, and reviewed the manuscript. ML developed experimental design, collected samples, assessed clinical health of koalas, reviewed the manuscript. AP developed experimental design, and reviewed the manuscript. JRG developed experimental design, reviewed the manuscript. FMS developed experimental design, analysed data, revised and reviewed the manuscript. JMD developed experimental design, analysed data, revised and reviewed the manuscript.

**Legione AR**, Patterson JLS, Whiteley P, Firestone SM, Curnick M, Bodley K, Lynch M, Gilkerson JR, Sansom FM, and Devlin JM (2017). "Koala retrovirus (KoRV) genotyping analyses reveal a low prevalence of KoRV-A in Victorian koalas and an association with clinical disease". *Journal of Medical Microbiology*. 66(2): 236-244

ARL developed experimental design, processed samples, conducted the experiments, analysed data, and wrote the manuscript. JLSP developed experimental design, collected samples, processed samples, assessed clinical health of koalas, reviewed the manuscript. PW collected samples, assessed clinical health of koalas, and reviewed the manuscript. SMF analysed data, revised and reviewed the manuscript. MC collected samples, assessed clinical health of koalas, and reviewed the manuscript. KB collected samples, assessed clinical health of koalas, and reviewed the manuscript. ML developed experimental design, collected samples, assessed clinical health of koalas, and reviewed the manuscript. JRG developed experimental design and reviewed the manuscript. FMS developed experimental design, analysed data, revised and reviewed the manuscript. JMD developed experimental design, analysed data, revised and reviewed the manuscript.

## Project funding

The Holsworth Wildlife Research Endowment, provided a total of $20,200 from 2014 – 2017 for this project. The Vizard Foundation and the Albert George and Nancy Caroline Youngman Trust provided funding for Dr Pam Whiteley's research projects from which samples were derived for this project. Alistair Legione's stipend was provided by the Australian Postgraduate Award. Travel funding to allow collaboration with the University of the Sunshine Coast was provided by the Faculty of Veterinary and Agricultural Science Engagement and Partnerships Visiting Scholars' Program.

# Acknowledgements

I would first like to acknowledge and thank the most important person: my amazing partner, friend, and fiancé Dr Olivia Marshall. You have supported me throughout my PhD, mentally, emotionally, and financially, and you’ve put my pursuits ahead of your own for nearly four years. I am incredibly lucky to have been able to grow with you, and share with you love & laughter, as well as frustrations & anxieties. You have encouraged me and helped me work through everything. I love you with all my being and look forward to a lifetime together, free of the yoke of higher education. You also brought into my life two great gifts, our beloved greyhounds Harvey and Chloe, who never failed to make me smile throughout this journey.

I thank my supervisors for their guidance and support throughout my candidature. In particular, I extend my gratitude to my primary supervisor Dr Fiona Sansom who has offered constant guidance, feedback and, ensured my progress was not delayed, and my co-primary Associate Professor Jo Devlin, who gave me my first job as a scientist and has continued to support my early career in research, in particular encouraging me in my pursuits of collaborations and projects only loosely associated with my PhD. Both have been wonderful supervisors, giving up large amounts of their time to listen to my concerns and help me through various stages of my PhD. They have covered expenses for my publications, and provided travel funding to conferences, allowing me to present my research to an international audience. Professor James Gilkerson and Dr Michael Lynch, who make up my supervisory team, have provided me with continued support, project feedback and encouragement. Dr Lynch also was crucial in providing samples from fieldwork. I also thank Dr Carol Hartley, my advisory committee chair, who has always been willing to offer experimental advice, or simply a sounding board for ideas or questions I have had in the lab.

I thank all my colleagues at the Asia Pacific Centre for Animal Health, who have ensured I’ve been able to conduct my research in an environment I find motivating, scientifically stimulating, enjoyable, and fun. In particular, I would like to thank Dr Mauricio Coppo Diez and Ms Paola Vaz. Mauricio assisted me in organising hundreds of samples when I first began my project, scribing details listed on each tube to make all further lab work run much smoother. Paola provided a continuous stream of project ideas, assistance with genomics, and assisted with the foundations of our wildlife

sample database. I thank Cynthia Brown, Nino Ficorilli and Joanne Allen, for their friendship and all of their early mornings making sure our labs were organised, and ensuring I had any equipment or reagents when I needed them.

My project encompasses a significantly large number of samples collected from koalas as part of prior studies, management projects, or submitted as diagnostic samples. This sample collection, along with the assessment of koala clinical health, was undertaken by a large team of collaborators including the following: Dr Jade Patterson, Dr Pam Whiteley, Dr Jemima Amery-Gale, Dr Kate Bodley, Dr Leesa Haynes, and Dr Megan Curnick. Additionally, Dr Pam Whiteley played a key role in providing samples from koalas undergoing post-mortem assessment. These post-mortems were performed by Dr Whiteley, as well as Dr Barbara Bacci, Dr Samoa Giovannini, and Dr Andrew Stent, all of whom deserve thanks and praise of donating their time and energy towards this and other wildlife research projects.

I extend thanks to all of my research collaborators who offered me their time and advice, as well as those who had a more in-depth role in my project. This includes, but is by no means limited to Associate Professor Adam Polkinghorne, Dr Martina Jelocnik, Dr Brendan Ansell, Professor Mark Stevenson, Dr Simon Firestone, Dr Alyce Taylor-Brown, Dr Neil Young, Dr Kath Handasyde, Associate Professor Damien Higgins, and Dr Iona Maher.

Funding for research aspects of my project was provided by the Holsworth Research Endowment, and for Pam Whiteley's research from which I obtained samples through both the Vizard foundation and the Albert George and Nancy Caroline Youngman Trust. Funding to attend an international conference was provided by the Holsworth Research Endowment, Faculty of Veterinary and Agricultural Sciences (FVAS) student travel scholarship, and the Wildlife Disease Association Student Award. Additionally, I was awarded the Dr Sue Newton Travel Scholarship, and I wish to extend a special thanks to the Newton family for their ongoing support of early career researchers in the field of veterinary science. Funding travel to the University of the Sunshine Coast to undertake collaborative work was provided by the FVAS engagement award. The Australian Postgraduate Award provided my stipend, and I implore future governments to continue to invest in science.

Finally, I would like to thank my family, particularly my parents Julie and Raymond Legione, for their encouragement of my curiosity and love of problem solving. They endowed me with a love of knowledge and encouraged me to work hard towards my passions, whilst never pushing me to a point of distaste in my pursuits. I hope that I've made you proud in my endeavours.

# Table of Contents

Abstract ........ i

Declaration ........ iii

Preface ........ iv

Acknowledgements ........ vii

Table of Contents ........ x

Abbreviations ........ xv

List of Tables ........ xix

List of Figures ........ xx

Third party copyright material ........ xxi

1. Literature Review ........ 1

1.1 The koala, an overview ........ 1

1.2 Management of koala populations ........ 1

1.3 Koalas in Victoria ........ 4

1.4 Impact of disease on wildlife population management ........ 6

1.4.1 Diseases impacting koalas that may affect population management ........ 8

1.5 *Chlamydia pecorum* ........ 9

1.5.1 *C. pecorum* in koalas ........ 11

1.5.2 *C. pecorum* variation in koalas and association with virulence ........ 13

1.6 Retrovirus in koalas ........ 17

1.6.1 What is a retrovirus? ........ 17

1.6.2 KoRV - a history of discovery ........ 18

1.6.3 KoRV endogenisation and epidemiology ........ 20

1.6.4 KoRV and disease in koalas ........ 22

1.6.5 KoRV genotypes and their potential role in disease ........ 24

1.6.6 Transmission of KoRV ........ 26

1.7 Recent research and knowledge gaps ........ 27

1.7.1 *C. pecorum* in Victorian koalas ... 27
1.7.2 KoRV and disease ... 28
1.7.3 Wet bottom ... 28
1.7.4 Bacterial diversity studies to identify aetiological agents ... 29
1.8 Aims and objectives ... 30
2. General Methodology ... 36
2.1 Koala sample collection ... 36
2.1.1 Koala data and clinical health ... 36
2.1.2 Ocular and urogenital/cloacal swabs for *Chlamydia* spp. testing ... 37
2.1.3 Blood and spleen samples for KoRV testing ... 37
2.2 Standard molecular biology techniques ... 38
2.2.1 Positive control samples for PCR and DNA extractions ... 38
2.2.2 Agarose gel electrophoresis ... 39
2.2.3 Sanger sequencing of PCR products ... 39
2.3 Molecular testing for *C. pecorum* ... 40
2.3.1 DNA extraction from swab samples ... 40
2.3.2 Detection of *C. pecorum* DNA using qPCR ... 41
2.3.3 Amplification of the *C. pecorum* MOMP gene, *ompA*, from *Chlamydia*-positive samples ... 42
2.3.4 Amplification of genes for MLST of koala *C. pecorum* ... 42
2.3.5 *ompA* genotyping of *C. pecorum* detected in koala samples ... 43
2.3.6 MLST of *C. pecorum* detected in koala samples ... 43
2.4 Molecular testing for KoRV ... 44
2.4.1 Extraction of DNA from blood samples ... 44
2.4.2 Extraction of DNA from tissue samples ... 45
2.4.3 Detection of KoRV using qPCR ... 45
2.4.4 KoRV genotyping using KoRV-A and KoRV-B specific qPCRs ... 45

2.5 Normalisation of qPCR results using β-actin copy numbers .......................... 46

3. *Chlamydia pecorum* Infection in free-ranging koalas (*Phascolarctos cinereus*) on French Island, Victoria, Australia ..................................................................... 51

4. Identification of unusual *Chlamydia pecorum* genotypes in Victorian koalas (*Phascolarctos cinereus*) and clinical variables associated with infection ................... 56

5. Koala retrovirus genotyping analyses reveal a low prevalence of KoRV-A in Victorian koalas and an association with clinical disease ............................................ 66

6. Variation in the microbiome of the urogenital tract of female koalas with and without 'wet bottom' ........................................................................................ 76

6.1 Introduction ........................................................................................ 76

6.2 Methods ............................................................................................ 76

6.2.1 Sample Collection and initial screening ................................................. 76

6.2.2 Amplification and sequencing ............................................................... 77

6.2.3 Quality filtering and OTU assignment ................................................... 77

6.2.4 Read normalisation and analysis .......................................................... 78

6.3 Results ............................................................................................... 80

6.3.1 Clinical status of koalas ......................................................................... 80

6.3.2 Analysis and processing of sequencing data .......................................... 80

6.3.3 Phylum presence and relative abundance .............................................. 81

6.3.4 Richness and diversity ........................................................................... 81

6.3.5 Comparisons between samples using DESeq2 normalised reads ............ 82

6.4 Discussion .......................................................................................... 82

6.5 Concluding remarks ........................................................................... 86

7. Genomic comparisons of *Chlamydia pecorum* infecting koalas across Australia .. 98

7.1 Introduction ........................................................................................ 98

7.2 Methodology ...................................................................................... 99

7.2.1 Swab sample collection and selection ................................................... 99

7.2.2 DNA extraction and target enrichment ................................................ 100

7.2.3 Analysis ........ 100
7.3 Results ........ 105
7.3.1 Genome assemblies ........ 105
7.3.2 Genomic differences between *C. pecorum* infecting koalas ........ 106
7.3.3 Draft genome variant analysis ........ 107
7.3.4 Novel MLST genes in *C. pecorum* ........ 107
7.3.5 Plasmid variation ........ 108
7.3.6 Gene variation between koala-derived *C. pecorum* ........ 108
7.3.7 Other genomic features ........ 110
7.3.8 Tree distances and selection pressure on genes ........ 112
7.4 Discussion ........ 112
7.4.1 Genome assembly comparisons ........ 113
7.4.2 Phylogenetics of *C. pecorum* ........ 114
7.4.3 Tissue tropism and strain type ........ 115
7.4.4 Genome annotation comparisons ........ 115
7.4.5 Variation within the plasticity zone ........ 116
7.4.6 Genomic variation and recombination ........ 117
7.4.7 Chlamydia plasmid ........ 119
7.5 Concluding remarks ........ 119
8. General Discussion ........ 144
8.1 Overview of research aims ........ 144
8.2 *C. pecorum*, KoRV and wet bottom ........ 144
8.3 Health status of Victorian koalas ........ 148
8.3.1 Implications for koala translocations ........ 149
8.3.2 Implications for KoRV and *C. pecorum* vaccine development ........ 150
8.3.3 Implications for future prevalence and genotyping studies ........ 152
8.4 Future research directions ........ 153

8.4.1 KoRV infection in southern koalas ........ 153

8.4.2 Diagnostic testing to aid translocation ........ 153

8.4.3 Cross species transmissions of *C. pecorum* ........ 154

8.4.4 Other potential pathogens in koalas ........ 155

8.5 Concluding remarks ........ 157

9. References ........ 158

Appendices ........ 180

# Abbreviations

| | |
|---|---|
| 2D | Two dimensional |
| 3D | Three dimensional |
| AIC | Akaike information criterion |
| AIDS | Acquired Immune Deficiency Syndrome |
| ANOVA | Analysis Of Variance |
| APCAH | Asia Pacific Centre for Animal Health |
| APE | Analyses of Phylogenetics and Evolution |
| BAM | Binary Alignment/Mapping |
| BDT | BigDye Terminator |
| BH | Benjamini and Hochberg |
| BLAST | Basic Local Alignment Search Tool |
| bp | Base Pair |
| BWA | Burrows Wheeler Aligner |
| CDS | Coding DNA Sequence |
| CI | Confidence Interval |
| cm | Centimetre |
| CoA | Coenzyme A |
| CTR | Coding Tandem Repeat |
| DESeq2 | Differential Expression analysis for Sequence count data 2 |
| df | Degrees of freedom |
| dN | Nonsynonymous substitutions per nonsynonymous site |
| DNA | Deoxyribonucleic Acid |
| dNTP | Deoxyribonucleotide triphosphate |
| dS | Synonymous substitutions per synonymous site |
| dsDNA | Double stranded DNA |
| EDTA | Ethylenediaminetetraacetic Acid |
| ELISA | Enzyme Linked Immunosorbent Assay |
| EMBOSS | European Molecular Biology Open Software Suite |
| ESU | Evolutionarily Significant Unit |
| ETE3 | Environment for tree exploration 3 |
| FBS | Foetal Bovine Serum |
| FeLV | Feline Leukaemia Virus |
| +G | Gamma distribution |
| *g* | relative centrifugal force |
| GALV | Gibbon Ape Leukaemia Virus |
| GC | Guanine-Cytosine |
| GE | General Electric |
| GTP | Guanosine-5'-triphosphate |
| GTR | General Time Reversible |
| HCl | Hydrochloric acid |

| | |
|---|---|
| HIV | Human Immunodeficiency Virus |
| HKY | Hasegawa-Kishino-Yano |
| HMM | Hidden Markov Model |
| hORF-PZ | Hypothetical protein within *C. pecorum* plasticity zone |
| hr | hour |
| HS | High Sensitivity |
| IFNγ | Interferon gamma |
| +I | Proportion of invariable sites |
| IgG | Immunoglobulin G |
| IL | Interleukin |
| IPTG | isopropyl β-D-1-thiogalactopyranoside |
| ITS | Internal Transcribed Spacer |
| IUCN | International Union for Conservation of Nature |
| kb | Kilobase |
| KoRV | Koala retrovirus |
| kV | Kilovolt |
| LAMP | Loop-Mediated isothermal Amplification |
| LAZ | Los Angeles Zoo |
| LB | Luria-Bertani |
| LE | Left Eye |
| LFC | $\log_2$ Fold Change |
| LPS | Lipopolysaccharide |
| LRT | Likelihood Ratio Test |
| LTR | Long Terminal Repeat |
| MACPF | Membrane Attack Complex/Perforin |
| MAFFT | Multiple Alignment using Fast Fourier Transform |
| Mb | Megabase |
| MbRV | *Melomys burtoni* retrovirus |
| MCMC | Markov Chain Monte Carlo |
| $MgCl_2$ | Magnesium Chloride |
| MHCII | Major Histocompatibility Complex class two |
| min | Minute |
| mL | Millilitre |
| MlRV | *Megaderma lyra* retrovirus |
| MLST | Multilocus Sequence Typing |
| mM | Millimole |
| MOMP | Major Outer Membrane Protein |
| $MQH_2O$ | Milli-Q filtered water |
| MU | Management Unit |
| MuLV | Murine leukaemia virus |
| NCBI | National Center for Biotechnology Information |
| ND | Not determined |
| ng | Nanograms |

| | |
|---|---|
| nM | Nanomole |
| NST | Novel Sequence Type |
| NSW | New South Wales |
| nt | Nucleotide |
| NWB | No wet bottom |
| OJ | Oji |
| OR | Odds Ratio |
| ORF | Open Reading Frame |
| OTU | Operational Taxonomic Unit |
| PAML | Phylogenetic Analysis by Maximum Likelihood |
| PBMC | Peripheral Blood Mononuclear Cell |
| PBS | Phosphate Buffered Saline |
| PCoA | Principal Coordinates Analysis |
| PCR | Polymerase Chain Reaction |
| PERMANOVA | Permutational Analysis Of Variance |
| PERV | Porcine endogenous retrovirus |
| pH | Potential of Hydrogen |
| PiT1 | Sodium-dependent phosphate cotransporter 1 |
| PLD | Phospholipase D |
| Pmp | Polymorphic membrane protein |
| PyNAST | Python Nearest Alignment Space Termination |
| PZ | Plasticity Zone |
| QIIME | Quantitative Insights Into Microbial Ecology |
| QLD | Queensland |
| qPCR | Quantitative Polymerase Chain Reaction |
| QUAST | Quality Assessment Tool for Genome Assemblies |
| RAxML | Randomized Axelerated Maximum Likelihood |
| RE | Right Eye |
| RNA | Ribonucleic Acid |
| rpm | Revolutions per minute |
| rRNA | Ribosomal RNA |
| RT | Room Temperature |
| s | Seconds |
| SA | South Australia |
| SAM | Sequence Alignment/Mapping |
| SD | Standard Deviation |
| SE | Standard Error |
| SNP | Single Nucleotide Polymorphism |
| sp. | Species (singular) |
| spp. | Species (plural) |
| ST | Sequence Type |
| TBE | Tris/Borate/EDTA buffer |
| TE | Tris EDTA |

| | |
|---|---|
| THTR1 | Thiamine transport protein 1 |
| Tris | Tris(hydroxymethyl)aminomethane |
| tRNA | Transfer RNA |
| U | Units |
| UGT | Urogenital tract |
| UK | United Kingdom |
| USA | United States of America |
| USC | University of the Sunshine Coast |
| UV | Ultraviolet |
| V | Volts |
| vs | Versus |
| w/v | weight of solution in the total volume of solution |
| WB | Wet bottom |
| X-gal | 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside |
| Ω | Ohms |
| μF | Microfarad |
| μg | Micrograms |
| μL | Microlitre |
| μM | Micromole |

# List of Tables

Table 1.1. ..................................................................................................... 32

Table 1.2. ..................................................................................................... 33

Table 1.3. ..................................................................................................... 34

Table 2.1. ..................................................................................................... 48

Table 6.1. ..................................................................................................... 88

Table 6.2. ..................................................................................................... 89

Table 6.3. ..................................................................................................... 91

Table 6.4. ..................................................................................................... 92

Table 7.1. ..................................................................................................... 121

Table 7.2. ..................................................................................................... 123

Table 7.3. ..................................................................................................... 124

Table 7.4. ..................................................................................................... 126

Table 7.5. ..................................................................................................... 128

Table 7.6. ..................................................................................................... 129

Table 7.7. ..................................................................................................... 132

## List of Figures


Figure 1.1....... 35

Figure 2.1....... 50

Figure 6.1....... 94

Figure 6.2....... 95

Figure 6.3....... 96

Figure 6.4....... 97

Figure 7.1....... 134

Figure 7.2....... 135

Figure 7.3....... 137

Figure 7.4....... 138

Figure 7.5....... 139

Figure 7.6....... 140

Figure 7.7....... 141

Figure 7.8....... 142

Figure 7.9....... 143

## Third party copyright material

Below are details of third party materials used in this thesis for which permission for inclusion has been sought.

| Third party material citation | Location | Permission granted |
|---|---|---|
| Legione, A. R., J. Amery-Gale, M. Lynch, L. Haynes, J. R. Gilkerson, F. M. Sansom & J. M. Devlin (2016). "*Chlamydia pecorum* infection in free-ranging koalas (*Phascolarctos cinereus*) on French Island, Victoria, Australia." *J Wildl Dis* 52(2), 426-429. | Chapter 3 | Yes |
| Legione, A. R., J. L. S. Patterson, P. L. Whiteley, J. Amery-Gale, M. Lynch, L. Haynes, J. R. Gilkerson, A. Polkinghorne, J. M. Devlin & F. M. Sansom (2016). "Identification of unusual *Chlamydia pecorum* genotypes in Victorian koalas (*Phascolarctos cinereus*) and clinical variables associated with infection." *J Med Microbiol* 65(5), 420-428. | Chapter 4 | Yes |
| Legione, A. R., J. L. S. Patterson, P. Whiteley, S. M. Firestone, M. Curnick, K. Bodley, M. Lynch, J. R. Gilkerson, F. M. Sansom & J. M. Devlin (2017). "Koala retrovirus genotyping analyses reveal a low prevalence of KoRV-A in Victorian koalas and an association with clinical disease." *J Med Microbiol* 66(2), 236-244. | Chapter 5 | Yes |

# 1. Literature Review

## 1.1 The koala, an overview

The koala (*Phascolarctos cinereus,* meaning pouched ashen bear) is the last remaining species of the family Phascolarctidae, within the order Diprotodontia. Its closest relative are the wombats, with whom they share the suborder Vombatiformes. The family was first named by the French zoologist Henri Marie Ducrotay de Blainville in 1816 (Plassan 1816). The species name *cinereus* arose from the German zoologist Georg August Goldfuss, who had named the species *Lipurus cinereus* in 1817. The species was listed by the International Union for Conservation of Nature (IUCN) as of 'least concern' in 1998, a category it has maintained until this day (Gordon *et al.* 2008), despite disagreements between stakeholder groups. This classification was established due to the koala's large population range (from south-east South Australia, to southern Queensland) and estimated total population size across this range.

Koalas are one of Australia's most iconic and recognisable wildlife species, along with the kangaroo, and are a highly sought after attraction for tourists. Their economic impact has been estimated to be as high as $1.8 billion annually (Hundloe *et al.* 1997), although other reports have suggested a more conservative value of around $336 million annually (Davis *et al.* 2001). Its iconic stature amongst Australian wildlife has led to community groups championing its cause, with repeated efforts to have the species listed as 'vulnerable' (Phillips 2000), despite the arguments put forth by the IUCN (Gordon *et al.* 2008) and koala researchers (Strahan & Martin 1981). Estimates from a koala focused non-government organisation in 1994 suggest that koala numbers in Australia are fewer than 100,000 (Phillips 2000). This non-government organisation has been active in the push to reclassify the species as vulnerable and thus their estimates of koala numbers should be viewed with caution due the potential for a conflict. Ideally, koala population estimates from a number of different sources should be considered instead.

## 1.2 Management of koala populations

In the past koalas in each state have been divided into three subspecies: *Phascolarctos cinereus cinereus* (New South Wales), *Phascolarctos cinereus adustus* (Queensland) and *Phascolarctos cinereus victor* (Victoria). However, these subspecies are now

considered arbitrary selections from a cline, rather than being individual taxa (Strahan 1978). In general, koalas in Victoria are larger than their northern counterparts and have grey rather than rust coloured fur (Martin & Handasyde 1999).

More recent investigation of this regional based system of classification utilised microsatellite technology on populations from Victoria, Queensland, and New South Wales (Houlden *et al.* 1996). This revealed that the Victorian, or southern, population clustered together as a single management unit (MU), which in turn was deemed to be separate to the populations of Queensland and New South Wales. It was tentatively suggested that these populations may represent separate evolutionarily significant units (ESUs). However, later mitochondrial DNA analysis demonstrated that despite their wide geographic distances, all populations in Australia tested represented a single ESU (Houlden *et al.* 1999). Furthermore, there was only a minor difference between haplotypic diversity between and within sub-species (55.36% and 44.64% respectively). These results were "consistent with a species comprised of highly differentiated populations, rather than a species with several highly-differentiated subspecies" (Houlden *et al.* 1999).

In the last decade, broader population analysis has also identified limited genetic diversity in Victorian koalas. Genetic diversity studies using microsatellite markers (Ruiz-Rodriguez *et al.* 2014) compared two Victorian koala populations with a population from Queensland, and also with captive populations in zoological collections in the USA (Ruiz-Rodriguez *et al.* 2016). Koala genetic diversity was found to cluster by state of origin. Haplotypic diversity using 662 koalas from Queensland, New South Wales, South Australia, and Victoria also supported the conclusions that koalas across the states should be classified as a single ESU, with no sub-speciation (Neaves *et al.* 2016). In addition, these haplotypic diversity studies suggested that historical geographic barriers in Queensland (including the Clarence and Brisbane rivers) caused some regional diversity clustering, however, there was no apparent effect of geographical barriers on gene flow between koalas in New South Wales and Victoria. Koalas in these two states clustered together in regard to haplotypes present, but with Victoria having a lower diversity.

Although now considered a single species, koalas in each state (Queensland, New South Wales, Victoria, and South Australia) are managed as different populations, as each fall under the jurisdiction of different state governments. In 2012, due to population

declines, koalas in certain regions of Queensland were listed as vulnerable by the federal government. One population in south-east Queensland was found to have declined by 51% in a three year span (Department of Environment and Resource Management 2009). Koala population decline can be caused by a number of factors, including disease (Polkinghorne *et al.* 2013), habitat fragmentation (Melzer *et al.* 2000), urbanisation (Smith & Smith 1990), motor vehicle trauma (Dique *et al.* 2003; Griffith *et al.* 2013) and dog attacks (Lunney *et al.* 2007). From the first arrival of European settlers, koala populations have gone through cycles of growth and decline. When early naturalists and zoologists surveyed the country, they predominately found koalas in areas of low human settlement, a factor that has been attributed to hunting by Indigenous Australians (Parris 1948). It is suggested that following Indigenous population decline, the number of koalas increased dramatically (Melzer *et al.* 2000), although this theory lacks supporting evidence. These population rises were punctuated with hunting 'open-seasons' in Queensland in 1919 and 1927, where fur sales data from the Queensland government at the time, suggested that up to 1,000,000 and 600,000 koalas were killed respectively (Hrdina & Gordon 2004).

Population crashes have also occurred due to disease, with early reports that koalas were almost driven to extinction around 1900 due to 'ophthalmic disease and periostitis of the skull', as well as the impacts of hunting (Troughton 1941). In addition to dramatic changes in koala population numbers there are also pressures acting more slowly on these populations. The impact of urbanisation as well as agricultural land use has led to increases in habitat fragmentation and deaths due to motor vehicle trauma and dog attacks. Koalas are a predominantly nocturnal species, and move from tree to tree at ground level. Areas of suitable habitat, containing a large supply of eucalypts, need to be connected by a corridor of trees for them to be able to move from location to location (de Oliveira *et al.* 2014). However, the majority of these corridors occur along roads, meaning they are prone to motor vehicle trauma. One five-year study in a section of south-east Queensland found an average of 281 koalas were hit by vehicles each year (Dique *et al.* 2003). The requirement to move from tree to tree at a ground level also leads to an increased chance of being victim of a dog attack. An investigation into the viability of koala populations at Port Stephens, New South Wales, showed that reducing dog predation was the key issue for population survival (Lunney *et al.* 2007). Due to the difficulty involved in moving from location to location for koalas, they are perhaps

more prone to stay in the same area, which puts pressure on their food sources, often leading to over-browsing and then tree die off (Martin 1985a). This in turn leads to starvation of animals and the decline of the population (Martin 1985b). Koalas suffering severe food shortages in Cape Otway, Victoria, have been shown to remain in areas of 0.9 – 1.0 hectares, despite sufficient woodland connectivity to move wider distances (Whisson *et al.* 2016).

## 1.3 Koalas in Victoria

Koalas in Victoria have undergone similar population trends to those in other states of Australia. There are early reports of locations that had no koalas present in the mid-1800s, such as on the lower Goulburn River, but were then found to contain koalas by the late 1800s (Parris 1948). The notion that koalas had a more limited historical range prior to European settlement has been contested outside of peer-reviewed literature, with critics of Parris' report maintaining that the information is merely an anecdote with no data attached (Tabart 2002). Whilst the criticism is correct in part, the same could be said of the majority of reports by early naturalists before the advent of scientifically determined population estimates. As stated previously, koalas declined in the early 1900s (Troughton 1941) but the documented evidence of this decline is minimal. There is no statement of the original estimated population size, how many cases of disease were seen, nor how the decline was measured. Hunting for the purpose of exporting koala fur also allegedly caused the death of 'millions' of koalas in Victoria (Troughton 1941). Once the species was afforded greater legislative protection, it is suggested that the animal was still hunted for fur, but exported under the guise of wombat fur (Mattingley 1901). Aside from the transcribed statement made by Mattingley in a meeting of the Victorian Naturalists, no evidence exists for this claim. Indeed, much of the early history and recordings of the koala in different regions are anecdotal observations and hearsay written in formal journals of the time. After population declines, the only remnant populations in Victoria were found on French Island and in Gippsland. The French Island population originated when unknown number of koalas were introduced in the late 1880s and quickly proliferated. This may have been undertaken by fisherman or farmers from Corinella as a means to protect the animal from an area prone to bushfires (Lewis 1954; Martin & Handasyde 1999). In an attempt to increase the population of koalas in Victoria and decrease the unsustainable number on French Island, koalas were subsequently translocated to other local islands, namely

Quail Island and Phillip Island (Warneke 1978). An earlier translocation from Gippsland to Phillip Island had also occurred in 1870 (Martin & Handasyde 1999). French Island koalas were also used to populate Kangaroo Island in South Australia, a state in which koalas were deemed extinct at the time (Masters *et al.* 2004). As few as 18 animals were translocated to Kangaroo Island between 1923 and 1925, and the population reached approximately 27,000 in roughly 75 years (Masters *et al.* 2004).

In the last 70 years koalas in Victoria have been translocated across the state as part of population management strategies and now occupy much of Victoria. As populations in each new location increased, koalas were translocated to unpopulated regions of suggested historic range to allow for further growth of the species (Martin & Handasyde 1990). Repopulation via translocations from island populations is likely to have resulted in a decreased genetic pool, potentially creating genetic bottlenecks, founder effects and increasing the likelihood of future inbreeding potential (Houlden *et al.* 1996). Evidence of this can be seen in the lack of diversity in the major histocompatibility complex class two molecules (MHCII) in koala populations in Victoria (Lau *et al.* 2014). Assessment of MHCII diversity in various regions revealed that koala populations in six New South Wales regions had an average of seven DA β1 (DAB) variants present, and the Queensland population tested had five. A subset of Victorian koalas from south-east Gippsland were tested and found to also have seven variants. In comparison, the French Island koala population, and three other populations founded from the French Island colony, had only three variants (Lau *et al.* 2014). This highlights that the Gippsland population is more genetically diverse than other koalas within the state. Lee *et al.* (2010) determined that major roads and rivers acted as barriers to gene flow in Queensland populations, which could be a factor in the low level of natural gene dispersal from south-east Gippsland koala populations to other mainland populations in Victoria. The management strategy of translocation, whilst effective in increasing the population of koalas in the Victoria, also led to at least one occasion in which disease was introduced into a naïve koala population. The introduction of *Chlamydia* infected individuals to an established koala population in the Grampians in western Victoria caused a population crash in this region (Martin & Handasyde 1990). The opposite situation, whereby naïve individuals are translocated to a region where disease is already present is also possible. Santamaria and Schlagloth (2016) documented a case where 30 naïve koalas were moved from an island population to mainland locations

where *Chlamydia* was already present. After less than two years, the majority of individuals that were recaptured were found to have antibodies to the pathogen (16/17), whilst 11/17 had detectable urogenital tract infections. Therefore, if translocation is to be used for koala population management, either to re-establish historic range or to reduce the density of some populations, knowledge of diseases within these populations plays an important role.

## 1.4 Impact of disease on wildlife population management

Worldwide, efforts to re-establish declining wildlife populations have been met with mixed success. In part, this is due to difficulties in assessing the success, or otherwise, of the different conservation projects. Kapos *et al.* (2008) highlighted that these difficulties can arise due to clear objectives not being set, the limited availability of resources to measure outcomes long term, and the lack of incentive to investigate a potentially poor outcome. Seddon (1999) suggested three objectives for wildlife conservation, for cases of population reintroduction: "the survival of the release generation; breeding by the release generation and their offspring; and persistence of the re-established population, perhaps assessed through extinction probability modelling". Translocation of individuals or groups of animals, either from threatened populations or captive breeding colonies, has been used to improve conservation outcomes. Management of these threatened populations is varied, but those at critically low levels will often be the subject of breeding programs, with controlled releases into the wild. This has occurred with a number of Australian species with mixed success, and few species have successfully recovered through this method. An example of an, as yet, unsuccessful recovery target is the orange-bellied parrot (*Neophema chrysogaster*), of which there are thought to be fewer than 70 wild individuals (Orange-bellied Parrot Recovery Team 2016a). Individuals have successfully been bred in captive populations, with approximately 350 birds in different colonies around the country (Orange-bellied Parrot Recovery Team 2016a). Although captive individuals are released to supplement the wild population, the number of wild breeding pairs have failed to increase (Starks *et al.* 1992; Pritchard 2012; Orange-bellied Parrot Recovery Team 2016a; Orange-bellied Parrot Recovery Team 2016b). The recovery efforts have been hampered by psittacine beak and feather disease virus, a circovirus that is lethal to parrot species. This virus first threatened recovery efforts in the 1980s, when the first captive breeding population was affected (Peters *et al.* 2014). The threat was minimised with the improvement of

management techniques, but recently re-emerged with the detection of the virus in the wild population (Peters *et al.* 2014). This detection is thought to have played a role in reducing the survival of juveniles in 2015-2016 to 12%, lower than the average for the previous two decades (Orange-bellied Parrot Recovery Team 2016a). As a result of this virus, management practices have been altered in an effort to reduce contamination of feeding stations and nest boxes. The prevalence of infectious diseases and parasites in a source population, as well as at the planned release site of any species undergoing translocations need to be considered.

Management strategies will differ when disease mitigation is the rationale for translocations or releases. Such an approach can be observed in the management of the Northern corroboree frog (*Pseudophryne pengilleyi*), whose natural populations have been devastated by chytrid fungus (*Batrachochytrium dendrobatidis*) (McFadden *et al.* 2016). All potential release sites for the species are considered to be contaminated with the fungus, although small populations have managed to persist. Therefore, a strategy of annually supplementing two field sites with captive bred eggs and tadpoles has been developed, with the knowledge that most released individuals will not survive to maturity (McFadden *et al.* 2016). For both the Northern corroboree frog and the orange-bellied parrot, environmental sampling still needs to be undertaken at the intended release sites prior to the release of naïve individuals. The ability of pathogens such as those listed here to persist within the environment, and the dispersal of released individuals outside the region covered by environmental sampling makes comprehensive reduction of disease risk unlikely.

Another management strategy used in conjunction with the establishment of naïve satellite populations is the concurrent development of a vaccine against the most prominent diseases associated with population decline. This is the case with the Tasmanian devil (*Sarcophilus harrisii*), which has been greatly affected by one of the few existing contagious cancers, known as devil facial tumour disease (Hawkins *et al.* 2006). Insurance populations of Tasmanian devils were established in zoological collections around Australia, as well as on islands free of predators (Hogg *et al.* 2016). These populations have been established using animals free of the tumour, which must be kept in quarantine for extended periods of time to ensure no disease develops. Insurance populations may either be used to help re-establish wild populations in Tasmania when the local population goes extinct, or when a viable vaccine candidate is

developed. The length of development time for a successful vaccine makes this strategy dependant on sufficient funding over a number of decades. In this case, the devil facial tumour was first described in 1996 (Hawkins *et al.* 2006), and only recently have potential vaccine candidates shown promising results (Kreiss *et al.* 2015; Tovar *et al.* 2017).

Further to managing population survival, the genetic diversity of species needs to be conserved to minimise potential founder effects. Genetic bottleneck events, whereby the broader gene pool is substantially reduced due to a population decline, is an issue in wildlife conservation. A recent example of this was identified in the red-billed chough (*Pyrrhocorax pyrrhocorax*). In these birds a recessive allele causing blindness in nestlings was found to be expressed in an inbred population (Trask *et al.* 2016). These founder effects can be addressed with strategic planning and recruitment of additional individuals, as has occurred for both the Tasmanian devils (Hogg *et al.* 2016) and orange-bellied parrot (Pritchard 2012). It has been hypothesised that a reduced gene pool, and subsequent inbreeding, would lead to increased susceptibility of a population to a catastrophic decline due to infectious disease (O'Brien & Evermann 1988). This could arise in a scenario where the genes characterising the immune system within a population would not have sufficient diversity to survive a novel pathogen. Spielman *et al.* (2004) sought to show this experimentally, noting that examples in the literature such as cheetah and panther declines (O'Brien *et al.* 1985; Roelke *et al.* 1993) were not repeatable, and were confounded by other co-variables. Using *Drosophila melanogaster,* they produced populations of varying genetic diversity, and through exposing them to pathogens, noted that lower diversity resulted in reduced resistance. Importantly, this was attributed to the loss of specific resistance alleles through genetic drift, not inbreeding depression (Spielman *et al.* 2004).

### 1.4.1 Diseases impacting koalas that may affect population management

It is clear from these examples in other species that knowledge of infectious diseases within a wildlife population is important in planning and implementing management strategies. Early records of disease in koalas are scarce, but the impact of disease on these animals appears to have been significant. As already mentioned, near extinction events around the turn of the twentieth century have been attributed to ophthalmic and bony disease, although no detail is known about the causative agents. Nonetheless, they suggest that disease has been a major threat to koala populations for well over 100 years

(Troughton 1941). One of the earliest publications specifically detailing causes of mortality in koalas found pathology suggesting hepatitis, leukaemia, pneumonia, anaemia, suppuratives cholangitis, ulcerative colitis, *Cryptococcus*, cystadenoma and sepsis post-trauma across 28 necropsies from Taronga Park Zoological Gardens (New South Wales, Australia) (Backhouse & Bolliger 1961).

While a number of known pathogens have been isolated from koalas, very few broad epidemiological surveys examining the effect of these pathogens on population health have been carried out, perhaps owing to the difficulty of accumulating sufficient sample numbers. Koalas have been found to be infected with up to three novel papillomavirus (Antonsson & McMillan 2006), two herpesviruses (Vaz *et al.* 2011; Vaz *et al.* 2012; Stalder *et al.* 2015), *Mycobacterium ulcerans* (Mitchell *et al.* 1987), *Toxoplasma gondii* (Canfield *et al.* 1990), *Cryptococcus* spp. (Connolly *et al.* 1999; Krockenberger *et al.* 2002), *Sarcoptes scabiei* (Obendorf 1983) and four *Trypanosoma* spp. (McInnes *et al.* 2011; Barbosa *et al.* 2016). In addition to these agents, two of the most prominent pathogens of koalas in modern times are *Chlamydia pecorum* and koala retrovirus (KoRV) (Blanshard & Bodley 2008).

## 1.5 *Chlamydia pecorum*

*C. pecorum* is a Gram-negative, obligate intracellular bacterium in the family *Chlamydiaceae*. All members of this family are intracellular parasites and rely on the host cell for key compounds, in particular adenosine triphosphate (Weiss 1965; Hatch *et al.* 1982). *Chlamydiaceae* have a biphasic development cycle (Tamura & Higashi 1963). The two stages comprise developmental forms known as elementary bodies and reticulate bodies and differ in morphology, ability to replicate, and ability to infect cells (Tamura & Manire 1967; Tamura *et al.* 1971). Elementary bodies are the predominantly extracellular infectious form, and are unable to replicate (Moulder 1991). Their extracellular survival is permissible through the use of a spore-like outer-membrane complex. The majority of this membrane complex consists of the major outer membrane protein (MOMP) and two cysteine rich outer membrane proteins (OmcA and OmcB), that undergo extensive disulphide bond cross-linking(Caldwell *et al.* 1981; Hatch *et al.* 1981; Hatch *et al.* 1984; Hatch 1996). Elementary bodies infect the host epithelial cell via adhesion to host cell receptors and parasite-induced phagocytosis (Byrne & Moulder 1978). Cell entry by the bacteria occurs within an inclusion, and requires Type III secretion system effectors known as inclusion membrane proteins (Subtil *et al.* 2001).

The bacteria then differentiate into the reticulate body form. The reticulate body can undertake cell division over the course of 24 – 72 hours (Bedson & Bland 1932), before returning to elementary bodies and subsequent spread through host cell lysis or extrusion from the host cell (Hybiske & Stephens 2007). Alternatively, after entry into the host cell the organism can enter a persistent state known as an aberrant body, with reduced metabolic activity (Moulder *et al.* 1980). Chlamydial infections are often asymptomatic, and can persist in a host for months or years if untreated (Meyer & Eddie 1933; Holland *et al.* 1992; Bazala & Renda 2005). Disease can be associated with the host inflammatory response, with proinflammatory cytokines released by infected cells leading to tissue damage in the infected region (Rasmussen *et al.* 1997). In general, members of the *Chlamydiaceae* family have been associated with diseases such as trachoma (Collier & Sowa 1958), pelvic inflammatory disease (Mårdh *et al.* 1977), pneumonia (Beem & Saxon 1977), and arthritis (Keat *et al.* 1980). The mechanisms of virulence have not been well established due to the difficulty of conducting *in vivo* studies (reviewed by Byrne (2010)). Candidate virulence factors include the type III secretion system, which is crucial in the pathogenesis of some other Gram-negative bacteria, potential cytotoxins that are similar to large clostridial cytotoxins, polymorphic membrane proteins and stress response proteins. A chlamydial plasmid has also been identified and is associated with virulence in *Chlamydia trachomatis*, and can act as a regulator of chromosomal genes (Carlson *et al.* 2008).

The genus *Chlamydia* contains an increasing number of species, with a wide host range (Appendix 1). *C. pecorum* was first proposed as a separate species to that of *Chlamydia psittaci*, *C. trachomatis* and *Chlamydia pneumoniae* in 1992 (Fukushi & Hirai 1992). An analysis of DNA-DNA homology, using Southern blotting, found that organisms isolated from cattle and sheep were considerably different to the three *Chlamydia* spp. identified at the time. *C. pecorum* was found in association with cases of sporadic encephalitis, infectious polyarthritis, pneumonia, and diarrhoea in livestock. For a brief period, *C. pecorum*, as well as *C. pneumoniae*, *C. psittaci*, *Chlamydia abortus*, *Chlamydia felis* and *Chlamydia caviae* were part of a genus known as *Chlamydophila*. This separation from the *Chlamydia* genus was based primarily on phylogenetic analysis of 16S and 23S ribosomal RNA sequences (Everett *et al.* 1999). However, within a decade, and in conjunction with more advanced knowledge of the family through full genome sequences, it was determined that the *Chlamydophila* genus should

be merged with the *Chlamydia* genus (containing *C. trachomatis*, *Chlamydia muridarum* and *Chlamydia suis*), unifying the different species under one genus, *Chlamydia*, again (Stephens *et al.* 2009). Since the initial detection of *C. pecorum* in cattle, the organism has been detected in a diverse range of species, with a wide range of clinical and sub-clinical infection sites (Table 1.1). Worldwide, the species perhaps most affected by *C. pecorum* infection, in terms of its known impact on wild populations, is the koala.

### 1.5.1 *C. pecorum* in koalas

*Chlamydia* in koalas was first isolated in the 1970s from a case of keratoconjunctivitis (Cockram & Jackson 1974) and then later found to be associated with infertility and reproductive disease (Brown & Grice 1984). At first the causative agent was named *C. psittaci*, as at the time all *Chlamydia* isolates of animal origin were assumed to be this organism. Further research suggested two different forms of the pathogen were associated with disease in koalas, which were named *C. psittaci* type-I and *C. psittaci* type-II (Girjes *et al.* 1988).

As research in the wider *Chlamydia* field advanced two novel species were found to be present among *Chlamydia* isolates. One species was *C. pecorum* which had been previously isolated from cattle (Fukushi & Hirai 1992), and the other was *C. pneumoniae* which was found previously in humans (Grayston *et al.* 1989). Through sequencing of the outer membrane protein 2 gene (*omp2*) of several koala *Chlamydia* samples, it was determined that these two species, not *C. psittaci*, were responsible for causing disease in koalas (Glassick *et al.* 1996). Sequencing revealed that *C. psittaci* type-I isolates were actually *C. pneumoniae,* whereas *C. psittaci* type-II isolates were in fact *C. pecorum*. Genus-specific PCR and species specific DNA probe hybridisation investigated the prevalence of these two organisms in two populations of free ranging koalas in Queensland (Jackson *et al.* 1999), with 85% (n = 33) of koalas infected in one population but only 10% (n = 20) in the second population. *C. pecorum* was the more common agent and was associated with all five instances of clinical signs of disease detected. There was no association between detection of *C. pneumoniae* and disease in koalas. Subsequent prevalence investigations have focussed on *C. pecorum*, with few published accounts of *C. pneumoniae* prevalence (Polkinghorne *et al.* 2013). At least two studies suggested that additional *Chlamydiaceae* species may exist within koala populations (Devereaux *et al.* 2003; Burach *et al.* 2014), but no disease has been

associated with these organisms. Five genotypes of *Chlamydiales* bacteria were also detected in tick species harvested from koalas (Burnard *et al.* 2017), but the detected families are not yet associated with disease in mammals, being primarily detected in insects.

*C. pecorum* is the primary pathogen impacting Queensland koalas and is generally widespread within northern koala populations (Kollipara *et al.* 2013b). *C. pecorum* is associated with ocular (Cockram & Jackson 1974; Cockram & Jackson 1981) and urogenital pathology (Hemsley & Canfield 1997). Ocular infection is characterised by conjunctivitis, and in severe cases can lead to blindness (Dickens 1976). A common clinical sign of urogenital tract infection is a staining or scalding of the rump area known as 'wet bottom' or 'dirty tail' (Dickens 1976), which is associated with urinary incontinence caused by cystitis. Urogenital tract infection was first hypothesised as a cause of infertility in female koalas, with a study finding significantly lower occurrence of pouch young in infected Victorian koalas on Phillip Island, compared to an uninfected population on French Island (McColl *et al.* 1984). As already mentioned, the accidental introduction of *Chlamydia* into a naïve koala population in the Grampians, Victoria, led to a severe population crash and a population wide fertility rate of zero two decades after introduction (Martin & Handasyde 1990). Recently a case of pneumonia in koalas was attributed to *C. pecorum* (Mackie *et al.* 2016), highlighting the diverse forms of disease caused by this organism.

Although research has focused on ocular and lower urogenital tract infections, limited studies suggest that *C. pecorum* causes disease in the upper urogenital tract. A strong association exists between *C. pecorum* presence in cells of the kidney, ovary and uterus, and disease in these tissues (Higgins *et al.* 2005). *Chlamydia* reticulate bodies were also detected in macrophages, suggesting that systemic spread could occur through dissemination of these cells. There has been limited research assessing the systemic spread of chlamydial infections in koalas. Burach *et al.* (2014) undertook histopathology, PCR and immuno-histochemistry assessment on samples from multiple tissues from 23 koalas, including eye, urogenital tract, gastrointestinal, lymphatic organ, lung, liver, and heart. A pan-*Chlamydiales* PCR detected 19 positive samples, only one of which was determined to be *C. pecorum* via sequencing (a gastrointestinal tract sample), with the other sequenced positives matching uncultured *Chlamydiales* spp. Immunohistochemical labelling detected *Chlamydiaceae* lipopolysaccharide (LPS) in

10 tissues from 9 koalas, and 6 of these koalas were positive in the urogenital tract samples. The authors suggested that the detection of *Chlamydiaceae* LPS in the gastrointestinal tract, spleen, lung, and cloaca, was evidence of systemic spread. It is unclear whether these were detections of the same organism in different tissues, or if they represented mixed infections. The gastrointestinal tract is a natural site of infection for *Chlamydia* spp. in cattle (Li *et al.* 2016) and has been shown to be a site of persistent infection in mice (Yeruva *et al.* 2013), so it is possible that the koala gastrointestinal tract may be infected independently of the urogenital tract or other sites.

### 1.5.2 *C. pecorum* variation in koalas and association with virulence

The diversity of *C. pecorum* has been investigated by sequencing single genes such as *ompA* (encoding MOMP) (Kollipara *et al.* 2013b), and with the more conservative multi-locus sequence typing (MLST) approach (Jelocnik *et al.* 2013). The intracellular nature of *C. pecorum* has made full genome analysis difficult in the past, with the first full genome not published until 2011 (Mojica *et al.* 2011) and the first strains of *C. pecorum* from koalas sequenced several years later (Bachmann *et al.* 2014). More recently, full genome analysis has become more accessible due to advances in next-generation sequencing technology, and the use of hybridisation technology (Seth-Smith *et al.* 2013; Bachmann *et al.* 2015). This has allowed the sequencing of near complete genomes from swab samples containing relatively low numbers of bacteria, and could now be considered the gold standard in comparative analysis of bacterial genetic diversity.

#### 1.5.2.1 Single gene analysis

The *ompA* gene has four variable domains and was first used for *C. pecorum* genotyping in koala samples by analysing the nucleotide sequence of the 4$^{th}$ variable domain (Jackson *et al.* 1997). This study identified five genotypes present in Australian populations, designated A, B, C, D, and E. The genotypes were found to have a sequence dissimilarity ranging from 8 – 30%. Only one sample from a southern koala was tested (from Raymond Island, Victoria) and was the only genotype B detected. The genotype most similar to genotype B was genotype D (8% dissimilarity), which was also only detected in one sample, for which the location was not recorded. More recent assessment of 12 different genes suggested that *ompA* is a good molecular marker for

epidemiological investigation, and is potentially associated with virulence (Yousef Mohamad *et al.* 2008).

*C. trachomatis* in humans and *C. pecorum* in koalas share similar gross and histopathological lesions, suggesting they may have similar pathogenic mechanisms. Diversity in the MOMP of *C. trachomatis* is thought to be a key factor in causing disease, as it can lead to immune evasion and subsequent infection by multiple MOMP serotypes. Analysis of a single population of koalas (47 individuals) in New South Wales with varying levels of disease found limited diversity of *ompA* genotypes amongst the population, with only a single genotype present (Higgins *et al.* 2012). The nucleotide sequence of this genotype was later classified as genotype F (Kollipara *et al.* 2013b). Within the same population, they identified a variable number of 15 nucleotide repeat motifs within ORF663, which encodes a hypothetical protein. ORF663 has been suggested as a virulence factor in *C. pecorum* infecting cattle, and Yousef Mohamad *et al.* (2008) found that repeat motif quantity was significantly associated with the presence or absence of clinical disease. A low number of tandem repeats (14 – 45) were detected in pathogenic isolates, and a high number of tandem repeats (53 – 68) were found in intestinal isolates from healthy cattle. In the koala study the number of repeats ranged from 17 to 77 across 25 samples, but no association analysis between OFR663 diversity and virulence was undertaken.

More recent *ompA* genotyping of *C. pecorum* in koala populations from Queensland, New South Wales, and South Australia detected additional genotypes of *C. pecorum* (Kollipara *et al.* 2013b). These researchers were investigating genetic diversity to inform the development of a potential vaccine against *C. pecorum*. In total, 45% of the 403 koalas from eight wild populations and three wildlife hospitals tested positive for *C. pecorum*. Only one southern population was included, with all four koalas from South Australia positive for *C. pecorum*. Based on near full *ompA* gene (1120/1170 bp) sequencing an additional six genotypes of *C. pecorum*, namely F, G, H, I, J, and K, were described. A new genotype was declared when nucleotide identity was less than 99% in comparison to previously defined types. This reflects a nucleotide difference of approximately 12 or more bases, based on a sequenced *ompA* length of 1120 bp. It is interesting to note the two genotypes found in the four South Australian koalas were B and G. Previously, genotype B had only been detected only in Victoria (Jackson *et al.* 1997).

Whilst *ompA* can be utilised as an epidemiological marker, it is a region of high nucleotide substitution rates compared to the rest of the genome and therefore may not be an optimal target for evolutionary analysis (Brunelle & Sensabaugh 2006). In an attempt to determine an appropriate marker for measuring *C. pecorum* genetic diversity, specifically in koalas, Marsh *et al.* (2011) assessed ten regions in *C. pecorum* (16S rRNA, 16S/23S intergenic spacer, *ompA*, *omcB*, *pmpD*, *incA*, *copN*, Membrane Attack Complex/Perforin (MACPF), ORF663 and *tarP*) for their usefulness in assessing phylogenetic relationships between isolates. The mean diversity of nucleotide sequences between isolates indicated that *ompA, incA,* ORF663, and *tarP* could potentially be used to discriminate isolates phylogenetically. Whilst neither *ompA* nor *incA* could be used as fine detailed markers, in the face of a large number of samples this analysis suggests that *ompA* would still be useful as a cost-effective comparative marker for epidemiological investigations.

#### 1.5.2.2 MLST analysis of *C. pecorum*

A robust MLST method was developed to type *Chlamydiales* (Pannekoek *et al.* 2008), and this method has been more recently refined to increase the specificity for *C. pecorum* (Jelocnik *et al.* 2013). MLST methods aim to sequence several conserved genes considered essential for cell function (housekeeping genes), and different nucleotide sequences within a target allele are given a specific allele number, with the combination of allele numbers giving rise to a ‘sequence type’ (ST) (Maiden *et al.* 1998). The *Chlamydiales* MLST utilises seven housekeeping genes (Table 1.2). Currently there are 156 *Chlamydiales* sequence types, 36 of which are *C. pecorum* types. MLST strategies have been used to differentiate *C. pecorum* types responsible for causing different clinical signs of disease in ruminants (Jelocnik *et al.* 2014b). The MLST types of *C. pecorum* in sheep were clustered phylogenetically into groups that correlated with specific clinical signs. For example, *C. pecorum* isolated from sheep suffering from conjunctivitis tended to cluster together, as did *C. pecorum* isolated from sheep suffering from both conjunctivitis and poly-arthritis (or poly-arthritis alone). Although MLST methodology is a robust means of assessing relatedness between isolates, the requirement to use Sanger sequencing on both DNA strands (i.e. forward and reverse) for seven genes makes it economically impractical for large numbers of samples.

#### 1.5.2.3 Full genome sequencing of *Chlamydia*

Full genome sequencing is required to fully understand phylogenetic relatedness and evolutionary history of different *Chlamydia* isolates, as well as to determine potential virulence genes. Genomic sequencing of *C. pecorum* isolates grown in cell culture has been limited to four complete sequences of ruminant origin, as well as a small number of draft sequences from both ruminants and koalas. The most complete sequence, of the type strain E58, was an isolate of bovine origin (McNutt & Waller 1940). This type strain was sequenced with a whole-genome shotgun approach (Mojica *et al.* 2011) and is 1,106,197 bp long, encoding 1073 putative coding sequences. Comparison with other available *Chlamydia* genomes revealed that the majority of genomic diversity was found in the plasticity zone (PZ). The PZ of E58 contained two complete cytotoxin genes and three phospholipase D genes, as well as a MACPF domain that is either truncated or missing in other members of the genus. A small number of genomes from *C. pecorum* from koalas are available, with the majority published in draft form (i.e. with gaps or ambiguous bases). Bachmann *et al.* (2014) used standard *Chlamydia* cell culture followed by Illumina HiSeq to sequence the majority of three genomes of *C. pecorum* from koalas (IPTaLE, MC/MarsBar, and DBDeUG). MC/MarsBar had a single nucleotide polymorphism (SNP) that caused a truncation in the second cytotoxin (or *toxB*) gene, resulting in a pseudogene. Deletions in the cytotoxin genes of most genitotrophic and oculotropic serovars of *C. trachomatis* have been thought to play a role in tissue tropism of *C. trachomatis* (Carlson *et al.* 2004; Carlson *et al.* 2005), but it is unclear whether similar differences impact tropism of *C. pecorum*.

One of the key barriers to full genome sequencing of *Chlamydia* spp. is the requirement for the organism to grow within a cell. Genomic analysis requires culture to a substantial yield in a permissive cell line, and it has been shown that each inclusion forming unit tends to produce fewer progeny than a standard extracellular bacteria would on a typical agar plate (Sabet *et al.* 1984). Furthermore, repeated passage of the organism may introduce adaptive genomic changes not present in the original sample, as seen in laboratory strains of *Escherichia coli* (Liu *et al.* 2017). A DNA hybridisation technique to bypass these barriers has now been used for *Chlamydia* genomes for a number of species, including *C. trachomatis* (Christiansen *et al.* 2014), *C. psittaci* (Branley *et al.* 2016), *C. pneumoniae* (Roulis *et al.* 2015), and *C. pecorum* (Bachmann *et al.* 2015). In this method, 120-mer hybridisation probes are designed against a

reference genome to bind small sheared DNA fragments of the target species. Separation of target DNA from unbound DNA by magnetic beads allows the target species to be preferentially sequenced with Illumina HiSeq technology. This technology allowed sequencing of a further four *C. pecorum* genomes, detected in samples taken from one koala from New South Wales and one from South Australia (Bachmann *et al.* 2015). Each koala was found to be infected with two distinct *C. pecorum* genomes. These results show that infection with one *C. pecorum* strain does not preclude co-infection with another. This has previously been identified in *C. trachomatis*, with infection with one serovar not precluding future infection with another (Dean & Stephens 1994), and has a significant impact on potential vaccine design (Murthy *et al.* 2012). The authors produced and compared the core genome (~280 kb) of koala-origin *C. pecorum* genomes (the four identified through their study and the three previously sequenced) with livestock origin *C. pecorum* genomes (four of Australian origin and five international). Four of the koala origin genomes formed their own cluster, whilst the remaining three fell into one of two distinct clusters with ovine and bovine origin *C. pecorum*. This suggests that *C. pecorum* has undergone multiple cross species transmission events from livestock to koalas, although more research is required in this area to more fully elucidate the evolutionary phylogeny of *C. pecorum*.

## 1.6 Retrovirus in koalas

### 1.6.1 What is a retrovirus?

Retroviruses are positive sense RNA viruses. Each virion contains two single stranded copies of a positive sense RNA genome. The nucleic acid is contained within a capsid which is in turn surrounded by a viral envelope. The retrovirus replication cycle involves the conversion of viral RNA into linear double stranded DNA via the action of a reverse transcriptase, creating a provirus. The provirus enters nucleus of the host cell and integrates with the genomic DNA of the host using the retrovirus integrase protein. The virus then replicates using the host cell mechanisms (Coffin *et al.* 1997). The family *Retroviridae*, contains two subfamilies: *Orthoretrovirinae* and *Spumaretrovirinae*. The subfamily *Orthoretrovirinae* contain six genera: *Lentivirus*, as well as the *Alpha-*, *Beta-*, *Gamma-*, *Delta-* and *Epsilonretroviruses*. The koala retrovirus (KoRV) belongs to the *Gammaretrovirus* genus (Adams *et al.* 2017), however, at the time of its discovery, it was classed as 'mammalian type C retrovirus' (Pringle 1998). Genomes of gammaretroviruses contain three central genes: *gag*, *pol* and *env*. These are

central genes flanked by identical long terminal repeats (LTRs) at the 5' and 3' ends of the genome. The *gag* gene encodes the core structural proteins of the viral particle (Delchambre *et al.* 1989), the *pol* gene encodes the reverse transcriptase and integrase proteins (Baltimore 1970; Temin & Mizutani 1970; Schwartzberg *et al.* 1984), and the *env* gene encodes the envelope proteins responsible for attachment and entry (Weiss 1993). The LTRs are responsible for retrovirus gene expression and contain enhancer and promotor sequences such as a TATA box (Wilson *et al.* 2003). The DNA sequence of the LTRs can be used as a tool for evolutionary analysis of endogenous retroviruses (Johnson & Coffin 1999). When the 5' and 3' LTRs of a provirus initially form they are identical, and any SNPs between them in an endogenous virus are therefore a result of host genome mutations, rather than viral mutations (Ishida *et al.* 2015b). This, in combination with knowledge of host genome mutation rates, can be used to approximate when a retrovirus first entered the germline of its host.

### 1.6.2 KoRV - a history of discovery

The first references to leukaemia in koalas were almost concurrently reported by both Heuschele & Hayes (1961), and Backhouse & Bolliger (1961). However, it was Heuschele & Hayes that suggested the leukaemia resembled that of viral origin seen in laboratory animals, in respect to tumour morphology, ecology and cytology, although no investigation of viral causes was reported in this study. Later studies (Canfield *et al.* 1987) identified 13 cases of lymphoid neoplasia out of 344 koala necropsies performed, and suggested a potential viral cause. Subsequently, evidence that leukaemia in a koala was associated with a retrovirus was reported, with transmission electron microscopic identification of a type C retrovirus in bone marrow from a koala that from leukaemia (Canfield *et al.* 1988).

The first successful virus isolation of KoRV used koala peripheral blood mononuclear cells (PBMCs) from an animal that was not suffering from lymphoma or leukaemia, using a media containing 5% foetal bovine serum (FBS), 5% koala serum, mitogen concanavalin A and IL-2 as key ingredients, as PBMCs from animals diagnosed with lymphoma or leukaemia grew slowly and died quickly (Robinson *et al.* 1996). Viral presence was initially confirmed in the cultured PBMCs using electron microscopy, and then by PCR using primers targeting a conserved region of retrovirus DNA. The sequence was similar to other mammalian type C retroviruses, but curiously was most closely related to gibbon ape leukaemia virus (GALV), a gammaretrovirus of simians

(O'Brien *et al.* 1997). This was independently confirmed two years later (Martin *et al.* 1999b). Sequencing of the highly conserved *pol* gene from 23 taxa, showed that GALV and KoRV shared 93% amino acid identity and 85% nucleic acid identity in this region. This suggests that the divergence between the two viruses has been a recent occurrence, since two genotypes of GALV also have a nucleic acid similarity as low as 87% in this region (Delassus *et al.* 1989). The similarity between these two viruses that infect host species with no ecological overlap suggests the involvement of an intermediate host. Martin *et al.* (1999b) hypothesised rodents may play a role, based on findings that a cell line derived from south-east Asian mice (*Mus caroli*) contained endogenous retrovirus more closely related to GALV than other murine origin retroviruses (Lieber *et al.* 1975). The first full genome sequence of KoRV confirmed its earlier classification as a type C retrovirus (later renamed as the genus *Gammaretrovirus*) (Hanger *et al.* 2000). KoRV was found to be closely related to GALV, with a 78% nucleotide identity across the complete genome, compared to feline leukaemia virus (FeLV) (56% similarity), porcine endogenous retrovirus (PERV) (44% similarity) and murine leukaemia virus (MuLV) (57% similarity). No KoRV-like sequences were detected in a small number of marsupial samples included in the study. The full genome of KoRV is 8,431 bp in length, and has the standard structure of a virus within the *Gammaretrovirus* genus, as described in section 1.6.1.

Studies have attempted to identify the intermediate host between koalas and gibbons. A related retrovirus was detected in greater false vampire bats (*Megaderma lyra*), named *Megaderma lyra* retrovirus (MlRV) (Cui *et al.* 2012). The genomes of MlRV grouped phylogenetically with GALV and KoRV, suggesting a common ancestor. Clustering with MlRV, and thus also sharing a common ancestor with KoRV, was *Mus dunni* endogenous virus. This retrovirus was previously isolated from laboratory cell lines derived from the South Asian wild mouse, *Mus terricolor* (formally *Mus dunni)* (Miller *et al.* 1996). Analysis of 42 other different vertebrate species including 19 rodents, 10 flying vertebrates (bats and birds), six feral (to Australia) vertebrates and seven marsupials detected only one retrovirus (MbRV) in 17 samples from grassland melomys (*Melomys burtoni*; a rodent) (Simmons *et al.* 2014). The partial sequence of MbRV (2880 bp) has 93% nucleotide identity with GALV strain GALV-SEATO and 83% identity with KoRV. The virus has not been fully sequenced and as yet is unculturable, so it is not known whether the virus is replication competent. Koalas and grassland

melomys share an overlapping range along the north-east coast of Australia suggesting MbRV could help to identify the link between KoRV and GALV similarities. However, as gibbons and grassland melomys share no overlapping range there is likely additional intermediate hosts involved. A recent investigation of 26 rodent species from south-east Asia detected novel retrovirus in a newly identified *Melomys burtoni* subspecies in Indonesia which was genetically similar to MbRV (Alfano *et al.* 2016). This mounting evidence suggests that rodents may play a role in the cross-species transmission of retroviruses and origin of KoRV.

### 1.6.3 KoRV endogenisation and epidemiology

KoRV has been described as a retrovirus that is undergoing endogenisation (Tarlinton *et al.* 2006). Originally KoRV was detected in all sampled koalas, and thus assumed to be endogenous (Hanger *et al.* 2000). Viral insertion sites in Queensland koalas are inherited, and proviral DNA is present in koala sperm cells (Tarlinton *et al.* 2006), demonstrating that in northern koala populations KoRV has integrated into the koala genome. However, the virus is still present as an uncorrupted whole genome in the germ line and is therefore capable of producing exogenous infectious viral particles. Tarlinton *et al.* (2006) identified KoRV in all koalas tested in Queensland (98/98), but only in 36% (8/22) of Victorian koalas and no South Australian koalas (0/26). Thus the virus is exogenous in some koala populations, suggesting that KoRV infection could be a recent occurrence in koalas, possibly arising in the last few hundred years, and has been spreading from northern to southern populations over this time.

A more recent study examining KoRV prevalence nationally tested 708 koala samples from Queensland (277), New South Wales (100), Victoria (and its islands; 169) and South Australia (specifically Kangaroo Island, 162) (Simmons *et al.* 2012). The prevalence on mainland Victoria was 73.0% (65/89) whereas the combined prevalence on Victorian islands was only 26.6% (22/80). Sample location data are missing from 41% of the positive samples (36/87) detected in Victoria, and the two locations listed for the Victorian mainland are “Gippsland” and the “Strzelecki ranges”. Gippsland is a broad region of Victoria that contains the Strzelecki ranges, and it is unclear where the geographic boundary has been determined between the two locations in the study. Curiously, no positive samples were obtained from Phillip Island, which could be a result of the small sample size, as only 11 animals were sampled. Previous studies with a small sample size (24) detected no KoRV on Kangaroo Island (Tarlinton *et al.* 2006),

whereas 14% of the 162 tested in this larger study were positive for KoRV. However, if Phillip Island is free of KoRV, it could be vital for future population management. Due to the known history of translocation between islands, absence of KoRV on Phillip Island could help determine an accurate history of the spread of KoRV. These results support the hypothesis that KoRV is endogenous in northern populations but has a mixed presence in southern koalas as it spreads north to south. The authors suggest that the southern population could have a genetic advantage over the northern population in regards to KoRV susceptibility, although this would be in contrast to the paradigm suggesting that a decreased genetic diversity increases disease susceptibility (Spielman *et al.* 2004). The authors also suggest that the spread of the virus could be influenced by an arthropod vector that is more abundant in tropical regions. However retroviruses have only been shown to be mechanically spread by arthropods (Ooshiro *et al.* 2013), and is considered an unlikely method of transmission (Day *et al.* 2008). Despite these prior studies there is still a large gap in the knowledge regarding KoRV prevalence in southern New South Wales and geographically separate mainland Victorian populations. Figure 1.1 demonstrates the geographical location of areas where the prevalence of KoRV has been assessed in Australia, highlighting the large regions where the prevalence of KoRV is unknown.

Analysis of DNA extracted from museum samples suggests KoRV has been present in northern koalas since at least 1891 (Ávila-Arcos *et al.* 2013). The KoRV *pol* gene was amplified from three Queensland koala museum specimens estimated to have been collected sometime between 1870 and 1891, and one New South Wales museum specimen from 1971. In addition, KoRV DNA was also detected in 12 koala specimens from Queensland collected between 1904 and the 1980s. A New South Wales sample from 1891 was negative by *env* PCR. The only Victorian sample tested, from 1930, was negative for KoRV. However, no mitochondrial DNA was amplified from this sample either, indicating that ancient DNA was not recoverable rather than necessarily indicating the true absence of KoRV. As the vast majority of historical samples are from Queensland it is difficult to reach conclusions regarding the early spread of KoRV in koala populations. More recent work has attempted to assess ancient copies of the entire KoRV genome from a subsample of the same museum specimens, as well as from modern samples, using the technique of hybridisation capture (Tsangaras *et al.* 2014). Comparison of the historical KoRV genomes to those detected in modern koalas

revealed a relatively low level of diversity, as determined by the number of the number of polymorphisms detected. Importantly, this study also examined the koala genome sequences flanking the KoRV provirus, and found that only 7.5% (32/429) and 7% (23/331) of the flanking sequences at the 5' and 3' end (respectively) were shared between two or more koalas. The low percentage of identical integration sites fits the previous hypothesis that KoRV integration is a recent or ongoing event (Tarlinton *et al.* 2006). Only northern koala samples were analysed in this study, highlighting the need for future studies to also include samples from southern koalas.

The detection of KoRV in museum samples sets 126 years as the minimum length of time koalas have been exposed to KoRV (Ávila-Arcos *et al.* 2013). A maximum estimate of 22,200 - 49,900 years has been calculated based on comparison of the nucleotide sequences from the LTR regions of KoRV provirus and an assumption of the rate of mutation of koala DNA (Ishida *et al.* 2015b). At present it is not possible to further define the point at which koalas were first truly exposed to KoRV.

### 1.6.4 KoRV and disease in koalas

The most commonly reported disease associated with KoRV is lymphoid neoplasia, leading to leukaemia and ultimately death (Hanger *et al.* 2000; Tarlinton *et al.* 2005; Hanger & Loader 2014). This association is only loosely supported in the current KoRV literature, but is strongly influenced by the typical pathogenesis of gammaretroviruses in other species. The majority of gammaretroviruses discovered to date have been associated with neoplasia, including GALV (Kawakami *et al.* 1972), FELV (Jarrett *et al.* 1964), PERV (Strandström *et al.* 1974), and MuLV (Moloney 1960). Prior to the identification of KoRV as a possible causative agent, lymphoid neoplasia had been identified as an issue in wild and captive koalas (Canfield 1987; Canfield *et al.* 1987). Spencer and Canfield (1996) summarised the results of 583 koala necropsies between 1982 and 1994, 112 of which were captive koalas. In total, 5.3% (31/583) of koalas were found to be affected by lymphoid neoplasia, with 7.1% (8/112) of captive koalas affected. The majority (18/31) of cases were multicentric lymphosarcoma with the superficial lymph nodes, liver, spleen, and bone marrow the organs most commonly affected. There was no sex predilection in cases of lymphoid neoplasia, but cases were most commonly found in adult animals. Immunophenotyping determined that most (26/38) cases were T cell immunophenotype and 12 were B cell immunophenotype (Connolly *et al.* 1998). Lymphoid leukaemia occurred in 63% of cases and was most

commonly associated with multicentric tumours. In each of these studies, the link between retrovirus and the lymphosarcoma was suggested but evidence of virus infection in the affected animals was not presented.

The prevalence of lymphoid and associated neoplasias in modern koala populations is unclear. Although the study reporting the first full genome of KoRV asserted that lymphomas were the cause of death for 80% of captive koalas in Queensland, no supporting data were supplied. (Hanger *et al.* 2000). Similarly, Shojima *et al.* (2013) stated that 10% of captive koalas in Japan succumb to leukaemia, but no supporting data were reported. Other studies have also mentioned prevalence rates in captive populations without reference to primary data (Tarlinton *et al.* 2005; Fiebig *et al.* 2006). Only recently have studies attempting to verify these figures emerged, although as yet these results have not been peer reviewed. Two recent conference presentations reported rates of neoplasia in captive and/or wild koala populations. Gillett (2014) surveyed 16 Australian institutes with captive koalas and found that tumours of some description were associated with 56% of disease cases in these captive populations, with lymphoma the most common diagnosis. However, most (95%) of the disease cases assessed were from Queensland and New South Wales koalas. Hanger and Loader (2014) reported that 287/6001 (4.8%) of koalas admitted to a Queensland based animal hospital were diagnosed with some form of neoplasia, whilst neoplasia affected 7/296 (2.4%) of the wild free ranging Queensland koalas they assessed between 2008 and 2013. These results, although preliminary, do suggest that the prevalence of lymphoma may be below the 80% suggested previously by Hanger *et al.* (2000)

Some evidence for a link between KoRV and neoplasia was provided through the use of a quantitative real time reverse transcriptase PCR for KoRV (Tarlinton *et al.* 2005). This study quantitatively measured proviral DNA and KoRV genomic RNA copy numbers (as a measure of viraemia) in 90 Queensland koalas, including 63 captive animals and 27 wild animals that were presented to veterinary clinics. All koalas in the study, regardless of the presence or absence of disease, had some level of viral RNA in their plasma. There was no link between levels of proviral DNA and clinical disease, age, body mass, sex, or viral RNA levels. However, a significant increase in viral RNA copy numbers occurred in animals with lymphoma and leukaemia, as well as a positive association between viral RNA copy numbers and age. There was no significant association between white cell count and viral RNA levels, suggesting viral RNA levels

weren't merely increasing as a side effect of leukaemia (as leukaemic animals would be expected to have more nucleated cells in their blood). Perhaps the most crucial component of this study was the resampling of ten captive animals over the span of 18 months. In this data set, three animals that had high levels of viraemia at the initial sampling point died of neoplasia during the study, but no animals with low levels of viraemia died or were diagnosed with neoplasia. Although untangling the association between neoplasia/leukaemia and KoRV infection is difficult, this research suggests a link between circulating viral load and clinical disease.

### 1.6.5 KoRV genotypes and their potential role in disease

Since the first full genome of KoRV was sequenced (Hanger *et al.* 2000) a number of genomic variants have been identified. In 2013 100% of captive koalas in the San Diego Zoo (28/28) and Los Angeles Zoo (13/13) (LAZ), USA, were shown to be positive for KoRV (Xu *et al.* 2013). A unique *env* sequence was identified in six of the thirteen animals from LAZ, three of which died of lymphoma. Full genome sequencing demonstrated this variant differed most in the *env* gene and the U3 region of the LTRs when compared to the original full genome sequence of KoRV. The new genotype was named KoRV-B, leading to the original sequenced strain being renamed KoRV-A. The authors suggested that this KoRV-B genotype could be the cause of the lymphoma in affected koalas, but cautioned that more data would be required to truly elucidate the cause. Neoplasia, including lymphoma, osteosarcoma, and osteochondroma-like tumours of the skull have previously been described in association with the detection of retrovirus in captive koalas at San Diego Zoo (Worley *et al.* 1993).

KoRV was also isolated from PBMCs from five captive koalas held in Kobe Municipal Oji Zoo, Japan (Miyazawa *et al.* 2011). From the five KoRV isolates found in the Japanese cohort (one from each koala, OJ-1 to OJ-5), a follow up study sequenced the full genome of OJ-4, isolated from the only koala with clinical disease (anorexia and pleural effusion), which was then designated KoRV-J (Shojima *et al.* 2013). A broader survey of 51 captive koalas in Japanese zoos found that none of the Victorian origin koalas (0/11) carried the KoRV-J genotype, whereas 67.5% of northern origin koalas (27/40) did.

Both the USA and Japanese studies used interference assays in human and mouse cell culture to determine which receptor was used for cell entry by their novel virus variants.

Previous interference assays had determined that KoRV-A used PiT1 (Oliveira *et al.* 2006), which is the same receptor that GALV (O'Hara *et al.* 1990) and FeLV-B (Takeuchi *et al.* 1992) utilise. In contrast, both KoRV-B and KoRV-J were found to utilise the thiamine transport protein 1 (THTR1) receptor (Shojima *et al.* 2013; Xu *et al.* 2013), the same receptor used by FeLV-A (Mendoza *et al.* 2006). This change is potentially linked to a change in 35/40 amino acids in the receptor binding domain of the envelope protein of KoRV-B/J (Xu *et al.* 2013). This amino acid similarity between KoRV-B and KoRV-J, combined with the use of the same cell receptor, suggested that they belong to the same subgroup (Shimode *et al.* 2014), and standard nomenclature has changed to encompass them both as KoRV-B variants. Two other isolates of KoRV from Japan (Miyazawa *et al.* 2011), taken from the same animal, are now classified as KoRV-C and KoRV-D respectively, based on the divergence of their *env* gene from both the KoRV-A and KoRV-B genotypes (Shimode *et al.* 2014). KoRV-B has since been detected in the koala transcriptome, taken from a wild male koala in Queensland (Hobbs *et al.* 2014) as well as multiple captive koalas, both in Australia (Maher & Higgins 2016) and internationally (Fiebig *et al.* 2016).

It has been hypothesised that KoRV-B arises in koalas already infected with KoRV-A through either mutation or recombination of KoRV-A with an as yet unidentified endogenous retrovirus (Fiebig *et al.* 2016). All koalas carrying KoRV-B were also infected with KoRV-A, and a similar phenomenon (transduction of the *env* gene of endogenous retroviruses) has given rise to new FELV genotypes (Anai *et al.* 2012). The changes in both the receptor binding properties of KoRV, which allow superinfection of host cells with these genotypes, in addition to duplications of repeats in the LTRs, which provides enhanced expression of these genotypes, have parallels to other gammaretroviruses associated with leukaemia (Fan 2014). For example, MuLVs can cause tumorigenesis by insertional activation of proto-oncogenes. The virus itself does not carry an oncogene, but proviral integration occurs at insertion sites upstream from proto-oncogenes of the host (Hayward *et al.* 1981). Duplicate copies of a promoter region are introduced due to a mutation in the LTRs that enhances the transcription of the retrovirus and in turn upregulates transcription of these proto-oncogenes (Lenz *et al.* 1984). This LTR mutation is also present in KoRV-B, and an assessment of the promoter activity in KoRV-B determined that it was more active than in KoRV-A, suggesting it could replicate more efficiently (Shimode *et al.* 2014). In murine

retroviruses changes in the LTR region can influence the type of leukaemia that arises (Li *et al.* 1987; Short *et al.* 1987). Similar differences in oncogenic potential may exist with different genotypes and variants of KoRV, which may in turn explain the various forms of leukaemia identified in koalas.

Retroviruses have the capacity to evolve quickly due to relatively high rates of mutation in their RNA genomes. This is reflected in the discovery of many new KoRV variants in koala PBMCs and by deep sequencing of KoRV-positive cases (summarised in Table 1.3). KoRV-E and KoRV-F, which differ to KoRV-A and KoRV-B in the variable region A of the *env* gene, were detected by PCR using koala PBMCs from the same captive population from which KoRV-B was initially discovered (Los Angeles Zoo, USA) (Xu *et al.* 2015). KoRV-F has the same repeated 18 nt promotor in the LTR region as KoRV-B, but with five copies, rather than the four copies present in KoRV-B. Recent deep sequencing of KoRV samples identified an even greater number of genotypes, and variants within these types. Chappell *et al.* (2017) investigated KoRV genotypes in 18 koalas from south-east Queensland and found 108 novel KoRV envelope sequences. Based on phylogenetic clustering these sequences were classified as KoRV-A, B, D, F, G, H, and I. Similar findings by Ishida *et al.* (2015a), who used four koalas and identified 163 novel amino acid sequences in the p15E region of *env*, highlight the 'moving target' nature of KoRV research that suggests there is a wide range of as yet undiscovered genotypes. Importantly, with the exception of KoRV-B, no potential change in the pathogenesis of KoRV has been associated (as yet) with these novel variants, and investigation of receptor binding and promoter activity would be useful in future characterisation of these and other variants.

### 1.6.6 Transmission of KoRV

Methods of KoRV transmission are not well understood, partly because most populations studied to date have a 100% KoRV prevalence. Infectious GALV is shed by gibbons in faeces and urine, even when they are outwardly healthy (Kawakami *et al.* 1977), and infected cats shed FeLV in saliva, urine and faeces, with more active infection resulting in larger quantities of virus being shed (Cattori *et al.* 2009). There is potential for similar mechanisms of shedding in koalas. Wedrowicz *et al.* (2016) screened koala faeces for KoRV DNA, and detected virus in 60% of samples (24/40). This suggests KoRV may be shed in faeces, but the study did not distinguish between provirus (present in host cells shed in faeces) and infectious virus particles. Vertical

transmission of KoRV-B has been reported, with an infected female (mated to an uninfected male) birthing two infected joeys. Conversely an uninfected female, mated with an infected male, birthed two uninfected joeys (Xu *et al.* 2013) suggesting that sexual transmission may not always occur. KoRV has also been detected in the mammary transcriptome and milk proteome from the same infected female (Morris *et al.* 2016). KoRV transcripts, collectively, made up 3% of the mammary transcriptome and KoRV proteins were detected in both the early and late lactation phase milk samples, suggesting that KoRV could be transmitted via milk. Shedding of KoRV in urine or saliva has not yet been assessed.

## 1.7 Recent research and knowledge gaps

### 1.7.1 *C. pecorum* in Victorian koalas

Modern PCR techniques have only been used to assess the prevalence of *C. pecorum* in three Victorian koala populations. Patterson *et al.* (2015) tested urogenital swabs and determined the prevalence at Mt Eccles, western Victoria, as 25% (30/120) and the prevalence on Raymond Island, eastern Victoria, as 41% (43/104). Samples (n = 63) were also collected from French Island and no *Chlamydia* was detected. French Island has been considered free of chlamydial disease throughout its history, correlating with the highly fecund koala population on the island, with one study reporting that French Island koalas were 3 times more likely to produce offspring annually than koalas on nearby Phillip Island (Martin 1981). Complement fixation testing identified 6/87 (6.8%) French Island koalas and 27/30 (90%) Phillip Island koalas as having anti-chlamydial complement fixing antibodies (McColl *et al.* 1984). However, complement fixation tests for *Chlamydia* spp. are considered neither highly specific nor sensitive due to issues with cross-reactivity (Brown & Grice 1984; White & Timms 1994; Verkooyen *et al.* 1998), and are not commonly used when other more reliable methods are available.

Patterson *et al.* (2015) detected no ocular pathology associated with *C. pecorum* infection in Victorian koalas. This contrasts with northern states, where ocular infection with *C. pecorum* is common (Wan *et al.* 2011; Higgins *et al.* 2012; Funnell *et al.* 2013), and this difference in tissue tropism warrants further investigation. Genetic investigation of *C. pecorum* found in Victorian koalas, and comparison with northern studies, may assist in understanding this difference. Patterson *et al.* (2015) also found a positive association between the presence of ‘wet bottom’ and *C. pecorum* in the Raymond

Island koalas, but not for koalas at Mt Eccles. Previous studies have also found that clinical signs such as conjunctivitis and wet bottom alone are not considered accurate measures of *C. pecorum* presence (White & Timms 1994). Furthermore, the presence of wet bottom was comparable in the French Island koala population and the mainland populations, despite the absence of *C. pecorum* in French Island koalas test in that study. This suggests another aetiological agent may be present. Whilst *C. pecorum* is the most commonly investigated cause of urogenital infections in koalas, it is likely they are caused by a range of different opportunistic or specific pathogens, as in other animal species.

### 1.7.2 KoRV and disease

Whilst studies have suggested that the newly described genotypes of KoRV may be associated with disease (Shojima *et al.* 2013; Xu *et al.* 2013), the sample sizes of studies are low. Larger epidemiological studies of wild/free ranging koalas may reveal if different genotypes and/or variants of KoRV are more strongly associated with lymphoma. Furthermore, although some evidence suggests higher levels of viraemia are associated with neoplastic disease (Tarlinton *et al.* 2005), no detailed study has compared clinical signs of disease and the presence or absence of KoRV infection. Victorian koalas offer a unique opportunity to investigate potential associations, and in the process undertake a broad genotyping study, which may provide further information regarding the virulence of different KoRV genotypes. In addition to leukaemia, KoRV has been implicated as a potential cause of a series of conditions collectively known as "koala AIDS" (Hanger & Loader 2014). In these cases, koalas present with clinical signs including stomatitis, dermatitis, and chronic ill thrift, as well as opportunistic fungal infections. Such a "syndrome" is yet to be reported in Victorian koalas, but the ability to associate these clinical signs of disease with the presence or absence of KoRV will increase our understanding of any AIDS-like condition in koalas.

### 1.7.3 Wet bottom

As discussed previously, mild wet bottom has been detected in individual koalas or in koala populations where *C. pecorum* has not been detected (Patterson *et al.* 2015), suggesting that there may be a currently undetected pathogens in koala populations causing these clinical signs. Details of wet bottom or cystitis presence in free ranging populations of northern koalas is sparse, despite being listed in multiple publications as

a key sign of chlamydiosis (Brown *et al.* 1987; Canfield *et al.* 1991; Girjes *et al.* 1993). There is a wet bottom scoring system available, established by Flanagan (2009), which ranges from 0 (absent) to 10 (severe). This system was modified by Griffith (2010) to include a score of 0.5 in cases of discolouration of fur around the cloaca, where urine leakage and odour were absent. The scoring criteria described by Griffith (2010) is present in Appendix 2. A large number of studies which refer to wet bottom are investigating koalas admitted to wildlife hospitals (Markey *et al.* 2007; Griffith & Higgins 2012; Marschner *et al.* 2014), and do not report the range of scores encountered. The scoring system has been criticised by some due to its subjective nature, and the reliance on a range of factors that are difficult to measure such as odour (Griffith 2010). Another study attempting to associate wet bottom severity with *Chlamydia* load (quantified using qPCR) used a different system (ranging from 0 – 3) that included both outwardly presenting clinical signs and sonography of the urogenital tract, making it difficult to compare results to the previously established methods (Wan *et al.* 2011). Since the identification of *C. pecorum* as an important causative agent, no studies have sought to identify other causative agents of this clinical sign of disease, despite other bacterial organisms being detected in early work investigating urogenital tract disease of koalas (Canfield *et al.* 1983; Obendorf 1983). The advancement of sequencing technology may allow the detection of these potential aetiological agents.

### 1.7.4 Bacterial diversity studies to identify aetiological agents

Next generation deep sequencing technology has allowed the rapid advancement of bacterial studies. These studies utilise the 16S rRNA gene of bacteria to survey the bacterial communities present in a sample. Through the use of universal primers targeting multiple variable domains of the 16S rRNA gene, DNA can be amplified, sequenced and compared to curated bacterial databases. Such methods can be used to study the impact of diet on the gut flora (Turnbaugh *et al.* 2009), as well as help to investigate the role of microbiota in clinical syndromes such as inflammatory bowel disease (Morgan *et al.* 2012). In wildlife research, the tool can be a useful method of obtaining large quantities of data for minimal cost and from limited sampling. Previous methodology required the use of intensive bacterial culture techniques that overlook the true abundance of bacteria in a sample. Previous work (Osawa *et al.* 1993a) isolated (using culture-based methods) 39 bacterial species from faecal samples from 12 koalas held at Lone Pine Koala Sanctuary, Queensland. In contrast, Barker *et al.* (2013) used

454 pyrosequencing of regions of the 16S rRNA gene to identify 1855 bacterial species in the caecum, colon, and faecal samples from two wild koalas from Queensland. The majority of operational taxonomic units (OTUs), which are clusters of closely related sequences, were classified as *Firmicutes*, with *Bacteroidetes* making up the second most common phyla. More recent research utilising the superior Illumina sequencing chemistry supported this finding in part, with the *Firmicutes* making up the majority of bacterial species within faecal samples (Alfano *et al.* 2015). Whilst this research used captive koalas in Austria, statistical comparison of the phyla in these individuals and wild koalas (Barker *et al.* 2013), using Jaccard´s coefficient of similarity, showed no statistically significant difference. The study focused on microbiomes of the koala oral cavity, mid gut, hind gut, and faecal pellets and found that faecal pellets were a useful proxy for hind gut microbiomes. The ocular microbiome was also investigated, with *Proteobacteria* the most abundant phyla. No study to date has investigated the urogenital microbiome of female or male koalas, using either traditional culture techniques or modern bacterial diversity analysis.

The power of microbiome research to produce large quantities of data per sample may prove to be a useful tool for investigating potential aetiological agents of wet bottom in koalas not infected by *C. pecorum*. This technique has been used in analogous studies in humans, whereby the potential cause of urinary urgency incontinence was investigated (Pearce *et al.* 2014). In particular, *Gardnerella* sp. were found to occur at a higher relative abundance in women suffering from the syndrome, compared to clinically normal women, suggesting a possible role for this bacteria in the disease. Although microbiome studies will not definitively identify causative agents, association studies are a valuable tool to determine which potential causative agents should be investigated further.

## 1.8 Aims and objectives

Further work is required in order to better understand any link between KoRV infection and clinical signs of disease. Furthermore, a link between KoRV and *Chlamydia* has been suggested, but as yet there is no known association between dual infections and increased disease in koalas (Timms 2014). As all northern koalas are considered to be infected with KoRV, comparisons of presence/absence of infection with presence/absence of clinical signs cannot be undertaken. Similarly, links between the genotypes of *C. pecorum* and particular signs of clinical disease has not been

established in southern populations, with past research focussing on populations from New South Wales (Higgins *et al.* 2012) and Queensland (Jackson *et al.* 1997), and current research focusing on the success of vaccines against different genotypes of the pathogen (Kollipara *et al.* 2012; Kollipara *et al.* 2013a; Kollipara *et al.* 2013b; Kollipara *et al.* 2013c). At present, it is not known which genotypes of KoRV and *C. pecorum* are present in Victorian populations, nor whether there are other aetiological agents causing urogenital disease in koalas. Such studies present an opportunity to improve management actions within and between Victorian populations, as well as potentially determine whether genotypic differences result in differing disease presentations.

The aims and objectives of this work are therefore:

1) To undertake a broad survey of the prevalence of both *C. pecorum* and KoRV in Victorian koala populations, extending beyond the populations and sample sizes previously investigated.
2) To genotype the detected *C. pecorum*, utilising the well established molecular typing method of sequencing *ompA*. The pathogenicity of the *ompA* types detected in each population can then be assessed by comparing *ompA* genotype information and the associated clinical disease recorded.
3) To assess the presence of the reference KoRV genotype (KoRV-A) and the neoplasia-associated KoRV genotype (KoRV-B) in Victorian koalas. The clinical signs that are associated with the presence of the KoRV genotypes can then be examined using clinical records.
4) To investigate potential novel bacterial aetiological agents causing 'wet bottom' through the use of 16S rRNA diversity analysis. A study of the urogenital microbiome of female koalas has not previously been undertaken. This component of the study aims to establish foundational knowledge on the 'normal' microbiota, as well as determine potential bacterial causes of wet bottom.
5) To compare the full genomes of *C. pecorum* detected in koalas from across their entire Australian range. This will enable phlyogenetic comparison of *C. pecorum* from Queensland, New South Wales, Victoria and South Australia at a genomic level and will also allow genetic determinants of virulence and tissue tropism to be investigated.

**Table 1.1.** Species detected with *C. pecorum* and associated clinical disease

| Species | Common name | Clinical disease |
|---|---|---|
| **Eutherian** | | |
| *Bos taurus* | Cattle | Encephalomyelitis [a] |
| | | Polyarthritis [a] |
| | | Conjunctivitis [a] |
| | | Pneumonia [a] |
| | | Endometritis [a] |
| | | Abortion [a] |
| *Ovis aries* | Sheep | Polyarthritis [a] |
| | | Conjunctivitis [a] |
| | | Abortion [a] |
| | | Orchitis [a] |
| *Bubalus bubalis* | Water buffalo | Abortion [b] |
| *Sus domesticus* | Pig | Pneumonia [a] |
| | | Polyarthritis [a] |
| | | Enteritis [a] |
| | | Abortion [a] |
| *Capra aegagrus hircus* | Goat | Abortion [a] |
| *Capra ibex* | Alpine Ibex | No clinical signs [c] |
| *Cervus elaphus* | Red deer | No clinical signs [d] |
| **Metatherian** | | |
| *Phascolarctos cinereus* | Koala | Conjunctivitis [e] |
| | | Cystitis (wet bottom) [e] |
| | | Reproductive abnormalities [e] |
| | | Pneumonia [f] |
| *Petauroides volans* | Greater glider | None described [g] |
| *Trichosurus caninus* | Short-eared possum[#] | None described [g] |
| *Perameles bougainville* | Western barred bandicoot | None described [g] |
| **Avian** | | |
| *Cyanocompsa brissonii* | Ultramarine Grosbeak | No clinical signs [h] |
| *Gubernatrix cristata* | Yellow cardinal | No clinical signs [h] |
| *Paroaria coronata* | Red-crested cardinal | No clinical signs [h] |
| *Sicalis luteola* | Grassland yellow finch | No clinical signs [h] |
| *Turdus* spp. | Thrush/blackbird | No clinical signs [h] |
| *Nymphicus hollandicus* | Cockatiel | No clinical signs [h] |
| *Psittacula* spp. | Ring-necked parakeets | No clinical signs [h] |
| *Rhea* spp. | Rhea | No clinical signs [h] |

[#] Referred to as a mountain brushtail possum in reference.

[a] Yousef Mohamad *et al.* (2014); [b] Greco *et al.* (2008); [c] Holzwarth *et al.* (2011); [d] Regenscheit *et al.* (2012); [e] Polkinghorne *et al.* (2013); [f] Mackie *et al.* (2016) ; [g] Bodetti *et al.* (2003); [h] Frutos *et al.* (2015)

**Table 1.2.** Descriptive details of the house keeping genes used in the *Chlamydiales* multilocus sequence typing scheme (Pannekoek *et al.* 2008; Jelocnik *et al.* 2013)

| Gene | Protein | Gene length (bp) | | Number of allele types | |
|---|---|---|---|---|---|
| | | MLST fragment | Full length | *C. pecorum* specific | Total |
| *enoA* | Enolase | 381 | 1281 | 9 | 36 |
| *fumC* | Fumarate hydratase class II | 465 | 1383 | 2 | 28 |
| *gatA* | Glutamyl-tRNA amidotransferase subunit A | 425 | 1461 | 15 | 44 |
| *gidA* | Glucose-inhibited division protein A | 474 | 1839 | 12 | 45 |
| *hflX* | GTP binding protein | 435 | 1350 | 8 | 39 |
| *hemN* | Oxygen-independent coproporphyrinogen III oxidase | 432 | 1395 | 4 | 30 |
| *oppA*_3 | Oligonucleotide-binding protein | 483 | 1356 | 6 | 36 |

**Table 1.3.** Summary of KoRV genotype discovery

| KoRV type | Location first described | Also detected in Australia? | Receptor |
|---|---|---|---|
| A | Queensland, Australia [a] | – [o] | PiT1 |
| B | Kobe, Japan [#] [b] | Yes ^ | THTR1 |
| | Los Angeles, USA [c] | | |
| C | Kobe, Japan [d] | Yes [+] | Unknown |
| D | Kobe, Japan [d] | Yes [*] | Unknown |
| E | Los Angeles, USA [e] | No | Unknown |
| F | Los Angeles, USA [e] | No | Unknown |
| F' [&] | Queensland, Australia [f] | – | Unknown |
| G | Queensland, Australia [f] | – | Unknown |
| H | Queensland, Australia [f] | – | Unknown |
| I | Queensland, Australia [f] | – | Unknown |

[o] Sample originally detected in Australian koala population

[#] Originally designated KoRV-J (Shojima *et al.* 2013)

[^] Detected in a koala transcriptome project (Hobbs *et al.* 2014)

[+] Detected in samples from Kangaroo Island, SA (Young 2014)

[*] Detected through deep sequencing techniques (Chappell *et al.* 2017)

[&] Genotype F was classified by Xu *et al.* (2015), but the name reused by Chappell *et al.* (2017)

[a] Hanger *et al.* (2000); [b] Shojima *et al.* (2013); [c] Xu *et al.* (2013); [d] Shimode *et al.* (2014); [e] Xu *et al.* (2015); [f] Chappell *et al.* (2017)

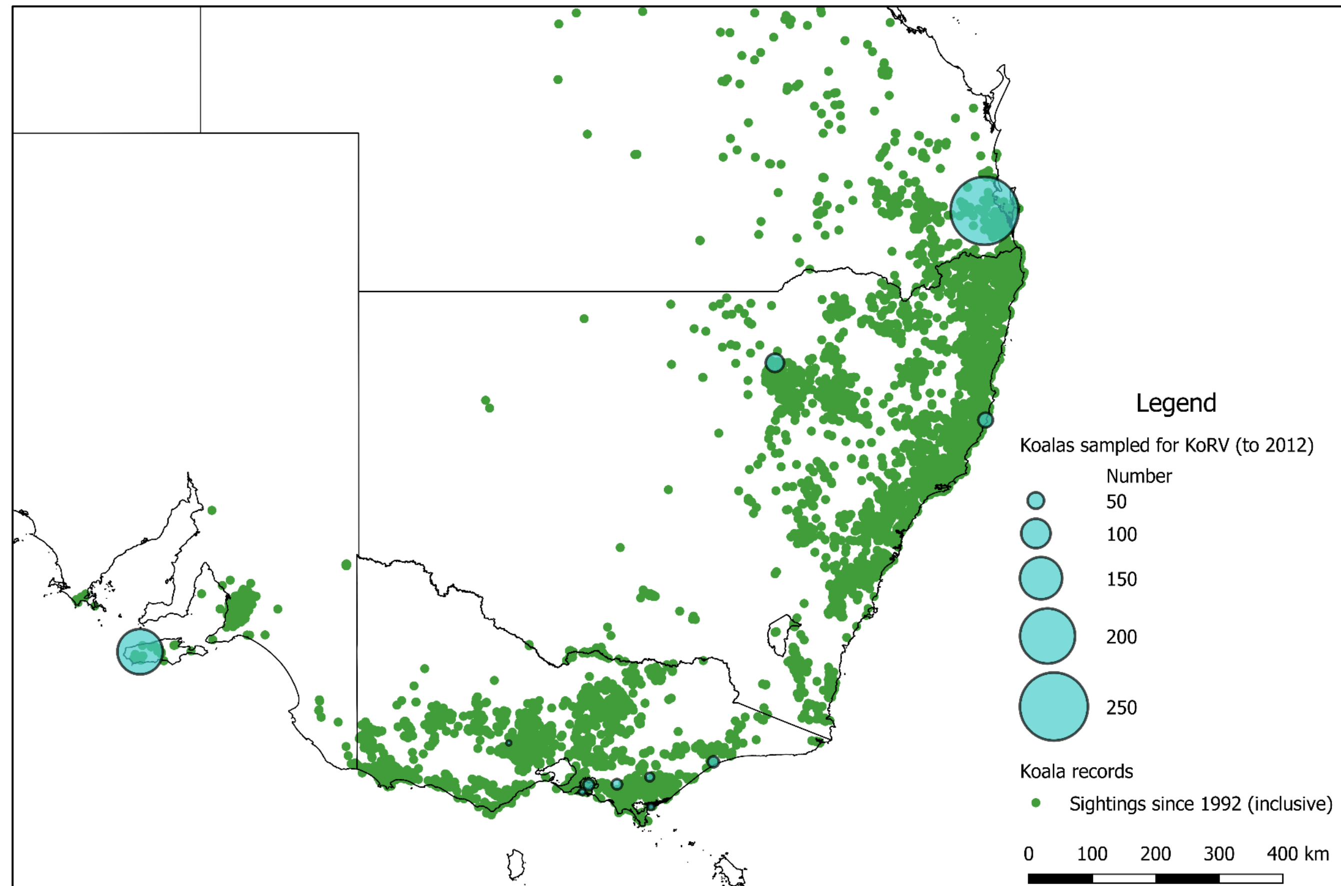



**Figure 1.1.** Map of koala sightings (green) since 1992 (~25 years) overlayed with KoRV sampling (blue) recorded by previous research (Tarlinton *et al.* 2006; Simmons *et al.* 2012). Koala sighting data downloaded from the Atlas of Living Australia (2017). Sampling locations are approximate, based on region/location names recorded in literature.

# 2. General Methodology

## 2.1 Koala sample collection

Samples for this study were sourced either from laboratory archives at the Asia Pacific Centre for Animal Health (APCAH) at the University of Melbourne, or collected by collaborators working with Zoos Victoria, the Department of Environment, Land, Water and Planning (DELWP), or Parks Victoria, during the course of the project. Archival samples were collected from past research field trips, management programs and post-mortem examinations, including those described in Patterson *et al.* (2015) and Stalder *et al.* (2015). For samples collected from live animals, during capture (where possible) animal health was assessed by a wildlife veterinarian and measures such as body condition score (Patterson *et al.* 2015), wet bottom score (Griffith 2010) and tooth wear class (Martin 1981) were recorded. Gross pathology of the urogenital tract was recorded when samples were collected at necropsy. For a subset of koalas surveyed in the field, ultrasound was used to assess the urogenital tract for any abnormalities including reproductive cysts and cystitis (Patterson *et al.* 2015). Sample collection was approved by the University of Melbourne Animal Ethics Committee (approval numbers 1011687.1 and 1312813.2) and DELWP (Research Permits 10004605, 10006948 and 10005388).

### 2.1.1 Koala data and clinical health

Koala-associated data were recorded for comparison with the level of pathogen detected. Data included sex and capture location (see Figure 2.1). Body condition score was measured based on palpation of the koala's scapula muscle, and ranged from 1 to 5. A score of 1 represents a severely emaciated koala in poor health; a score of 3 represents a well-conditioned koala of normal health; a score of 5 represents an over-conditioned koala (Patterson *et al.* 2015). Scores of 2 and 4 allow for intermediates between emaciated and over-conditioned koalas. Koala tooth wear was recorded as a proxy for koala age, as per Martin (1981) and McLean (2003). Koalas with a tooth wear score below III were classified as young/juvenile koalas; scores between III and V (inclusive) were classified as adult/mature adults; scores greater than V and up to the maximum score of VII were classified as senior adults/old adults. Wet bottom scores were allocated based on the system developed by (Flanagan 2009), and modified by Griffith (2010), which ranges from 0 – 10 (Appendix 2). Examples of this scoring

system include a score of 0.5, which represents discolouration of fur around the cloaca, a score of 4, which represent “greasy, stained fur around the cloaca, a strong odour, possible inflammation of the cloacal margins, and discharge containing urinary debris”, and scores greater than 7, which represent “progressive decline, possible infestation with maggots and ultimately death if intervention does not occur” (Griffith 2010). Not all veterinarians undertaking clinical health assessment were trained in this scoring system. In such cases, the presence or absence of wet bottom was recorded, but no associated score was recorded.

### 2.1.2 Ocular and urogenital/cloacal swabs for *Chlamydia* spp. testing

Ocular and urogenital/cloacal swab samples were collected by a veterinarian or field assistant using sterile plastic applicator rayon tipped swabs (Copan). For ocular swabs, a dry swab was used to sample the conjunctiva of koalas. For urogenital samples, the swab was inserted into the urogenital sinus (females) or into the penile opening (males). In some instances, a cloacal swab was taken rather than a direct urogenital swab, which involved swabbing inside the cloacal opening, without entering the urogenital tract. All samples were stored at 2 – 8°C in the field, before being transferred to -20°C for long term storage until further processing.

### 2.1.3 Blood and spleen samples for KoRV testing

Blood samples were collected by veterinarians into EDTA tubes to prevent clotting. Where possible, blood was separated in the field via centrifugation and separate fractions of plasma and buffy coat were removed and placed in 1.5 mL tubes containing RNA*later*® stabilising solution (Thermo Fisher Scientific). When centrifugation facilities were not available, whole blood was collected for downstream extraction of nucleic acid. For whole blood samples that were transferred to the laboratory within 24 hours of collection, Ficoll-Paque PLUS (GE Healthcare Life Sciences) separation of PBMCs was undertaken using the following methods. Blood samples were mixed 1:1 with cold phosphate buffered saline (PBS), containing 1% FBS, and layered onto an equal volume of equilibrated Ficoll-Paque PLUS. Centrifugation was performed at 400 × *g* at room temperature (RT) for 40 min, without braking. 1 mL of plasma was then collected and stored at -20°C, whilst 600 µL of the plasma fraction was added to 400 µL of RNA*later*® and stored at -20°C. The buffy coat layer was removed using a plastic transfer pipette and added to 3 volumes of cold 1% FBS/PBS. This was centrifuged

(1000 × *g*, RT, 15 min), the supernatant removed and the cells washed in 2 mL of cold 1% FBS/PBS. Following centrifugation (1000 × *g*, RT, 15 min) the supernatant was removed and the cells resuspended in 600 µL of RNA*later*® and stored at -20°C for future nucleic acid extraction.

Spleen tissue samples were collected from koalas during necropsy by a wildlife veterinarian, placed in 1.5 mL screw cap tubes and stored at -80°C for downstream extraction of nucleic acid.

## 2.2 Standard molecular biology techniques

### 2.2.1 Positive control samples for PCR and DNA extractions

Plasmids containing target genes (KoRV-*pol*, KoRV-A-*env*, β-actin and *ompA*, detailed in following sections) were generated for use as positive control samples for DNA extractions, as positive control template in PCR reactions, or to generate standard curves to quantitate DNA in test samples. Known positive samples from previous research (Patterson 2012) were used as template to amplify each target gene using PCRs described in section 2.3, 2.4 and 2.5. Products of the appropriate size were visualised after electrophoresis (section 2.2.2), the band excised and DNA was purified using the QIAquick Gel Extraction Kit (Qiagen), and concentration determined using a Nanodrop 1000 spectrometer (Thermo Fisher Scientific). The gene of interest was ligated into pGEM®-T vector (Promega) as per manufacturer's instructions, using an overnight incubation protocol and the plasmid transformed into either DH5α or JM109 electro-competent *Escherichia coli* cells. Briefly, 2 µL of ligation mixture and 40 µL of *E. coli* were mixed in a 1.5 mL tube and incubated on ice for up to 30 min. The transformation mix was then transferred to a 0.2 cm cuvette and electroporated using a Gene-Pulser (Bio-Rad Laboratories), set to 2.5 kV, 25 µF and 200 Ω, for ~4.5 s. The bacteria were then resuspended in RT Luria-Bertani (LB) broth, transferred to a 1.5 mL tube and incubated at 37°C for 1 hr. 100 µL of this culture was then spread onto LB agar plates containing ampicillin (100 µg/mL), X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) (20 µg/mL) and isopropyl β-D-1-thiogalactopyranoside (IPTG) (100 µM). After overnight incubation at 37°C, colonies were screened visually to allow colour (blue/white) selection, with up to eight white colonies picked using a plastic pipette tip and placed in individual McCartney bottles containing 5 mL LB broth with ampicillin (100 µg/mL). After overnight culture using shaking incubation at 37°C at 200

rpm, plasmid was extracted from the cultures using the Wizard Plus SV Miniprep DNA Purification Kit (Promega) per manufacturer's instructions. Confirmation of the presence of the target gene DNA within the resulting plasmid preparation was undertaken using Sanger sequencing. Sequencing PCRs used primers targeting plasmid regions flanking the inserted DNA (T7 and SP6, Table 2.1). Sequencing reactions were performed as per section 2.2.3. The resulting nucleotide sequence was confirmed to be the desired DNA insert using a BLAST query (Altschul *et al.* 1990) of the NCBI nucleotide database (Clark *et al.* 2016). Plasmids containing the target gene for the 16SG qPCR (section 2.3.2) were produced through prior work, as described in (Patterson 2012). A plasmid containing the entire KoRV-B-env gene was produced commercially by GenScript as no positive controls from naturally infected koalas were available.

### 2.2.2 Agarose gel electrophoresis

To visualise (when required) amplified PCR products, standard agarose gel electrophoresis was used. Gels were prepared using SYBR safe DNA gel stain in 0.5X Tris/Borate/EDTA (TBE) buffer (Invitrogen) with 1% agarose (Scientifix) and transferred to appropriately sized electrophoresis gel trays. Once set, gels were transferred to electrophoresis chambers, submerged in 0.5X TBE buffer, loaded with 3 - 5 µL of each sample and an electric field of 10 V/cm was applied for 30 – 60 min. Each gel was also loaded with either the Hyperladder™ 1 kb or 100 bp molecular weight marker (Bioline), as appropriate. Gels were then visualised using the ChemiDoc XRS+ System (Bio-Rad Laboratories).

### 2.2.3 Sanger sequencing of PCR products

Target PCR products for Sanger (dideoxy) sequencing were purified directly from the PCR reaction mix using the QIAquick PCR Purification Kit (Qiagen). In instances where multiple bands were present following agarose gel electrophoresis, the remaining PCR products were loaded on a 1% agarose gel as described in section 2.2.2 and the band of the correct size was viewed on a UV transilluminator and excised from the gel using a scalpel blade. This PCR product was then purified using the QIAquick Gel Extraction Kit per manufacturer's instructions. Purified DNA was quantified using a Nanodrop 1000 spectrophotometer, with the resulting concentration used to prepare sequencing reactions. Sanger sequencing of DNA was performed using the following

reaction: 0.125 U of BigDye Terminator (BDT) v3.1, 1× BDT dilution buffer, 250 nM primer, 1 – 10 µL of DNA template and Milli-Q filtered water (Millipore Corporation) ($MQH_2O$) to a final reaction volume of 20 µL. 100 ng/kb of the target region was used as template. Sequencing PCR cycling conditions were as follows: 94°C for 5 min, followed by 30 cycles of 94°C for 10 s, 50°C for 5 s and 60°C for 3 min.

Reactions were then cleaned and DNA precipitated using sodium acetate/ethanol to eliminate unincorporated dyes. 2 µL of 3 M sodium acetate (pH 5.6) was added to the reaction after transfer to a 1.5 mL tube, followed by the addition of 50 µL of chilled absolute ethanol. After vortexing (Xtron Vortex Mixer, Bartelt Instruments Pty Ltd) and overnight incubation at -20°C to allow the DNA to flocculate, samples were centrifuged (20,000 × *g*, 30 min, RT), the supernatant removed and the pellets resuspended by vortexing in 100 µL of 70% w/v ethanol. Following centrifugation (20, 000 × *g*, 5 min, RT) the supernatant was removed and the tubes pulse spun (20,000 × *g*, 5 – 10 s, RT). Any remaining supernatant was removed via micropipette and the DNA pellet air dried for 10 – 15 min before being submitted for fragment analysis via capillary electrophoresis at the Centre for Translational Pathology at the University of Melbourne.

## 2.3 Molecular testing for *C. pecorum*

### 2.3.1 DNA extraction from swab samples

The majority of swabs were stored as dry (no buffer or medium) samples at -20°C for up to three years before thawing and processing. A subset of 142 samples from French Island, Victoria, were stored in RLT buffer (Qiagen) with β-mercaptoethanol. For DNA extraction, the dry swab tips were cut with sterile (autoclaved) surgical scissors into 1.5 mL tubes containing 800 µL of either Tris-EDTA (TE) buffer (pH 8.0), PBS (PH 7.4), or RNA*later*®. The tubes were vortexed for 5 – 10 s before DNA was extracted using a Qiaxtractor robot (Qiagen) and the Qiaxtractor VX extraction kit. 200 µL of the buffer/swab solution was transferred to a 96-well lysis block and extractions carried out according to manufacturer's instructions, with 100 µL of Tris-HCl (pH 8.5) used to elute the DNA at the end of the extraction process. Each 96-well lysis plate contained one positive extraction control and twelve negative extraction controls. The positive extraction control was a diluted broth of transformed *E. coli* cells containing the target *C. pecorum* 16SG DNA fragment in the pGEM®-T plasmid. The negative extraction

controls were either sterile TE buffer, water or PBS. After extraction, the eluted DNA was stored at -20°C.

### 2.3.2 Detection of *C. pecorum* DNA using qPCR

Extracted DNA from swabs was tested for the presence of *C. pecorum* using the *Chlamydiaceae* 16SG qPCR first described by Robertson *et al.* (2009). This PCR targets a conserved region of the 16S rRNA gene in *Chlamydiaceae* species and allows for speciation based on the melt profile of the PCR product. Amplification of DNA was performed using the GoTaq Flexi DNA polymerase PCR kit (Promega) and fluorescence was detected using SYTO 9 green fluorescent nucleic acid stain (Life Technologies). Each reaction contained 2 µM of each forward and reverse 16SG primer (Table 2.1), 2 mM magnesium chloride ($MgCl_2$), 0.2 mM dNTP mix, 1× GoTaq Flexi Buffer, 10 µM SYTO 9, 1.2 U GoTaq DNA polymerase and 5 µL of extracted DNA template, with $MQH_2O$ making the reaction mix up to the required 25 µL. The qPCR cycling conditions were as follows: denaturing at 96°C for 2 min, followed by 40 cycles of denaturing at 96°C for 25 s, annealing at 58°C for 20 s and extension at 72°C for 20 s. After 40 cycles, a final extension at 72°C occurred for 2 min. Amplification and detection cycling was followed by the generation of a melt curve using the temperature range of 75°C to 95°C with fluorescence measurements recorded every 0.3°C increase. All qPCR reactions were performed using the Mx3000P qPCR System (Stratagene). Samples were deemed to be positive if DNA amplification was detected during the qPCR and if the melting temperature (measured by the dissociation curve) of the product was the same (+/– 1°C) as the positive control. Each qPCR contained three negative controls (Tris-HCl buffer, pH 8.5 or $MQH_2O$). The positive controls for the qPCR consisted of purified pGEM®-T containing the target 16SG region. To produce a standard curve, the concentration of a purified preparation of pGEM®-T containing the target DNA was determined using the Qubit dsDNA HS Assay Kit (Life Technologies) in conjunction with the Qubit 2.0 Fluorometer (Life Technologies). The purified plasmid was tested in triplicate and the average concentration was used to determine the total copy numbers per microlitre of sample, based on the following formula:

$$x = \frac{C \times 6.022 \times 10^{23}}{l \times 660 \times 10^{9}}$$

Where $x$ = double stranded DNA copy numbers

$C$ = Concentration (ng/μL) of plasmid DNA

$l$ = Length (bp) of plasmid (with insert)

Serial ten-fold dilutions of the plasmid ($10^7$ – $10^1$ copies per 5 μL) were then prepared using a Qiagility high precision automated PCR system (Qiagen) and used in triplicate as template in qPCR reactions to produce a standard curve.

### 2.3.3 Amplification of the *C. pecorum* MOMP gene, *ompA*, from *Chlamydia*-positive samples

Extracted DNA from swabs positive for *Chlamydia* using the 16SG qPCR (section 2.3.2) was then used as template in a conventional PCR to amplify *ompA*, the gene encoding the MOMP of *C. pecorum*. This PCR, first described by Kollipara *et al.* (2013b), targets the entire *ompA* gene and is specific to *C. pecorum*. Amplification was performed using the GoTaq Flexi DNA polymerase PCR kit. Each reaction contained the following: 1 μM of each forward and reverse *ompA* primer (Table 2.1), 2 mM $MgCl_2$, 0.2 mM dNTP mix, 1× GoTaq Flexi Buffer, 2.5 U GoTaq DNA polymerase, 10 μL of extracted DNA template, and $MQH_2O$ up to 50 μL. The PCR cycling conditions were as follows: denaturing at 95°C for 10 min, followed by 40 cycles of denaturing at 95°C for 1 minute, annealing at 56°C for 1 minute and extension at 72°C for 2 min. This was followed by a final extension at 72°C for 7 min. All PCR reactions were performed using either a DNA Engine Thermal Cycler or a T100 Thermal Cycler (both Bio-Rad Laboratories). Each PCR contained one positive template control and one negative template control. The positive template control was diluted purified pGEM®-T containing the target *ompA* gene, whilst the negative template control was $MQH_2O$. The amplified products were then used for DNA sequencing (section 2.2.3) and *ompA* genotyping (section 2.3.5)

### 2.3.4 Amplification of genes for MLST of koala *C. pecorum*

Extracted DNA from selected swabs that were positive for *C. pecorum* using qPCR was used as template in conventional PCRs to amplify the seven housekeeping genes used for the *Chlamydiales* MLST. Amplification was performed using the GoTaq Flexi DNA polymerase PCR kit. Each reaction contained the following: 300 nM of each forward and reverse gene specific primers (Table 2.1), 2 mM $MgCl_2$, 0.2 mM dNTP mix, 1× GoTaq Flexi Buffer, 1.25 U GoTaq DNA polymerase, 2 μL of extracted DNA template, and $MQH_2O$ up to a total volume of 25 μL. The PCR cycling conditions were as

follows: denaturing at 95°C for 5 min, followed by 40 cycles of denaturing at 95°C for 30 s, annealing for 30 s (temperatures as per Table 2.1) and extension at 72°C for 1 minute. This was followed by a final extension at 72°C for 7 min. All PCR reactions were performed using either a DNA Engine Thermal Cycler or a T100 Thermal Cycler. The amplified products were then used for DNA sequencing (section 2.2.3) and MLST (section 2.3.6)

### 2.3.5 *ompA* genotyping of *C. pecorum* detected in koala samples

The *ompA* DNA sequences obtained from Sanger sequencing (described in 2.2.3) were analysed using Geneious 7 software (Biomatters Ltd). Sanger sequencing was performed in forward and reverse directions using *ompA* primers, in addition to internal primers (Table 2.1) to obtain high quality consensus sequence. The raw sequences were trimmed using the default algorithm within Geneious (removing nucleotides with greater than 5% chance of error from 5' and 3' ends) and then aligned, allowing the complete *ompA* gene sequence to be assembled. These *ompA* sequences were then aligned, using the Clustal W algorithm (Thompson *et al.* 1994), to a series of near complete (1115 – 1121 bp) *C. pecorum ompA* sequences available in Genbank (Genotype A [accession number: KF150132], B [KF150133], E' [KF150134], F [KF150135], G [KF150137], H [KF150138], I [KF150139], J [KF150140], K [KF150141]). In addition to these near complete sequences in Genbank, portions of *ompA* genotypes C (382 bp) and D (198 bp) were entered manually based on previously published sequences (Jackson *et al.* 1997). An extended 854 bp version of genotype C was developed by aligning the 699 bp sequence Ko3, described by Higgins *et al.* (2012), with the smaller genotype C portion, describe by Jackson *et al.* (1997). These sequences overlapped by 227 bp with only two nucleotide differences. Genotypes were assigned to the new sequences obtained in this project based on similarity to published reference genotypes, with a nucleotide difference of greater than 1% from known genotypes considered the threshold for classifying a new genotype, as described previously (Kollipara *et al.* 2013b).

### 2.3.6 MLST of *C. pecorum* detected in koala samples

The sequences resulting from Sanger sequencing (described in 2.2.3) of the housekeeping genes that were amplified for MLST (see section 2.3.4) were assessed using Geneious 7 software. All sequences were trimmed using the default algorithm

within Geneious, and then forward and reverse sequences of each gene product were aligned using Geneious alignment tool. All genes were trimmed to their appropriate size for analysis (Table 1.2)(Jelocnik *et al.* 2013) and then concatenated in alphabetical order of gene name (*enoA*, *fumC*, *gatA*, *gidA*, *hemN*, *hflX*, *oppA*). The concatenated sequences were queried against an online MLST database (http://pubmlst.org/chlamydiales/) (Jolley & Maiden 2010) to identify SNPs within each gene, compared to previously published sequence types. The concatenated sequences were also assessed phylogenetically using Bayesian analysis, employed in the software package 'MrBayes' (Huelsenbeck & Ronquist 2001) following the method of (Jelocnik *et al.* 2013). In brief, concatenated sequences were aligned using Clustal W (Thompson *et al.* 1994), and a phylogenetic tree was built using the HKY85 nucleotide substitution model (Hasegawa *et al.* 1985), using the Markov chain Monte Carlo (MCMC) method, including two runs of 1,000,000 generations, with 4 heated chains, a burn-in length of 100,000, and a subsampling frequency of 100.

## 2.4 Molecular testing for KoRV

### 2.4.1 Extraction of DNA from blood samples

Available blood samples from koalas included whole blood, centrifuge-separated buffy coat, Ficoll-Paque PLUS-separated PMBCs, and plasma. All blood samples were stored at -20°C until DNA extraction. Extraction of DNA was performed using a Qiaxtractor robot, utilising the Qiaxtractor VX extraction kit. For this, 200 µL of each sample was transferred to a 96-well lysis block. Samples of whole blood were diluted 1:4 with PBS (pH 7.4) and samples of centrifuge-separated buffy coat, which were heavily contaminated with red blood cells compared to Ficoll-Paque PLUS-separated PMBCs, were diluted 1:1 with PBS (pH 7.4). Extractions were then carried out according to manufacturer's instructions, with 100 µL of Tris-HCl (pH 8.5) used to elute the DNA. Each 96-well lysis plate contained one positive extraction control and twelve negative extraction controls. The positive extraction control was a diluted broth of *E. coli* cells containing the target KoRV-*pol* DNA fragment cloned into the pGEM®-T plasmid. The negative extraction controls were $MQH_2O$. After extraction, eluted DNA was stored at -20°C.

### 2.4.2 Extraction of DNA from tissue samples

Spleen samples were stored at -80°C prior to processing. Samples were thawed and vigorously homogenised using a sterile plastic applicator, rayon tipped swab. Each swab was then placed in a 1.5 mL screw cap tube containing 800 µL of PBS. Each tube containing the swab/liquid combination was briefly vortexed and DNA extracted as per the protocol described in section 2.4.1. After extraction, eluted DNA was stored at -80°C.

### 2.4.3 Detection of KoRV using qPCR

The extracted DNA was tested for the presence of KoRV proviral DNA using qPCR, as previously described (Tarlinton *et al.* 2005). This PCR targets a conserved region of the KoRV-*pol* gene. Amplification was performed using the SYBR Green Platinum Master Mix (Applied Biosystems). Each reaction contained the following: 1 µM of each forward and reverse primer (Table 2.1), 2× SYBR Green Master Mix, 5 µL of extracted DNA template, and MQ$H_2O$ up to 20 µL. The qPCR cycling conditions were as follows: 50°C for 2 min, denaturing at 95°C for 5 min, followed by 40 cycles of denaturing at 95°C for 30 s, and a combined annealing and extension at 60°C for 30 s. Amplification and detection cycling was followed by the generation of a melt curve using the temperature range of 75°C to 95°C with fluorescence measurements recorded every 0.3°C increase. All qPCR reactions were performed using the Mx3000P qPCR System. Samples were deemed to be positive if DNA amplification was detected during the qPCR and if the melting temperature (measured by the dissociation curve) of the product was the same (+/– 1°C) as the positive control. Each qPCR contained three non-template controls (Tris-HCl buffer, pH 8.5). A standard curve was produced using seven 10-fold dilutions of a positive control ranging from $10^7$ to $10^1$ copies/reaction, in triplicate. The positive control consisted of purified pGEM®-T containing the target KoRV-*pol* region.

### 2.4.4 KoRV genotyping using KoRV-A and KoRV-B specific qPCRs

Genotyping qPCRs for KoRV-A and KoRV-B have previously been established using hydrolysis probes specific to each genotype (Xu *et al.* 2013). These qPCRs were applied, with some modifications, to the samples that were positive for KoRV DNA using the PCR targeting the KoRV-*pol* gene (section 2.4.3). Genotyping qPCRs were performed using the SYBR Green Platinum Master Mix. Each reaction contained the

following: 0.5 µM of each forward and reverse primer (Table 2.1), 2× SYBR Green Master Mix, 2 µL of extracted DNA template, and $MQH_2O$ up to 20 µL. The qPCR cycling conditions were as follows: 50°C for 2 min, denaturing at 95°C for 5 min, followed by 40 cycles of denaturing at 95°C for 30 s, and a combined annealing and extension at 60°C for 30 s. Amplification and detection cycling was followed by the generation of a melt curve using the temperature range of 75°C to 95°C with fluorescence measurements recorded every 0.3°C increase. All qPCR reactions were performed using the Mx3000P qPCR System. Samples were deemed to be positive if DNA amplification was detected during the qPCR and if the melting temperature (measured by the dissociation curve) of the product was the same (+/– 1°C) as the positive control. Each qPCR contained three non-template controls (Tris-HCl buffer, pH 8.5). A standard curve was produced using seven 10-fold dilutions of a positive control ranging from $10^7$ to $10^1$ copies/reaction, in triplicate. The positive control consisted of either purified pGEM®-T containing the target KoRV-A-*env* region, or purified pUC17 vector plasmid containing the whole KoRV-B-*env* gene. Determination of the concentration of the control plasmid, and calculation of the DNA used in the standard curve, was performed as described for *C. pecorum* qPCR (section 2.3.2).

## 2.5 Normalisation of qPCR results using β-actin copy numbers

Following DNA extraction from swabs or blood, samples were assessed for the presence of host cells using a qPCR targeting the koala housekeeping gene encoding β-actin, as first described by Shojima *et al.* (2013). Amplification and detection was performed using SYBR Green Platinum Master Mix. Each reaction contained the following: 1 µM of each forward and reverse primer (Table 2.1), 2× SYBR Green Master Mix, 2 µL of extracted DNA template, and $MQH_2O$ to a final reaction volume of 20 µL. The qPCR cycling conditions were as follows: 2 min at 50°C followed by initial denaturing at 96°C for 2 min, followed by 40 cycles of denaturing at 96°C for 25 s, and a combined annealing/extension step at 60°C for 30 s. Amplification and detection cycling was followed by the generation of a melt curve using the temperature range of 75°C to 95°C with fluorescence measurements recorded every 0.3°C increase. All qPCR reactions were performed using the Mx3000P qPCR System. Samples were deemed to be positive if DNA amplification was detected during the qPCR and if the melting temperature (measured by the dissociation curve) of the product was the same (+/– 1°C) as the positive control. Each qPCR contained three non-template controls

(Tris-HCl buffer, pH 8.5). A standard curve was produced using serial 10-fold dilutions ($10^7$ to $10^1$ copies/reaction), in triplicate, of purified pGEM®-T containing the target β-actin region. Determination of the concentration of the control plasmid, and calculation of the DNA used in the standard curve, was performed as described for *C. pecorum* qPCR (section 2.3.2). Copy numbers detected in qPCRs for KoRV and *Chlamydia* were standardised between samples by dividing their raw values by the copy numbers of β-actin from the same DNA extracts.

**Table 2.1.** Primer sequences utilised in the described studies

| Target gene | Primer Name | Sequence (5' – 3') | Product size (bp) | Annealing temp. (°C) | Reference |
|---|---|---|---|---|---|
| **pGEM-T plasmid** | | | | | |
| Insertion region | T7 | TAATACGACTCACTATAGGG | 161 | 55 | Promega, USA |
| | SP6 | TATTTAGGTGACACTATAG | | | |
| **Prokaryotes** | | | | | |
| 16S rRNA | PRK341F | CCTAYGGGRBGCASCAG | ~465 | 56 | Yu *et al.* (2005) |
| | PRK806R | GGACTACNNGGGTATCTAAT | | | and Xu *et al.* (2012) |
| **Koala** | | | | | |
| β-actin | KoBactin.F | GAGACCTTCAACACCCCAGC | 111 | 60^ | Shojima *et al.* (2013) |
| | KoBactin.R | GTGGGTCACACCATCACCAG | | | |
| ***Chlamydia* spp.** | | | | | |
| 16S rRNA | Chlam.16SG.F | TGATGAGGCATGCAAGTC | ~460 | 58 | Robertson *et al.* (2009) |
| | Chlam.16SG.R | TTACCTGGTACGCTCAAAT | | | |
| ***C. pecorum*** | | | | | |
| *ompA* | Cpec.ompA.F | ATGAAAAAACTCTTAAAATCGG | ~1170 | 56 | Kollipara *et al.* (2013b) |
| | Cpec.ompA.R | TTAGAATCTGCATTGAGCAG | | | |
| | Bov.Cpec.intR | TTAGTGAACCATTCTGCATC | -# | | This project |
| | Cpec.ompAint2.F | AGAGCTACTTTTGATGCAGA | -# | | This project |
| | Cpec.ompAint.R | TTTGTGAACCACTCCGCATC | -# | | This project |
| *gatA* | Cpec.MJgatA1.F | GCTTTAGAGTTGAGAGAAGCT | 512 | 54 | Pannekoek *et al.* (2008) |
| | Cpec.MJgatA1.R | GATCCTCCTGTATCTGATCC | | | |
| *oppA*_3 | Cpec.MJoppA1.F | ATGTGCAAGATCCCAGTGGG | 605 | 58 | Pannekoek *et al.* (2008) |
| | Cpec.MJoppA1.R | GGCGCTACTTGTTATGGG | | | Jelocnik *et al.* (2013) |
| *hflX* | Cpec.MJhflX1.F | TGAGGAGATCTCTGCATCG | 607 | 58 | Jelocnik *et al.* (2013) |
| | Cpec.MJhflX1.R | ATCTTCATGCAAAGCAGCC | | | |
| *gidA* | Cpec.MJgidA1.F | GCGTCACAACAAAAGAAGGC | 560 | 60 | Jelocnik *et al.* (2013) |
| | Cpec.MJgidA1.R | TGACGCTGTATATCACACGG | | | |
| *enoA* | Cpec.MJenoA1.F | CCTATGATGAACCTTATCAATGG | 431 | 58 | Pannekoek *et al.* (2008) |

| Target gene | Primer Name | Sequence (5' – 3') | Product size (bp) | Annealing temp. (°C) | Reference |
|---|---|---|---|---|---|
| | Cpec.MJenoA1.R | TCTTCCTCCGCTAAGCCATCC | | | |
| *hemN* | Cpec.MJhemN1.F | GATCGCGATAGAGATAGACCC | 634 | 54 | Jelocnik *et al.* (2013) |
| | Cpec.MJhemN1.R | ATCTTCTCCTGATAGATATCG | | | |
| *fumC* | Cpec.MJfumC1.F | TGATTAAGAAATGTGCAGC | 572 | 54 | Jelocnik *et al.* (2013) |
| | Cpec.MJfumC1.R | CCTTCAGGTACATTAAGCC | | | Pannekoek *et al.* (2008) |
| **KoRV** | | | | | |
| *pol* | KoRV.pol.F | TTGGAGGAGGAATACCGATTACAC | 110 | 60^ | Tarlinton *et al.* (2005) |
| | KoRV.pol.R | GCCAGTCCCATACCTGCCTT | | | |
| **KoRV-A** | | | | | |
| *env* | KoRV.envA.F | GCCAGGCCCCCTGATTCAA | 117 | 60^ | Xu *et al.* (2013) |
| | KoRV.envA.R | GCACACGTAGAACTGGGACC | | | |
| **KoRV-B** | | | | | |
| *env* | KoRV.envB.F | GCCAGAATCTCAACAGTCTGC | 157 | 60^ | Xu *et al.* (2013) |
| | KoRV.envB.R | GGGACACACATAGAACTGAGATTG | | | |

# Internal primer used for Sanger sequencing only

^ Combined annealing and extension temperature

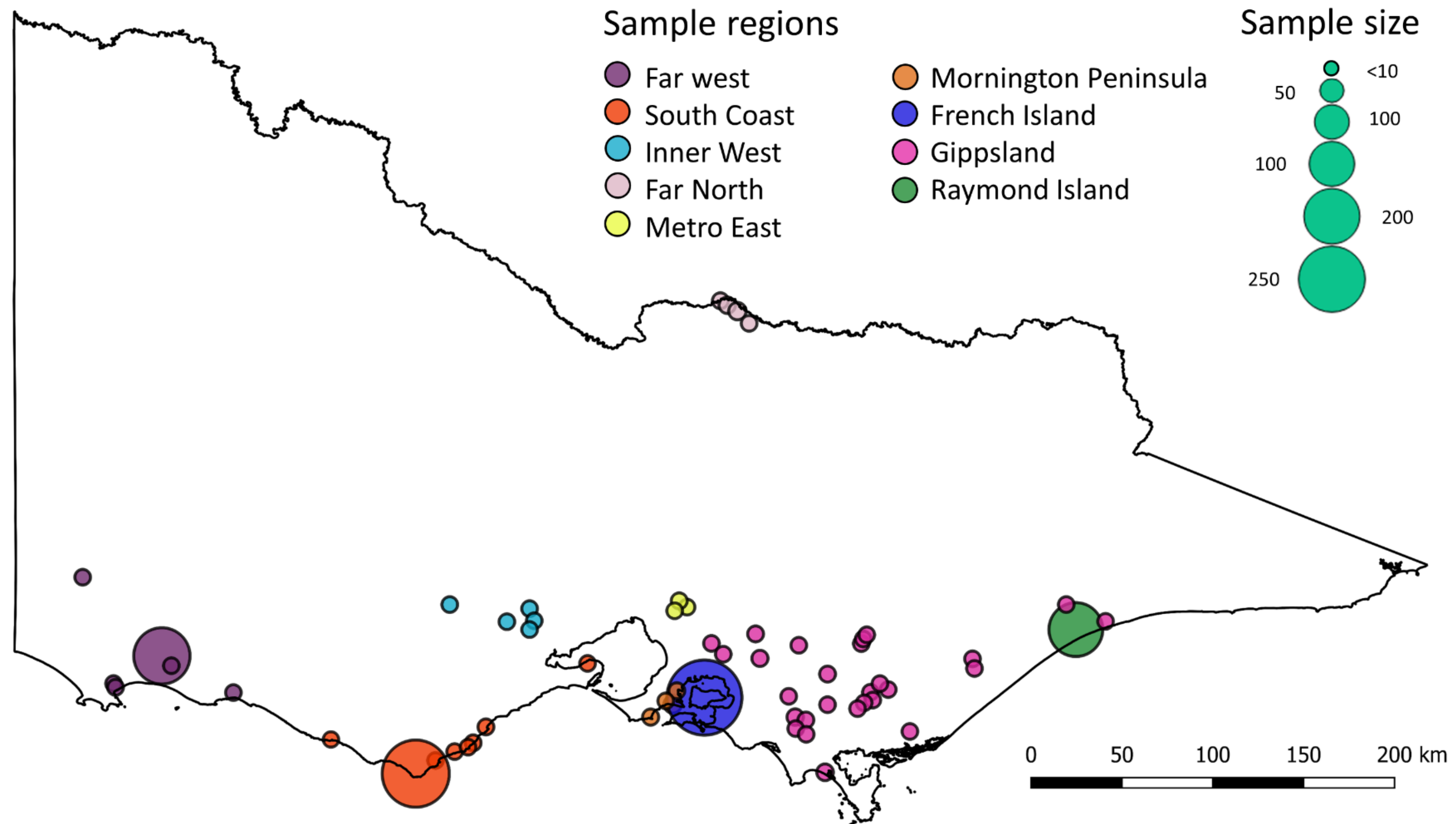



**Figure 2.1.** Map of the state of Victoria, Australia, overlayed with locations that koalas were sampled from, colour coded by region. Size of circle scales with the number of samples taken at the location, with the centre of each circle marking approximate sample location.

# 3. *Chlamydia pecorum* Infection in free-ranging koalas (*Phascolarctos cinereus*) on French Island, Victoria, Australia

**Published by:** Journal of Wildlife Diseases, volume 52, issue 2, 2016.

**Co-authors:** J. Amery-Gale, M. Lynch, L. Haynes, J. R. Gilkerson, F. M. Sansom and J. M. Devlin

The original publication can be obtained from the below DOI:

dx.doi.org/10.7589/2015-10-276

The following material has been included with permission from the journal.

DOI: 10.7589/2015-10-276

*Journal of Wildlife Diseases*, 52(2), 2016, pp. 426–429
© Wildlife Disease Association 2016

# *Chlamydia pecorum* Infection in Free-ranging Koalas (*Phascolarctos cinereus*) on French Island, Victoria, Australia

**Alistair R. Legione,[1,7] Jemima Amery-Gale,[1,2] Michael Lynch,[3] Leesa Haynes,[4] James R. Gilkerson,[5] Fiona M. Sansom,[1,6] and Joanne M. Devlin[1,6]** [1]Asia Pacific Centre for Animal Health, The University of Melbourne, Cnr. Park Dr. and Flemington Rd., Parkville, Victoria 3052, Australia; [2]Australian Wildlife Health Centre, Healesville Sanctuary, Badger Creek Rd., Healesville, Victoria 3777, Australia; [3]Veterinary Department, Melbourne Zoo, Elliott Ave., Parkville, Victoria 3052, Australia; [4]Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, 250 Princes Hwy., Werribee, Victoria 3030, Australia; [5]Centre for Equine Infectious Diseases, The University of Melbourne, Cnr. Park Dr. and Flemington Rd., Parkville, Victoria 3052, Australia; [6]These authors contributed equally; [7]Corresponding author (email: alegione@student.unimelb.edu.au)

ABSTRACT: We detected *Chlamydia pecorum* in two koalas (*Phascolarctos cinereus*) from a closed island population in Victoria, Australia, previously free of *Chlamydia* infection. The *ompA* and multilocus sequence type were most closely related to published isolates of livestock rather than koala origin, suggesting potential cross-species transmission of *C. pecorum*.

*Chlamydia pecorum*, originally identified in cattle (Fukushi and Hirai 1992), is associated with ocular and urogenital pathology in koalas (*Phascolarctos cinereus*; Glassick et al. 1996). The capability for *C. pecorum* to cross species from livestock to koalas has been hypothesized but no direct link has been established. This hypothesis is based on *ompA* sequence similarity between isolates of different host origin (Jackson et al. 1997; Kaltenboeck et al. 2009). More recently, multilocus sequence typing (MLST) has placed *C. pecorum* from koalas into a cluster within a greater livestock clade (Jelocnik et al. 2013).

The koala population on French Island, Victoria, Australia (38°21′0″S, 145°22′12″E), established in 1898 by translocation of koalas from the mainland, has long been considered to be free from *Chlamydia* (Martin and Handasyde 1999). French Island koalas are highly fecund compared to other populations (61% compared to 22% on Phillip Island; Martin 1981) and postmortem necropsies found no evidence of urogenital pathology in this population compared to other Victorian koalas (Obendorf 1981, 1983). In 2010, no *Chlamydia* spp. or significant reproductive tract disease was found in French Island koalas (Patterson et al. 2015).

After the Victorian koala population declined dramatically in the late 1890s, the French Island population was initially used to reestablish populations across Victoria (Martin and Handasyde 1999). However, it is thought that during the 1900s, koalas carrying *C. pecorum* were moved from a remnant Gippsland population to Phillip Island, and from there the pathogen was spread via further translocation to the rest of Victoria (Martin and Handasyde 1999). No koalas were subsequently moved to French Island from mainland Victoria, preventing transfer of *Chlamydia* to this population.

During population health management in 2011, urogenital swabs and clinical data were collected from 142 female koalas on French Island (approved by the University of Melbourne Animal Ethics Committee, 1011687.1). Swabs were stored at –20 C until DNA extraction using a QIAGEN Xtractor (QIAGEN, Hilden, Germany). Samples were tested for *Chlamydia* using quantitative PCR targeting the 16S rRNA gene (Patterson et al. 2015). Multilocus sequence typing and *ompA* PCR were applied to positive samples (Jelocnik et al. 2013; Kollipara et al. 2013). Products were purified (QIAquick Gel Extraction Kit; QIAGEN) and sequenced using BigDye® Terminator v3.1 (Life Technologies, Carlsbad, California, USA). For sequence and phylogenetic analysis we used Geneious 7 (Biomatters Ltd., Auckland, New Zealand) and compared to sequences in GenBank or PubMLST (Jolley

TABLE 1. Summary of clinical data and multilocus sequence typing housekeeping gene allele types (Pannekoek et al. 2008) from two *Chlamydia pecorum*–positive koalas (*Phascolarctos cinereus*) detected on French Island, Victoria, Australia, 2011.

| Animal ID | Weight (kg) | Body condition score[a] | Wet bottom score[a] | Offspring | Multilocus sequence type allele number[b] | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | enoA | fumC | gatA | gidA[c] | hemN | hflX | oppA |
| FrIs/K20 | 8.0 | 2.5 | 3 | Pouch young | 20 | 9 | 21 | 23 ($^{219}$C » T) | 8 | 10 | 12 |
| FrIs/K26 | 7.3 | 2 | 1 | None | 20 | 9 | 21 | 23 ($^{219}$C » T) | 8 | 10 | 12 |

[a] Methodology from Patterson et al. (2015). Body condition was scored from 1 (emaciated) to 5 (excellent). Wet bottom was scored from 0 (absent) to 10 (severe).

[b] enoA = enolase; fumC = fumarate hydratase class II; gatA = glutamyl-tRNA amidotransferase subunit A; gidA = glucose-inhibited division protein A; hemN = oxygen-independent coproporphyrinogen III oxidase; hflX = GTP-binding protein; oppA = oligonucleotide-binding protein.

[c] Closest allele type, with single-nucleotide polymorphism represented in parentheses.

and Maiden 2010) archives (*ompA* and MLST genes, respectively). Phylogenetic analysis of the MLST concatenated sequence followed the method of Jelocnik et al. (2013).

Two koalas were positive for *C. pecorum*, from which identical *ompA* nucleotide sequence and MLST housekeeping gene sequences were obtained. The *ompA* gene shared 100% nucleotide identity to a cattle isolate from Japan (strain Maeda; GenBank accession AB512085) and had 98% identity to European livestock isolates 3638/3, iC2, and iC4 (GenBank accessions GQ228186, EU684931, and EU684933, respectively). The closest nucleotide identity to a koala isolate was 86% (genotype I; GenBank accession KF150139). The housekeeping gene sequence type (ST) was novel, because of a synonymous single-nucleotide polymorphism in the glucose-inhibited division protein A gene (Table 1). The phylogenetic analysis output clustered the novel ST with livestock origin isolates (Fig. 1).

French Island has been home to livestock since settlement in 1850 (Department of Environment, Land, Water and Planning 2011). Cattle (*Bos taurus*) and sheep (*Ovis aries*) are farmed on the island. Feral pigs (*Sus scrofa*), goats (*Capra aegagrus hircus*), and Sambar deer (*Rusa unicolor*) are also present. The similarity of the *ompA* and MLST sequences to livestock, rather than koala, *Chlamydia* sequences, as well as the fact that this is a closed koala population, suggests cross-species transmission has occurred. The low prevalence may represent isolated transmission events, or the pathogen may have recently infected the koala population. Both koalas displayed signs of urinary incontinence and were in poor body condition. One had pouch young, suggesting fertility may not yet have been affected (Table 1).

Detection of similar *C. pecorum* types in livestock and koalas has been assessed using techniques such as partial *ompA* sequence comparison and MLST (Jackson et al. 1997; Jelocnik et al. 2013). The discovery of *C. pecorum* in a closed koala population previously considered free from *Chlamydia*, with a novel ST clustering with livestock STs, strongly supports the hypothesis of pathogen transmission from livestock to koalas. Koala and livestock ranges overlap on French Island, with koalas sampled in this survey along roadsides next to paddocks. The *C. pecorum* infection status of French Island livestock is currently unknown, and no sequences of *C. pecorum* isolated from Victorian livestock are available. The only Australian livestock isolates publically available originate from Queensland, New South Wales, and Western Australia. Future work testing Victorian livestock, including French Island livestock, would enhance our understanding of potential cross-species transmission events. Whole-genome sequencing of *C. pecorum* from Victorian koala and livestock would also increase

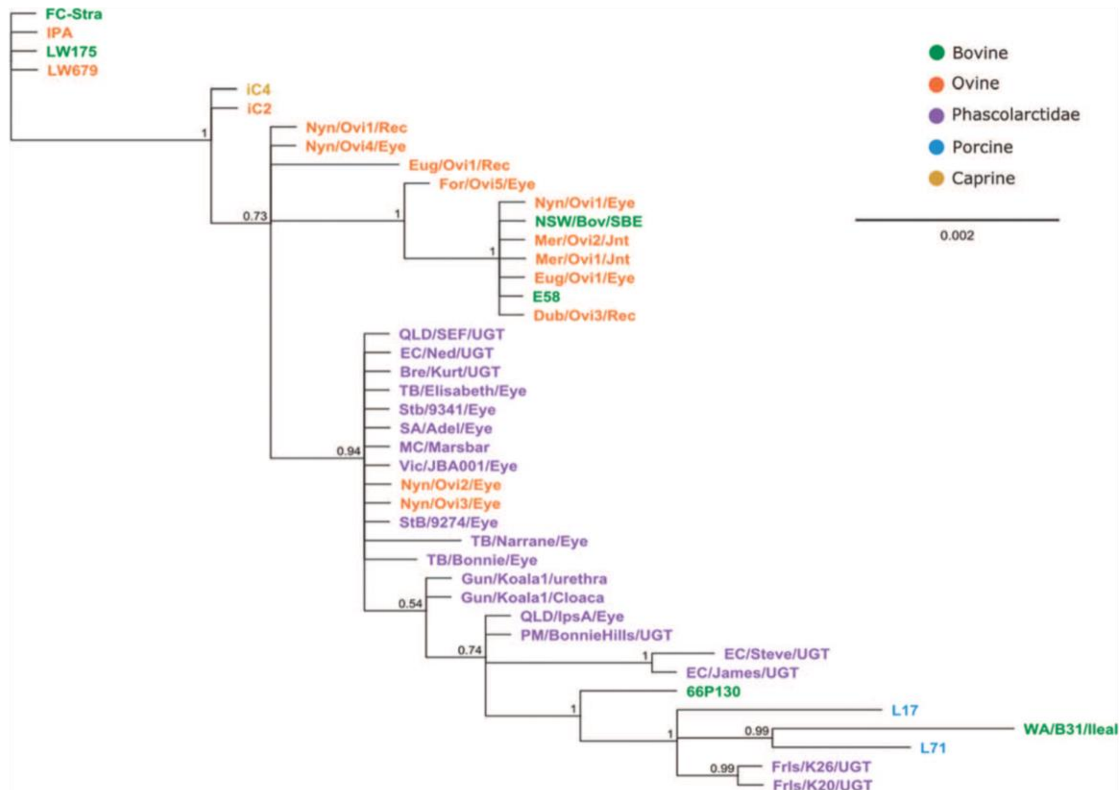



FIGURE 1. Bayesian phylogenetic analysis (Huelsenbeck and Ronquist 2001) of concatenated sequences of seven housekeeping gene fragments of French Island koala (*Phascolarctos cinereus*) *Chlamydia pecorum* (FrIs/K20/UGT, FrIs/K26/UGT), French Island, Victoria, Australia, 2011. Other sequences, obtained from PubMLST (Pannekoek et al. 2008), included in analysis were ovine/caprine isolates with high *ompA* similarity (iC2 and iC4), European porcine isolates (L71 and L17), North American cattle isolates (66P130 and E58), and Australian livestock and koala isolates. Tree node labels are posterior probabilities. Outgroup consists of livestock isolates Fc-Stra, IPA, LW623, and LW679 as per methodology of Jelocnik et al. (2013).

our understanding of the relationship between strains infecting different species.

The detection of *C. pecorum* in koalas on French Island may have significant conservation impacts. French Island koalas have previously been used as a source population for southern Australia. The detection of a *C. pecorum* type not previously associated with koalas has implications for any future translocations. There are potential risks in introducing novel *C. pecorum* types to koala populations. Previous research has highlighted the range of pathogenicity of *C. pecorum* isolates from livestock (Mohamad et al. 2014), but a greater understanding of koala *C. pecorum* types and their pathogenicity is required before the impact of this finding can be fully known. The discovery of *C. pecorum* in this koala population also provides a unique opportunity to understand the clinical significance of infection in a previously naive population.

Alistair Legione is supported by an Australian Postgraduate Award and the Holsworth Wildlife Research Endowment. We thank everyone involved in the collection of koala samples, particularly Parks Victoria.

## LITERATURE CITED

Department of Environment, Land, Water and Planning. 2011. Heritage and culture. In: *French Island & Sandstone Island planning scheme*. Victoria State Government, Melbourne, Australia, pp. 142–146.

Fukushi H, Hirai K. 1992. Proposal of *Chlamydia pecorum* sp. nov. for *Chlamydia* strains derived from ruminants. *Int J Syst Bacteriol* 42:306–308.

Glassick T, Giffard P, Timms P. 1996. Outer Membrane Protein 2 gene sequences indicate that *Chlamydia*

*pecorum* and *Chlamydia pneumoniae* cause infections in koalas. *Syst Appl Microbiol* 19:457–464.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.

Jackson M, Giffard P, Timms P. 1997. Outer Membrane Protein A gene sequencing demonstrates the polyphyletic nature of koala *Chlamydia pecorum* isolates. *Syst Appl Microbiol* 20:187–200.

Jelocnik M, Frentiu FD, Timms P, Polkinghorne A. 2013. Multilocus sequence analysis provides insights into molecular epidemiology of *Chlamydia pecorum* infections in Australian sheep, cattle, and koalas. *J Clin Microbiol* 51:2625–2632.

Jolley KA, Maiden MC. 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11:595.

Kaltenboeck B, Heinen E, Schneider R, Wittenbrink MM, Schmeer N. 2009. *OmpA* and antigenic diversity of bovine *Chlamydophila pecorum* strains. *Vet Microbiol* 135:175–180.

Kollipara A, Polkinghorne A, Wan C, Kanyoka P, Hanger J, Loader J, Callaghan J, Bell A, Ellis W, Fitzgibbon S, et al. 2013. Genetic diversity of *Chlamydia pecorum* strains in wild koala locations across Australia and the implications for a recombinant *C. pecorum* major outer membrane protein based vaccine. *Vet Microbiol* 167:513–522.

Martin RW. 1981. Age-specific fertility in three populations of the koala, *Phascolarctos cinereus* Goldfuss, in Victoria. *Wildl Res* 8:275–283.

Martin RW, Handasyde KA. 1999. *The koala: Natural history, conservation and management*, 2nd Ed. UNSW Press Ltd., Sydney, New South Wales, Australia, 132 pp.

Mohamad KY, Kaltenboeck B, Shamsur Rahman K, Magnino S, Sachse K, Rodolakis A. 2014. Host adaptation of *Chlamydia pecorum* towards low virulence evident in co-evolution of the *ompA*, *incA*, and ORF663 Loci. *PLoS One* 9:e103615.

Obendorf DL. 1981. Pathology of the female reproductive tract in the koala, *Phascolarctos cinereus* (Goldfuss), from Victoria, Australia. *J Wildl Dis* 17:587–592.

Obendorf DL. 1983. Causes of mortality and morbidity of wild koalas, *Phascolarctos cinereus* (Goldfuss), in Victoria, Australia. *J Wildl Dis* 19:123–131.

Pannekoek Y, Morelli G, Kusecek B, Morre S, Ossewaarde J, Langerak A, van der Ende A. 2008. Multi locus sequence typing of Chlamydiales: Clonal groupings within the obligate intracellular bacteria *Chlamydia trachomatis*. *BMC Microbiol* 8:42.

Patterson JLS, Lynch M, Anderson GA, Noormohammadi AH, Legione A, Gilkerson JR, Devlin JM. 2015. The prevalence and clinical significance of *Chlamydia* infection in island and mainland populations of Victorian koalas (*Phascolarctos cinereus*). *J Wildl Dis* 51:309–317.

*Submitted for publication 13 October 2015.*
*Accepted 23 November 2015.*

# 4. Identification of unusual *Chlamydia pecorum* genotypes in Victorian koalas (*Phascolarctos cinereus*) and clinical variables associated with infection

**Published by:** Journal of Medical Microbiology, volume 65, issue 5, 2016.

**Co-authors:** J. L. S. Patterson, P. L. Whiteley, J. Amery-Gale, M. Lynch, L. Haynes, J. R. Gilkerson, A. Polkinghorne, J. M. Devlin and F. M. Sansom

The original publication can be obtained from the below DOI:

dx.doi.org/10.1099/jmm.0.000241

The following material has been included with permission from the journal.

Supplementary files listed in manuscript are in Appendix 3 and Appendix 4.

*Journal of Medical Microbiology* (2016), **65**, 420–428 **DOI** 10.1099/jmm.0.000241

# Identification of unusual *Chlamydia pecorum* genotypes in Victorian koalas (*Phascolarctos cinereus*) and clinical variables associated with infection

Alistair R. Legione,[1] Jade L. S. Patterson,[1,2] Pam L. Whiteley,[1,3] Jemima Amery-Gale,[1,4] Michael Lynch,[2] Leesa Haynes,[5] James R. Gilkerson,[6] Adam Polkinghorne,[7] Joanne M. Devlin[1]† and Fiona M. Sansom[1]†

Correspondence
Alistair R. Legione
alegione@student.unimelb.edu.au

[1]Asia Pacific Centre for Animal Health, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria, Australia

[2]Veterinary Department, Melbourne Zoo, Parkville, Victoria, Australia

[3]Wildlife Health Surveillance Victoria, The University of Melbourne, Werribee, Victoria, Australia

[4]Australian Wildlife Health Centre, Healesville Sanctuary, Healesville, Victoria, Australia

[5]Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Werribee, Victoria, Australia

[6]Centre for Equine Infectious Diseases, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria, Australia

[7]Centre for Animal Health Innovation, University of the Sunshine Coast, Sippy Downs, Queensland, Australia

Received 15 December 2015
Accepted 28 February 2016

*Chlamydia pecorum* infection is a threat to the health of free-ranging koalas (*Phascolarctos cinereus*) in Australia. Utilizing an extensive sample archive we determined the prevalence of *C. pecorum* in koalas within six regions of Victoria, Australia. The *ompA* genotypes of the detected *C. pecorum* were characterized to better understand the epidemiology of this pathogen in Victorian koalas. Despite many studies in northern Australia (i.e. Queensland and New South Wales), prior *Chlamydia* studies in Victorian koalas are limited. We detected *C. pecorum* in 125/820 (15 %) urogenital swabs, but in only one ocular swab. Nucleotide sequencing of the molecular marker *C. pecorum ompA* revealed that the majority (90/114) of *C. pecorum* samples typed were genotype B. This genotype has not been reported in northern koalas. In general, *Chlamydia* infection in Victorian koalas is associated with milder clinical signs compared with infection in koalas in northern populations. Although disease pathogenesis is likely to be multifactorial, the high prevalence of genotype B in Victoria may suggest it is less pathogenic. All but three koalas had *C. pecorum* genotypes unique to southern koala populations (i.e. Victoria and South Australia). These included a novel *C. pecorum ompA* genotype and two genotypes associated with livestock. Regression analysis determined that significant factors for the presence of *C. pecorum* infection were sex and geographical location. The presence of 'wet bottom' in males and the presence of reproductive tract pathology in females were significantly associated with *C. pecorum* infection, suggesting variation in clinical disease manifestations between sexes.

## INTRODUCTION

The koala (*Phascolarctos cinereus*) is an iconic arboreal marsupial, native to Australia. Koalas are considered a vulnerable species in parts of Australia due to rapid population decline. Whilst the contraction of populations is primarily attributed to the impacts of urbanization (Lunney *et al.*, 2007; de Oliveira *et al.*, 2014), disease

†These authors contributed equally to this work.

Abbreviation: qPCR, quantitative PCR.

The GenBank/EMBL/DDBJ accession numbers for the *ompA* sequences of *Chlamydia pecorum* are KU214244–KU214251.

One supplementary table and one supplementary figure are available with the online Supplementary Material.

from pathogens such as *Chlamydia pecorum* may play an important role (Polkinghorne *et al.*, 2013).

*C. pecorum*, an intracellular bacteria of the family *Chlamydiaceae*, predominantly infects the urogenital tract of koalas (Brown *et al.*, 1984), although in northern populations (i.e. Queensland and New South Wales) it is sometimes associated with ocular pathology (Polkinghorne *et al.*, 2013). Urogenital infection with *C. pecorum* can be associated with 'wet bottom' or 'dirty tail', which refers to urine staining or scalding of the rump (Dickens, 1976), due to the presence of cystitis. In females, reproductive tract abnormalities such as paraovarian cysts are also associated with *C. pecorum* infection (Obendorf, 1981) and often cause irreversible infertility (McColl *et al.*, 1984). At a population level, such infertility can result in a rapid decline over a short time span, with introduction of *C. pecorum* in naive koala populations reducing fecundity to zero in as little as 25 years (Martin & Handasyde, 1999).

Our previous study of clinical signs associated with *Chlamydia* infection in Victorian koalas suggested that disease is more mild than that seen in northern populations (Patterson *et al.*, 2015), although further studies of koala populations are needed to definitively confirm this. The reasons behind these apparently milder signs are unclear, and may be due to less virulent *C. pecorum* and/or variation in the prevalence of different *Chlamydia* species, but could also be due to other factors such as lower rates of koala retrovirus in Victoria (Simmons *et al.*, 2012). Variation in host and environmental factors between southern and northern populations would also be expected to result in variation of disease expression (Patterson *et al.*, 2015). Previous studies assessing the diversity of *C. pecorum* in Australian koala populations have focused on populations in Queensland and New South Wales (Kollipara *et al.*, 2013). Comprehensive *C. pecorum* genotyping studies in Victorian koalas are currently lacking, but would be useful in understanding the observed differences in disease syndromes between koala populations.

In *C. pecorum*, *ompA* encodes a major outer membrane protein associated with virulence (Fitch *et al.*, 1993). The nucleotide sequence of *ompA*, which has four variable domains, has been used frequently to genotype *C. pecorum* samples collected from koalas, leading to the detection of 11 koala-associated genotypes, named A–K (Jackson *et al.*, 1997; Kollipara *et al.*, 2013). *C. pecorum* in livestock species such as cattle, pigs, sheep and goats have also been genotyped using *ompA* (Jackson *et al.*, 1997; Mohamad *et al.*, 2014). Although a number of different methods are now available to genotype *C. pecorum* samples (Marsh *et al.*, 2011; Jelocnik *et al.*, 2013), the existing catalogue of *ompA* sequences from northern koala populations makes it a useful tool to undertake comparisons with Victorian koala populations. Interestingly, differences in the presence of a chlamydial plasmid, a known virulence factor in some chlamydial species (O'Connell *et al.*, 2007), was also found to be a distinguishing feature of *C. pecorum* strains from northern populations compared with those from southern populations in South Australia (Jelocnik *et al.*, 2015). Whilst some Victorian koala *C. pecorum* samples were found to be PCR-positive for this plasmid (p*Cpec*), only a relatively small number of animals were sampled.

In this study, *C. pecorum ompA* molecular typing and p*Cpec* screening were performed on a large number of koala samples collected from different free-ranging populations across Victoria. Furthermore, using our newly collected data, reported here, combined with our previously reported data (Patterson *et al.*, 2015), we analysed clinical information and *C. pecorum* genome copy number to detect any significant associations between *C. pecorum* infection load and clinical disease, and to identify any significant associations between *C. pecorum* infection and a range of host-related variables.

## METHODS

**Sample collection.** Sample collection was approved by The University of Melbourne Animal Ethics Committee (approval numbers 1011687.1 and 1312813.2) and Parks Victoria (Research Permits 10004605, 10006948 and 10005388). In total, urogenital swabs from 820 koalas and ocular swabs from 459 koalas were used in this study. These included 430 urogenital samples collected as a component of previously described research (Patterson *et al.*, 2015; Legione *et al.*, 2016). Other samples were collected between 2010 and 2015, inclusive, during a variety of research field trips, management programmes and post-mortem examinations. Of the ocular swabs, 456 came from koalas from which urogenital swabs were also collected. Samples were collected using either an aluminium or plastic-shafted rayon swab (Copan Italia). Clinical examinations of live captured animals were performed by veterinarians, including assessment of body condition score (Patterson *et al.*, 2015), wet bottom score (Griffith, 2010) and, for a subsample of koalas, the presence or absence of urogenital tract pathology by ultrasound (Patterson *et al.*, 2015). All koalas also had a suite of other parameters recorded, including tooth wear class (Martin, 1981) and the presence or absence of young. Gross pathology of the urogenital tract of any koalas that required euthanasia for health and welfare reasons was also recorded. All but six koalas used in the study were from free-ranging populations, with five koalas held in care for $>1$ month prior to euthanasia and one koala raised in captivity.

**DNA extraction from swab samples.** Swabs were added to 1.5 ml tubes containing 800 µl either PBS (137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$, 2 mM $KH_2PO_4$, pH 7.4), Tris/EDTA buffer (10 mM Tris, 1 mM EDTA, pH 8.0) or RLT buffer (Qiagen) and mixed briefly using an Xtron Vortex Mixer (Bartelt Instruments) prior to processing. DNA extraction was then carried out using an X-tractor robot (Qiagen), utilizing a Qiaxtractor VX extraction kit as per the manufacturer's instructions. Each extraction contained both positive and negative extraction controls. The positive control was a diluted broth of *E. coli* containing a pGEM-T (Promega) plasmid, which in turn contained the *C. pecorum* 16S rRNA target, whilst the negative control was either sterile water or PBS.

**Quantitative PCR for *Chlamydia* prevalence.** Extracted DNA was tested for the presence of *Chlamydia* spp. using the *Chlamydia* 16SG nucleotide set qPCR first described by Robertson *et al.* (2009). The resulting melt curves from this qPCR, coupled with appropriate controls, can be used to distinguish detected *Chlamydia* species (Fig. S1, available in the online Supplementary Material). The koala

housekeeping gene $\beta$-actin was utilized to standardize genome copy numbers detected in each sample, as described previously (Shojima *et al.*, 2013). Each sample was standardized based on the number of 16S rRNA copies per $\beta$-actin copies in the extracted liquid sample. A standard curve was employed for each qPCR consisting of 10-fold dilutions, in triplicate, of purified plasmid containing either the 16S rRNA or $\beta$-actin gene from $10^7$ to $10^1$ copies per reaction. Copy numbers were calculated using a Qubit 3.0 fluorometer (Invitrogen).

**Genotyping of *C. pecorum* using *ompA*.** *C. pecorum*-positive DNA was used as template for conventional PCR to amplify *ompA*, as described previously (Kollipara *et al.*, 2013). PCR products of the predicted size ($\sim$1170 bp) were purified using a QIAquick PCR purification kit (Qiagen) and sequenced using BigDye Terminator version 3.1 (Life Technologies) using *ompA* PCR primers (Kollipara *et al.*, 2013), and internal primers 5′-AGAGCTACTTTTGATGCAGA-3′ and 5′-TTTGTGAACCACTCCGCATC-3′. Geneious 7 software (Biomatters) was used for sequence analysis, and CLUSTAL W (Thompson *et al.*, 1994) was used to align results with published *ompA* sequences available in GenBank and in the literature (Jackson *et al.*, 1997). Genotypes were assigned based on similarity to published genotypes. A difference of $>$1 % nucleotides across the $\sim$1170 bp gene was used as the threshold for classifying a new genotype, as described previously (Kollipara *et al.*, 2013).

**p*Cpec* screening in *C. pecorum*-positive Victorian koala samples.** The prevalence of p*Cpec* across the *C. pecorum*-positive samples identified in this study was detected by use of p*Cpec*-specific conventional PCR, as described previously (Jelocnik *et al.*, 2015).

**Assessment of factors associated with chlamydial infection.** Statistical analysis was conducted using Minitab 17 software (Minitab). The analysed parameters included sex, age (based on tooth wear), body condition score, reproductive and urinary tract gross pathology, fecundity (based on the presence of back or pouch young), wet bottom presence/absence, and time of year and location collected. Not all parameters were available for all animals. Koalas were classified as young (tooth wear less than class III), mature (tooth wear class III–V) or old (tooth wear greater than class V) (Patterson *et al.*, 2015). Specific location data were pooled into broader regional locations (Fig. 1).

Binomial logistical regression univariable analysis was performed on each variable in relation to the presence or absence of *C. pecorum*. Variables with $P \leq 0.25$ were included in the multivariable analysis. A backward elimination method was used to determine the final multivariable logistic regression model to identify which variables were either significant risk factors for, or useful predictors of, *C. pecorum* infection.

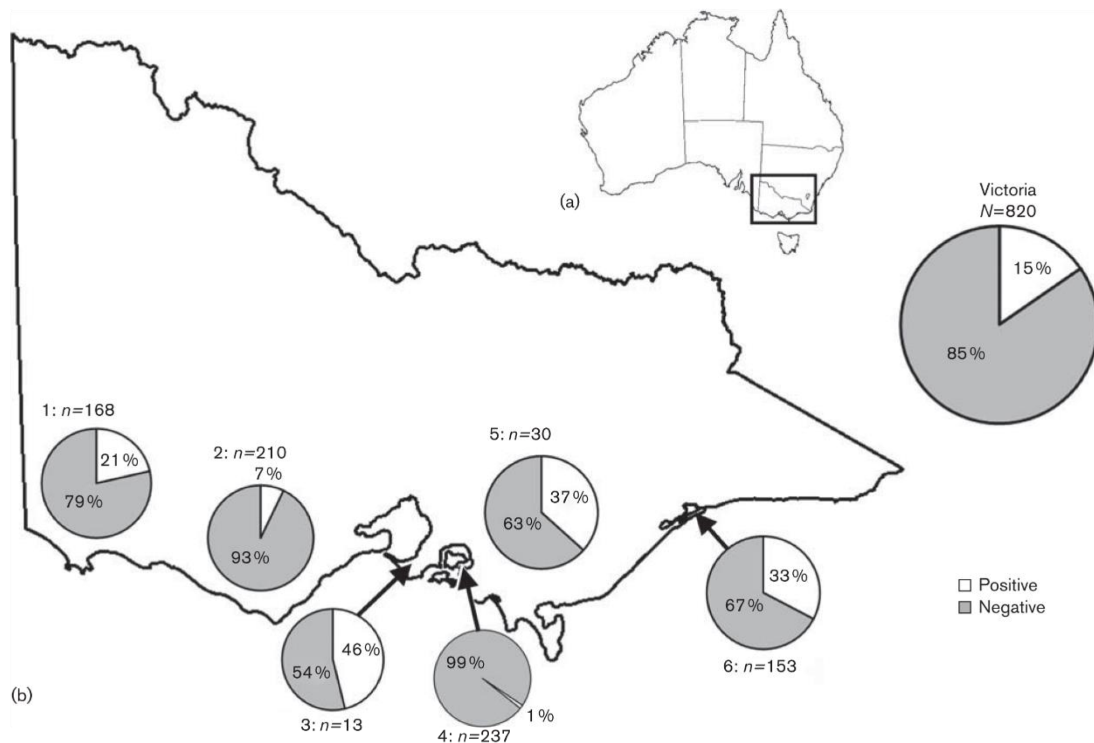



**Fig. 1.** Prevalence of *C. pecorum* infection in Victorian koala populations. (a) Map of Australia highlighting the state of Victoria (boxed). (b) Regional prevalence of *C. pecorum* detected in Victorian koala urogenital swabs. Regions: 1, Far Western Victoria; 2, South West Coast; 3, Mornington Peninsula; 4, French Island; 5, Greater Gippsland; 6, Raymond Island. Map not to scale.

Standardized copy numbers were exponentially transformed to normally distribute the data. Data >2SD from the mean were considered outliers and removed from the analysis. *C. pecorum* genome copies in each sample in relation to clinical signs (body condition, wet bottom and urinary and reproductive tract pathology), as well as organism-specific variables, such as *ompA* genotype and p*Cpec* presence, were compared using Student's *t*-test. Linear regression was used to determine which variable had the strongest effect on the amount of *C. pecorum* DNA detected via qPCR.

## RESULTS

### Prevalence of *Chlamydia* infection in Victorian koalas varied between populations

We detected *C. pecorum* in 15.2 % (125/820) of the urogenital tract samples tested (Fig. 1). Only 0.4 % (2/459) of ocular swabs were positive for *Chlamydia*, with one case of *Chlamydia pneumoniae* detected in the South West Coast region and one *C. pecorum*-positive ocular swab from the Greater Gippsland region. Substantial variation in prevalence was seen across the different geographical regions. The lowest prevalence of *C. pecorum* was 0.8 % (2/237) in the French Island koala population and 7.1 % (15/210) in the South West Coast region, which incorporates the Great Otway National Park and its surrounding coastline (Fig. 1). The highest prevalence was found in the Greater Gippsland (36.7 %, 11/30) and Mornington Peninsula (46.1 %, 6/13) regions (Fig. 1).

### Sequencing of *ompA* revealed that genotype B dominates in Victorian koalas, but also identified a novel genotype and two genotypes not previously associated with koalas

The *C. pecorum ompA* sequence was determined for 114 *C. pecorum*-positive samples. A summary of the different genotypes detected, and their geographical distribution, is shown in Table 1. Genotype B predominated in Victorian koalas, occurring in 90/114 *ompA* genotyped urogenital cases (78.9 %). Nine koalas were infected with genotype C and only three koalas, all from the Greater Gippsland region, were infected with genotype F. A novel genotype, designated 'M', was detected in one urogenital sample from the Greater Gippsland region. This genotype had 98.6 % identity to a *C. pecorum ompA* sequence from a livestock isolate (strain M14; GenBank accession number EU684920.1) (Mohamad *et al.*, 2008). The single detected case of ocular *C. pecorum* infection, also occurring in the Greater Gippsland region, was also genotype M. Nine koalas that tested positive for *C. pecorum* from the Great Otway National Park region on the South West Coast of Victoria were found to have a genotype not previously found in koalas. This genotype, which we designated 'L', has 99.9 % nucleotide identity to a livestock-associated *C. pecorum* (strain DC49; GenBank accession number GQ228195) (Mohamad *et al.*, 2014).

### *C. pecorum* plasmid p*Cpec* is present in 90 % of Victorian samples

We screened the *C. pecorum*-positive samples for the *C. pecorum* plasmid p*Cpec*. This revealed a p*Cpec* prevalence of 90 % (113/125) (Table 1). This high prevalence precluded statistical evaluation of the contribution of the plasmid to the pathogenicity of *C. pecorum*, although it is interesting that in all cases where pCpec was absent, wet bottom was also absent. There was no significant relationship between p*Cpec* presence and the number of bacterial genome copies detected from swabs ($P$=0.326). Univariable binomial logistic regression was conducted to test the relationship between p*Cpec* presence and gender or

**Table 1.** *C. pecorum ompA* genotype in urogenital swabs from Victorian koalas and prevalence of p*Cpec*

| Region | *C. pecorum*-positive | *ompA* genotype* | | | | | | | p*Cpec*-positive |
|---|---|---|---|---|---|---|---|---|---|
| | | B | C | F | L | M | N† | Unknown‡ | |
| Far Western Victoria | 36 | 35 | 0 | 0 | 0 | 0 | 0 | 1 | 33 |
| South West Coast | 15 | 2 | 0 | 0 | 9 | 0 | 0 | 4 | 11 |
| Mornington Peninsula | 6 | 1 | 4 | 0 | 0 | 0 | 0 | 1 | 6 |
| French Island | 2 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| Greater Gippsland | 11 | 1 | 3 | 3 | 0 | 1§ | 0 | 3 | 10 |
| Raymond Island | 50 | 49 | 0 | 0 | 0 | 0 | 0 | 1 | 47 |
| Other | 5 | 2 | 2 | 0 | 0 | 0 | 0 | 1 | 4 |
| Total | 125 | 90 | 9 | 3 | 9 | 1 | 2 | 11 | 113 |

*Examples of *ompA* genotypes submitted to GenBank under accession numbers KU214244 (genotype N), KU214245 (genotype C), KU214246 (genotype F), KU21427 (genotype M), KU214248 (genotype B), KU214249 (genotype B), KU214250 (genotype L) and KU214251 (genotype B).
†This genotype was previously described in Legione *et al.* (2016); however, it was not named as genotype N.
‡Unknown genotypes were *C. pecorum*-positives from which *ompA* could not be amplified. All samples were confirmed to be *C. pecorum* via sequencing of the 16S rRNA region and/or presence of *C. pecorum* plasmid p*Cpec*.
§The only ocular swab positive for *C. pecorum* was also genotype M from the Greater Gippsland region.

body condition score, but neither were significant factors ($P=0.104$ and $P=0.785$, respectively).

### Male koalas had a higher likelihood of urogenital infection with *C. pecorum* than female koalas

Binary logistic regression analysis, with the presence or absence of *C. pecorum* infection as an outcome, was used to assess animal signalment and associated sampling information. Univariable analysis determined that sex, age of animal, season and region captured, and the presence of back or pouch young (in females only) were significant predictive factors for *C. pecorum* infection (Table 2). In the multivariable model ($n=614$), after stepwise backward elimination, only capture region was identified as a significant factor ($P<0.001$). Removing capture region from the multivariable model to assess the variables at a population-wide level, and repeating the stepwise elimination, identified sex as a significant factor ($P<0.001$), with male koalas 2.7 times more likely (95 % confidence interval 1.60–4.45) to be positive for *C. pecorum.*

### Urogenital *C. pecorum* infection was significantly associated with wet bottom in male koalas and with reproductive tract pathology in female koalas

Assessment of clinical signs and pathological findings, through binary logistic regression analysis using the presence or absence of *C. pecorum* infection as an outcome, was conducted with male and female animals pooled, and with male and female animals separated. The results from this univariable analysis are summarized in Tables 3 and S1,

**Table 2.** Univariable analysis assessing selected epidemiological variables relating to animal signalment and sample collection as predictors for the presence of *Chlamydia* DNA

| Variable | *Chlamydia*-positive | Prevalence (%) | Odds ratio* | 95 % Confidence interval | Coefficient *P*-value | Likelihood ratio *P*-value† |
|---|---|---|---|---|---|---|
| **Sex** | | | | | | *0.020* |
| Female | 84/604 | 13.9 | 1.00 | – | – | |
| Male | 39/184 | 21.2 | 1.67 | 1.09–2.54 | 0.018 | |
| Not recorded | 2/32 | | | | | |
| **Age** | | | | | | *0.007* |
| Young | 13/124 | 10.5 | 0.75 | 0.40–1.41 | 0.400 | |
| Mature adult | 70/519 | 13.5 | 1.00 | – | – | |
| Old adult | 16/55 | 29.1 | 2.63 | 1.40–4.96 | 0.002 | |
| Not recorded | 27/122 | | | | | |
| **Season** | | | | | | *0.107* |
| Summer | 1/7 | 14.3 | 1.18 | 0.14–9.94 | 0.880 | |
| Autumn | 30/156 | 19.2 | 1.68 | 1.05–2.70 | 0.030 | |
| Winter | 0/24 | 0 | – | – | – | |
| Spring | 69/557 | 12.4 | 1.00 | – | – | |
| Not recorded | 25/76 | | | | | |
| **Back/pouch young or lactation (females)** | | | | | | *<0.001* |
| No | 36/247 | 14.6 | 1.00 | – | – | |
| Yes | 11/251 | 4.4 | 0.27 | 0.13–0.54 | <0.001 | |
| Not recorded | 37/106 | | | | | |
| **Region‡** | | | | | | *<0.001* |
| Far Western Victoria | 36/168 | 21.4 | 3.55 | 1.87–6.73 | <0.001 | |
| South West Coast | 15/210 | 7.1 | 1.00 | – | – | |
| Mornington Peninsula | 6/13 | 46.2 | – | – | – | |
| French Island | 2/237 | 0.8 | – | – | – | |
| Greater Gippsland | 11/30 | 36.7 | 7.53 | 3.03–18.69 | <0.001 | |
| Raymond Island | 50/153 | 32.7 | 6.31 | 3.38–11.78 | <0.001 | |
| Other/not recorded | 5/9 | | | | | |

*Reference levels are indicated by odds ratio of 1.0.

†Results in italics (log-likelihood $P \leq 0.25$) represent variables included in the initial multivariable model, with the exception of the presence or absence of young/lactation, as this correlated with sex and was thus excluded. In the final model ($n=614$, after stepwise backward elimination), only 'region' was identified as a significant factor ($P<0.001$). Removing 'region' from the multivariable model to assess the variables at a population-wide level and repeating the stepwise elimination identified sex as a significant factor ($P<0.001$).

‡Only regions with >10 positive cases were included in the analysis.

**Table 3.** Univariable analysis assessing selected epidemiological variables relating to animal health and disease as predictors for the presence of *Chlamydia* DNA in koalas of each sex

| Variable | *Chlamydia*-positive | Prevalence (%) | Odds ratio* | 95 % Confidence interval | Coefficient *P*-value | Likelihood ratio *P*-value† |
|---|---|---|---|---|---|---|
| **Female koalas** | | | | | | |
| Body condition score | | | | | | 0.734 |
| ≤2 | 8/80 | 10 | 0.87 | 0.40–1.92 | 0.737 | |
| ≥3 | 53/470 | 11.3 | 1.00 | – | – | |
| Unknown | 23/54 | | | | | |
| Wet bottom | | | | | | 0.277 |
| Absent | 57/444 | 12.8 | 1.00 | – | – | |
| Present | 26/159 | 16.4 | 1.33 | 0.80–2.20 | 0.271 | |
| Not recorded | 1/1 | | | | | |
| Urinary tract pathology | | | | | | 0.511 |
| Absent | 27/124 | 21.8 | 1.00 | – | – | |
| Present | 9/33 | 27.3 | 1.35 | 0.56–3.24 | 0.505 | |
| Not recorded | 48/447 | | | | | |
| Reproductive tract pathology | | | | | | *<0.001* |
| Absent | 16/110 | 14.5 | 1.00 | – | – | |
| Present | 19/45 | 42.2 | 4.29 | 1.94–9.50 | <0.001 | |
| Not recorded | 49/449 | | | | | |
| **Male koalas** | | | | | | |
| Body condition score | | | | | | *0.008* |
| ≤2 | 2/32 | 6.3 | 0.19 | 0.04–0.85 | 0.030 | |
| ≥3 | 29/112 | 25.9 | 1.00 | – | | |
| Unknown | 8/40 | | | | | |
| Wet bottom | | | | | | *<0.001* |
| Absent | 24/155 | 15.5 | 1.00 | – | – | |
| Present | 15/28 | 53.6 | 6.30 | 2.66–14.90 | <0.001 | |
| Not recorded | 0/1 | | | | | |
| Urinary tract pathology | | | | | | *0.644* |
| Absent | 13/44 | 29.5 | 1.00 | – | – | |
| Present | 3/13 | 23.1 | 0.72 | 0.17–3.03 | 0.649 | |
| Not recorded | 23/127 | | | | | |

*Reference levels are indicated by odds ratio of 1.0.

†Results in italics (log-likelihood $P \leq 0.25$) represent variables included in the initial multivariable model. For female koalas, only one variable was significant at the univariable level so multivariable analysis was not conducted. For males in the final model ($n=144$), after stepwise backward elimination the only significant factor ($P<0.001$) was the presence or absence of wet bottom.

respectively. After separating data by sex, a multivariable analysis found that for male koalas, wet bottom was the only significant indicator of *C. pecorum* infection ($P<0.001$). For females, reproductive tract pathology was the only significant indicator ($P<0.001$).

### *C. pecorum* genome copy numbers were significantly higher in swabs from koalas infected with genotype F

The geometric mean of bacterial genome copies for each *C. pecorum* genotype was determined (Table 4). Swabs collected from koalas infected with genotype F had a significantly higher mean genome copy number than swabs collected from koalas infected with genotypes B and L. Linear regression analysis with stepwise backward elimination was conducted with genome copy number as an outcome, and genotype, p*Cpec*, wet bottom presence and gender as categorical predictors. This showed that genotype was the only significant predictive factor of genome load ($P=0.034$).

## DISCUSSION

Our *Chlamydia* prevalence data incorporated three koala populations not previously investigated for these pathogens (Greater Gippsland and Mornington Peninsula in eastern Victoria, and the South West Coast in western Victoria). In addition we have expanded on previously published results for three populations [Mt Eccles (in the Far Western region), French Island and Raymond Island (Patterson *et al.*, 2015)]. From Patterson *et al.* (2015) we included

**Table 4.** *C. pecorum* load for different *ompA* genotypes

Genome copies were quantified from urogenital swabs via 16S rRNA (16SG nucleotide set) qPCR and standardized using koala $\beta$-actin qPCR.

| Genotype | Samples* | Load ($\log_{10}$16S rRNA/$10^6$ $\beta$-actin copies) (mean ± SD)† |
|---|---|---|
| B | 87 | $2.52 \pm 0.96^{a}$ |
| C | 7 | $2.84 \pm 1.07^{a,b}$ |
| F | 3 | $3.73 \pm 0.13^{b}$ |
| L | 9 | $2.78 \pm 0.74^{a}$ |
| M | 1 | 1.95‡ |
| N | 2 | 1.01 ± 0.20‡ |

*Values that fell >2SD from the mean were removed from the data set before analysis. These included three samples from genotype B and two from genotype C.
†Values with the same superscript symbol were not significantly different using the two-sample Student's *t*-test ($P \geq 0.05$).
‡Analysis was not performed due to insufficient samples ($n<3$).

288 koalas from which we previously reported prevalence, 286 from which the presence or absence of wet bottom was reported and 117 koalas for which urogenital tract pathology was reported. These data were included to improve the power of our study and increase the likelihood of accurate correlations being described. For similar reasons, here we also include the clinical data from two *Chlamydia*-positive koalas from French Island, whose *ompA* genotypes we reported recently (Legione *et al*., 2016).

To the best of our knowledge, this is the largest molecular study of *Chlamydia* in free-ranging koala populations reported to date. Prevalence of *C. pecorum* varied substantially across the state, ranging from 1 to 46 %, with the highest prevalence of *C. pecorum* occurring in the Mornington Peninsula and Greater Gippsland regions. All but two samples from these high-prevalence regions were obtained post-mortem from free-ranging koalas after euthanasia was required because of injury or disease (the other two samples, both *Chlamydia*-negative, were obtained from koalas that had been kept in captivity for up to 1 year). If euthanasia was performed due to disease related to *Chlamydia* infection, or resulted in immune suppression that enhanced bacterial shedding, the likelihood of detecting *Chlamydia* would increase, which may in part explain the higher prevalence in these populations. Easier access to anatomical sampling sites during post-mortem examination may also have increased detection rates. Our findings reflect previous investigations of *C. pecorum ompA* genotypes in northern koala populations (Kollipara *et al*., 2013), which also found variable *C. pecorum* prevalence (28–61 % in Queensland koalas and 20–63 % in New South Wales koalas), and noted that *C. pecorum* was more commonly detected in samples taken from sick or injured koalas. In contrast, the prevalence of *C. pneumoniae* in northern populations ranges from 4 to 23 % (Polkinghorne *et al*., 2013), but we detected *C. pneumoniae* in only one ocular swab. Interestingly, ocular pathology is commonly associated with *C. pecorum* infection in northern koalas (Polkinghorne *et al*., 2013), but no koalas with ocular pathology in our study ($n=44$) were infected with *Chlamydia*, suggesting other causes were responsible.

The existence of a large catalogue of *ompA* genotyped samples from northern koala populations makes *ompA* genotyping a useful tool for undertaking comparisons between northern and southern populations. Recent research also suggests *ompA* is a valuable epidemiological marker for phylogenetic analysis (Marsh *et al*., 2011), although it may overestimate whole-genome evolution due to its exposure to strong selection pressure and recombination (Harris *et al*., 2012). Here, we used *ompA* genotyping to reveal that, in comparison with the considerable heterogeneity seen in northern koala populations (Kollipara *et al*., 2013), *C. pecorum ompA* genotype B dominates in samples collected from Victorian koalas. Genotype B has previously been found only in other southern koalas (Jackson *et al*., 1997; Higgins *et al*., 2012; Kollipara *et al*., 2013). The lower diversity of *C. pecorum* genotypes in Victorian populations could be reflective of the translocation history of koalas in Victoria. Following a near-extinction event in Victorian koalas around 1900 (Troughton, 1941), koalas from French Island and remnant Gippsland koala populations were used to repopulate Victoria (Martin, 1989). It is possible that koalas infected with *C. pecorum* were translocated from the Gippsland region and from there the pathogen was moved into the newly established Victorian populations (Martin & Handasyde, 1999). The observed homogeneity provides support for this hypothesis, although we cannot rule out that these infections arose due to separate transmission events.

Recorded *C. pecorum*-associated clinical signs are more severe in northern populations than those observed to date in infected Victorian koalas and we isolated *C. pecorum* from only one ocular swab, which was not associated with clinical disease (Griffith, 2010; Polkinghorne *et al*., 2013; Patterson *et al*., 2015). Although *C. pecorum* disease in koalas is no doubt multifactorial, the dominant presence of genotype B across Victoria, and its absence from northern populations, may suggest this *C. pecorum* genotype is less pathogenic for koalas, particularly relating to the lack of ocular infection and/or pathology. However, even amongst genotype B-infected koalas we saw both asymptomatic and diseased animals, with 38 % (34/89) displaying wet bottom and 38 % (13/34) with urogenital pathology such as paraovarian cysts.

Genotype F, found in all northern populations tested previously (Kollipara *et al*., 2013), was only detected in Greater Gippsland, one of the remnant Victorian populations. This may suggest it has been present in this population since before the near-extinction event. Genotype C, found

previously only in koalas of Victorian origin (Jackson *et al.*, 1997; Higgins *et al.*, 2012), and the novel genotype M were also found in the Greater Gippsland population. This region appears to have greater genetic diversity than the populations in the rest of Victoria, and additional future sampling from the Gippsland region would provide greater insight into *C. pecorum* genetic variation in this population.

The South West Coast koala population suffers from severe overpopulation. *C. pecorum* prevalence was significantly lower here when compared with other mainland populations, which may explain why there is no apparent loss of fertility. The lower prevalence may also suggest recent acquisition of *Chlamydia* by this population. The presence of genotype L in this population, not previously detected in koalas and not found in other Victorian populations in this study, suggests this *C. pecorum* strain was not moved with the koalas when this population was established by translocation of koalas from French Island in the 1980s (Martin, 1989). The origin of genotypes L, M and N in koalas is unknown, although their close similarity to *C. pecorum ompA* sequences from livestock raises interesting questions over the potential for cross-host transmission from livestock, for which molecular evidence continues to grow (Jelocnik *et al.*, 2013; Bachmann *et al.*, 2015; Legione *et al.*, 2016). Future sampling of livestock that occupy ranges overlapping those of koala populations would be invaluable in investigating this hypothesis further.

In addition to analysing *ompA* genotype, we examined the presence of p*Cpec* in our samples. p*Cpec* is more commonly present in *C. pecorum* from koalas than in *C. pecorum* present in other animal species, and its prevalence in koala *C. pecorum* strains from South Australia appears to be much lower than those from northern populations (Jelocnik *et al.*, 2015). In contrast to the results observed in these populations, 90 % of Victorian koala *C. pecorum* samples in this study were positive for p*Cpec*. This is the highest prevalence of the plasmid seen to date, and higher than previous studies in Victorian (79 %) and northern (76 %) populations (Jelocnik *et al.*, 2015). The high prevalence of p*Cpec* in our sample set made it difficult to identify significant effects on virulence, but analysis of genome copy numbers showed no significant difference between cases with or without p*Cpec*. The function of p*Cpec* in koala *C. pecorum* pathogenesis therefore remains unknown.

Previous studies in koalas and other species show that differences in clinical signs are not explained by *ompA* genotype alone (Higgins *et al.*, 2012; Mohamad *et al.*, 2014). Despite the predominance of genotype B, we observed variation in the presence of clinical signs in animals carrying this genotype. Finer detailed evolutionary analysis such as multilocus sequence typing (Jelocnik *et al.*, 2013), which has been used to investigate the pathogenesis of *C. pecorum* infections in sheep (Jelocnik *et al.*, 2014), or newly developed full-genome sequencing techniques (Bachmann *et al.*, 2015) could be employed to further examine the relationship between *C. pecorum* strain type and the observed clinical signs. An analysis of host genetic factors, such as major histocompatibility complex class II variation (Lau *et al.*, 2014), would also be useful for understanding the host–pathogen interaction. The publication of the koala genome will accelerate efforts in this area (Johnson *et al.*, 2014).

*Chlamydia* in koalas has long been associated with clinical disease, primarily ocular and urogenital pathology (Jackson *et al.*, 1999). Our findings partially support this, but also suggest that these clinical signs are not pathognomonic of *Chlamydia* infection. This is an important factor for koala conservation, with recent research advocating the culling of infertile koalas and treatment of *Chlamydia*-infected koalas as a means to reverse population declines (Wilson *et al.*, 2015). In female koalas we found that reproductive tract pathology is a significant factor for predicting the presence of *C. pecorum*. However, there were still a large number of koalas with evidence of reproductive tract pathology where qPCR did not detect *C. pecorum* (26/45), suggesting other aetiological agent(s) may be present that should also be considered in conservation programs. Signs of disease due to other pathogens, such as koala herpesvirus (Stalder *et al.*, 2015) or koala retrovirus (Denner, 2014), could also overlap with *C. pecorum* infection. Comprehensive investigations into these and other pathogens may also further our understanding of the true significance of *C. pecorum* in Victorian koalas.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the assistance with collection of samples and clinical information by staff and students of the University of Melbourne, Wildlife Health Surveillance Victoria, Zoos Victoria, Victorian Department of Environment, Land, Water and Planning, and Parks Victoria. We also thank Mark Stevenson for advice and assistance with statistics. A. L. is supported by an Australian Postgraduate Award. Funding was provided for this research by the Holsworth Wildlife Research Endowment, the Vizard Foundation and the Albert George and Nancy Caroline Youngman Trust.

## REFERENCES

**Bachmann, N. L., Sullivan, M. J., Jelocnik, M., Myers, G. S. A., Timms, P. & Polkinghorne, A. (2015).** Culture-independent genome sequencing of clinical samples reveals an unexpected heterogeneity of infections by *Chlamydia pecorum*. *J Clin Microbiol* **53**, 1573–1581.

**Brown, A. S., Carrick, F. N., Gordon, G. & Reynolds, K. (1984).** The diagnosis and epidemiology of an infertility disease in the female koala *Phascolarctos cinereus* (*Marsupialia*). *Vet Radiol* **25**, 242–248.

**de Oliveira, S. M., Murray, P. J., de Villiers, D. L. & Baxter, G. S. (2014).** Ecology and movement of urban koalas adjacent to linear infrastructure in coastal south-east Queensland. *Aust Mammal* **36**, 45–54.

**Denner, J. (2014).** Immunization with envelope proteins of the KoRV as a basis for a preventive vaccine. *Tech Rep Aust Mus* **24**, 71–77.

**Dickens, R. K. (1976).** The koala in health and disease. In *Refresher Course for Veterinarians Proceedings 29*, pp. 105–117, Sydney: Taronga Zoo/The University of Sydney.

**Fitch, W. M., Peterson, E. M. & de la Maza, L. M. (1993).** Phylogenetic analysis of the outer-membrane-protein genes of *Chlamydiae*, and its implication for vaccine development. *Mol Biol Evol* **10**, 892–913.

**Griffith, J. E. (2010).** *Studies into the diagnosis, treatment and management of chlamydiosis in koalas.* PhD Thesis, The University of Sydney, Sydney, New South Wales, Australia.

**Harris, S. R., Clarke, I. N., Seth-Smith, H. M. B., Solomon, A. W., Cutcliffe, L. T., Marsh, P., Skilton, R. J., Holland, M. J., Mabey, D. & other authors (2012).** Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. *Nat Genet* **44**, 413–419, S1.

**Higgins, D. P., Beninati, T., Meek, M., Irish, J. & Griffith, J. E. (2012).** Within-population diversity of koala *Chlamydophila pecorum* at *ompA* VD1-VD3 and the *ORF663* hypothetical gene. *Vet Microbiol* **156**, 353–358.

**Jackson, M., Giffard, P. & Timms, P. (1997).** Outer membrane protein A gene sequencing demonstrates the polyphyletic nature of koala *Chlamydia pecorum* isolates. *Syst Appl Microbiol* **20**, 187–200.

**Jackson, M., White, N., Giffard, P. & Timms, P. (1999).** Epizootiology of *Chlamydia* infections in two free-range koala populations. *Vet Microbiol* **65**, 255–264.

**Jelocnik, M., Frentiu, F. D., Timms, P. & Polkinghorne, A. (2013).** Multilocus sequence analysis provides insights into molecular epidemiology of *Chlamydia pecorum* infections in Australian sheep, cattle, and koalas. *J Clin Microbiol* **51**, 2625–2632.

**Jelocnik, M., Walker, E., Pannekoek, Y., Ellem, J., Timms, P. & Polkinghorne, A. (2014).** Evaluation of the relationship between *Chlamydia pecorum* sequence types and disease using a species-specific multi-locus sequence typing scheme (MLST). *Vet Microbiol* **174**, 214–222.

**Jelocnik, M., Bachmann, N. L., Kaltenboeck, B., Waugh, C., Woolford, L., Speight, K. N., Gillett, A., Higgins, D. P., Flanagan, C. & other authors (2015).** Genetic diversity in the plasticity zone and the presence of the chlamydial plasmid differentiates *Chlamydia pecorum* strains from pigs, sheep, cattle, and koalas. *BMC Genomics* **16**, 893.

**Johnson, R. N., Hobbs, M., Eldridge, M. D. B., King, A. G., Colgan, D. J., Wilkins, M. R., Chen, Z., Prentis, P. J., Pavasovic, A. & other authors (2014).** The Koala Genome Consortium. *Tech Rep Aust Mus* **24**, 91–92.

**Kollipara, A., Polkinghorne, A., Wan, C., Kanyoka, P., Hanger, J., Loader, J., Callaghan, J., Bell, A., Ellis, W. & other authors (2013).** Genetic diversity of *Chlamydia pecorum* strains in wild koala locations across Australia and the implications for a recombinant *C. pecorum* major outer membrane protein based vaccine. *Vet Microbiol* **167**, 513–522.

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

**Legione, A. R., Amery-Gale, J., Lynch, M., Haynes, L., Gilkerson, J. R., Sansom, F. M. & Devlin, J. M. (2016).** *Chlamydia pecorum* infection in free-ranging koalas (*Phascolarctos cinereus*) on French Island, Victoria, Australia. *J Wildl Dis* **52** doi:10.7589/2015-10-276 [Epub ahead of print].

**Lunney, D., Gresser, S., O'Neill, L. E., Matthews, A. & Rhodes, J. (2007).** The Impact of fire and dogs on koalas at Port Stephens, New South Wales, using population viability analysis. *Pac Conserv Biol* **13**, 189–201.

**Marsh, J., Kollipara, A., Timms, P. & Polkinghorne, A. (2011).** Novel molecular markers of *Chlamydia pecorum* genetic diversity in the koala (*Phascolarctos cinereus*). *BMC Microbiol* **11**, 77.

**Martin, R. W. (1981).** Age-specific fertility in three populations of the koala, *Phascolarctos cinereus* Goldfuss, in Victoria. *Wildl Res* **8**, 275–283.

**Martin, R. W. (1989).** *Draft Management Plan for the Conservation of the Koala* (Phascolarctos cinereus) *in Victoria.* Melbourne: Arthur Rylah Institute for Environmental Research.

**Martin, R. W. & Handasyde, K. A. (1999).** *The Koala: Natural History, Conservation and Management*, 2nd edn. Sydney: UNSW Press.

**McColl, K. A., Martin, R. W., Gleeson, L. J., Handasyde, K. A. & Lee, A. K. (1984).** *Chlamydia* infection and infertility in the female koala (*Phascolarctos cinereus*). *Vet Rec* **115**, 655.

**Mohamad, K. Y., Kaltenboeck, B., Rahman, Kh. S., Magnino, S., Sachse, K. & Rodolakis, A. (2014).** Host adaptation of *Chlamydia pecorum* towards low virulence evident in co-evolution of the *ompA, incA*, and ORF663 loci. *PLoS One* **9**, e103615.

**Obendorf, D. L. (1981).** Pathology of the female reproductive tract in the koala, *Phascolarctos cinereus* (Goldfuss), from Victoria, Australia. *J Wildl Dis* **17**, 587–592.

**O'Connell, C. M., Ingalls, R. R., Andrews, C. W., Jr, Scurlock, A. M. & Darville, T. (2007).** Plasmid-deficient *Chlamydia muridarum* fail to induce immune pathology and protect against oviduct disease. *J Immunol* **179**, 4027–4034.

**Patterson, J. L. S., Lynch, M., Anderson, G. A., Noormohammadi, A. H., Legione, A., Gilkerson, J. R. & Devlin, J. M. (2015).** The prevalence and clinical significance of *Chlamydia* infection in island and mainland populations of Victorian koalas (*Phascolarctos cinereus*). *J Wildl Dis* **51**, 309–317.

**Polkinghorne, A., Hanger, J. & Timms, P. (2013).** Recent advances in understanding the biology, epidemiology and control of chlamydial infections in koalas. *Vet Microbiol* **165**, 214–223.

**Robertson, T., Bibby, S., O'Rourke, D., Belfiore, T., Lambie, H. & Noormohammadi, A. H. (2009).** Characterization of *Chlamydiaceae* species using PCR and high resolution melt curve analysis of the 16S rRNA gene. *J Appl Microbiol* **107**, 2017–2028.

**Shojima, T., Yoshikawa, R., Hoshino, S., Shimode, S., Nakagawa, S., Ohata, T., Nakaoka, R. & Miyazawa, T. (2013).** Identification of a novel subgroup of Koala retrovirus from Koalas in Japanese zoos. *J Virol* **87**, 9943–9948.

**Simmons, G. S., Young, P. R., Hanger, J. J., Jones, K., Clarke, D., McKee, J. J. & Meers, J. (2012).** Prevalence of koala retrovirus in geographically diverse populations in Australia. *Aust Vet J* **90**, 404–409.

**Stalder, K., Vaz, P. K., Gilkerson, J. R., Baker, R., Whiteley, P., Ficorilli, N., Tatarczuch, L., Portas, T., Skogvold, K. & other authors (2015).** Prevalence and clinical significance of herpesvirus infection in populations of Australian marsupials. *PLoS One* **10**, e0133807.

**Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994).** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

**Troughton, E. (1941).** *Furred Animals of Australia.* Sydney: Angus & Roberson.

**Wilson, D. P., Craig, A. P., Hanger, J. & Timms, P. (2015).** The paradox of euthanizing koalas (*Phascolarctos cinereus*) to save populations from elimination. *J Wildl Dis* **51**, 833–842.

**Yousef Mohamad, K., Roche, S. M., Myers, G., Bavoil, P. M., Laroucau, K., Magnino, S., Laurent, S., Rasschaert, D. & Rodolakis, A. (2008).** Preliminary phylogenetic identification of virulent *Chlamydophila pecorum* strains. *Infect Genet Evol* **8**, 764–771.

# 5. Koala retrovirus genotyping analyses reveal a low prevalence of KoRV-A in Victorian koalas and an association with clinical disease

**Published by:** Journal of Medical Microbiology, volume 66, 2017.

**Co-authors:** J. L. S. Patterson, P. Whiteley, S. M. Firestone, M. Curnick, K. Bodley, M. Lynch, J. R. Gilkerson, F. M. Sansom and J. M. Devlin

The original publication can be obtained from the below DOI:

dx.doi.org/10.1099/jmm.0.000416

The following material has been included with permission from the journal.

Supplementary table 1 listed in manuscript is in Appendix 5.

JOURNAL OF MEDICAL MICROBIOLOGY

RESEARCH ARTICLE
Legione et al., Journal of Medical Microbiology 2017;66:236–244
DOI 10.1099/jmm.0.000416

MICROBIOLOGY SOCIETY

# Koala retrovirus genotyping analyses reveal a low prevalence of KoRV-A in Victorian koalas and an association with clinical disease

Alistair R. Legione,[1,*] Jade L. S. Patterson,[1,2] Pam Whiteley,[1,3] Simon M. Firestone,[1] Megan Curnick,[1,4] Kate Bodley,[2] Michael Lynch,[1,2] James R. Gilkerson,[5] Fiona M. Sansom[1]† and Joanne M. Devlin[1]†

**Abstract**

**Purpose.** Koala retrovirus (KoRV) is undergoing endogenization into the genome of koalas in Australia, providing an opportunity to assess the effect of retrovirus infection on the health of a population. The prevalence of KoRV in north-eastern Australia (Queensland and New South Wales) is 100 %, whereas previous preliminary investigations in south-eastern Australia (Victoria) suggested KoRV is present at a lower prevalence, although the values have varied widely. Here, we describe a large study of free-ranging koalas in Victoria to estimate the prevalence of KoRV and assess the clinical significance of KoRV infection in wild koalas.

**Methodology.** Blood or spleen samples from 648 koalas where tested for KoRV provirus, and subsequently genotyped, using PCRs to detect the *pol* and env genes respectively. Clinical data was also recorded where possible and analysed in comparison to infection status.

**Results.** The prevalence of KoRV was 24.7 % (160/648). KoRV-A was detected in 141/160 cases, but KoRV-B, a genotype associated with neoplasia in captive koalas, was not detected. The genotype in 19 cases could not be determined. Genomic differences between KoRV in Victoria and type strains may have impacted genotyping. Factors associated with KoRV infection, based on multivariable analysis, were low body condition score, region sampled, and 'wet bottom' (a staining of the fur around the rump associated with chronic urinary incontinence). Koalas with wet bottom were nearly twice as likely to have KoRV provirus detected than those without wet bottom (odds ratio=1.90, 95 % confidence interval 1.21, 2.98).

**Conclusion.** Our findings have important implications for the conservation of this iconic species, particularly regarding translocation potential of Victorian koalas.

## INTRODUCTION

The current process of koala retrovirus (KoRV) endogenization into the koala genome makes KoRV a gammaretrovirus of significant general interest [1]. While the virus was first observed using electron microscopy over 20 years ago [2] and the first full genome published just over a decade later [3], the impact of KoRV on koala health, and the role of viral genotypes in affecting immune function or causing disease, is still limited. Although early studies suggested that leukaemia in koalas may be linked to KoRV [4], this was based on the observation that other gammaretroviruses (gibbon ape leukaemia virus, murine leukaemia virus, feline leukaemia virus) are associated with leukaemia, but no causal relationship between KoRV and leukaemia was established at the time. Later studies linked increased levels of circulating virus in plasma, but not provirus in white blood cells, with development of neoplasia [5]. The detection of a neoplasia-associated genotype (KoRV-B) in captive koalas [6] allowed for the absence of clinical signs in some koalas infected with KoRV to be better understood. KoRV-B differs from KoRV-A (the original virus sequenced) in the envelope

Received 14 October 2016; Accepted 23 December 2016
Author affiliations: [1]Asia Pacific Centre for Animal Health, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria, Australia; [2]Veterinary Department, Melbourne Zoo, Parkville, Victoria, Australia; [3]Wildlife Health Surveillance Victoria, The University of Melbourne, Werribee, Victoria, Australia; [4]Australian Wildlife Health Centre, Healesville Sanctuary, Healesville, Victoria, Australia; [5]Centre for Equine Infectious Diseases, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria, Australia.
*Correspondence: Alistair R. Legione, alegione@student.unimelb.edu.au
Keywords: koala retrovirus; Victoria; genotype; prevalence; KoRV-A; KoRV-B.
Abbreviations: BCS, body condition score; KoRV, koala retrovirus.
†These authors contributed equally to this work.
One supplementary table is available with the online Supplementary Material.

protein-encoding *env* gene. This change in the envelope protein alters the host receptor used for virus attachment. Differences also exist in the long terminal repeats, with KoRV-B containing four 18-nucleotide long tandem repeats and KoRV-A containing only a single copy. More recently, a number of other KoRV genotypes (based on the *env* nucleotide sequence) have been described, namely KoRV-C and -D [7] and KoRV-E and -F [8], as well as variants within the genotypes [9, 10], but the clinical significance of these remains unclear.

In Australian koala populations, it appears that KoRV is undergoing endogenization in a southward direction [1]. This hypothesis is based on investigations of KoRV prevalence variance in wild populations, which ranges from 100 % in northern populations [Queensland and New South Wales (NSW)] to, on average, 15 and 51 % in southern states (South Australia and Victoria, respectively) [11]. Investigations utilizing museum specimens detected KoRV infection in Queensland koalas as early as the late 1800s [9]. However, only two specimens from NSW and none from Victoria or South Australia were used in that investigation, making it difficult to draw conclusions about the presence of the infection across the rest of the country throughout history. It has been shown that in Victorian koalas, the number of proviral copies per genomic unit is significantly less than that seen in northern populations, suggesting an exogenous infection [11, 12]. Information on KoRV genotypes present in modern wild populations is limited. Five unique museum specimens (of wild origin) have been used to investigate genomic variation and no historical evidence of KoRV-B and a low number of single nucleotide polymorphisms across the entire genome were found [9, 13]. KoRV-B has been detected in the transcriptome of a wild koala from Queensland [14] and KoRV-C-like virus detected in a South Australian koala [15], but there are no published large-scale studies of the prevalence of different genotypes. Here, we focus on Victorian populations, of particular interest due to the variation in KoRV infection levels in previous studies [1, 11, 12]. To better estimate the prevalence of KoRV in Victorian koalas, we undertook a rigorous investigation of the KoRV prevalence in a large number of individual koalas in multiple populations. We also aimed to estimate the prevalence of both KoRV-A and KoRV-B in Victorian koalas, and combine this with clinical data collected from free-ranging koalas to assess the clinical significance of KoRV infection. In the absence of a laboratory animal model to study KoRV infection, and with experimental studies in the natural host impracticable, these field-based studies represent the best approach to assessing the clinical significance of KoRV infection.

## METHODS

### Sample collection

Sample collection was approved by The University of Melbourne Animal Ethics Committee (approval numbers 1011687.1 and 1312813.2) and Parks Victoria (Research Permits 10004605, 10006948 and 10005388). Blood samples were collected from 530 live koalas and spleen samples were collected from 135 necropsied koalas, with a total of 648 koalas sampled. A small number of live koalas were later euthanized and necropsied ($n$=17), thus providing both blood and spleen samples. All of the necropsied animals were euthanized for health or welfare reasons unrelated to this study. Koala samples used in this study were collected between 2010 and 2016. Live koalas were sampled either as part of previous research efforts focusing on three regions in Victoria ($n$=271) [16] or as part of management operations ($n$=259). Blood samples were collected from anaesthetized koalas and separated (where possible) into buffy coat and plasma fractions. Both fractions were added to RNAlater stabilization solution (ThermoFisher Scientific). Spleen tissue was stored in 1.5 ml tubes at $-80\,^{\circ}$C until processing for extraction. Clinical examinations of live captured animals were performed by veterinarians, including assessment of body condition score (BCS) (1 to 5) [16], wet bottom presence/absence [17] and, for a subsample of koalas, the presence/absence of urogenital tract disease by ultrasound [16, 18]. Wet bottom is a clinical sign of disease characterized by a soiling of fur, or scalding, around the rump, as a result of urinary incontinence [19]. In contrast, urogenital tract disease represents gross changes to the urogenital tract, such as reproductive cysts. Other parameters that were also recorded included tooth wear class, as a measure of age [20], and the presence or absence of young, as a measure of fecundity. Lymph node enlargement was recorded as a proxy for potential lymphoid leukaemia. Previous data on detecting *Chlamydia pecorum* prevalence were utilized here [16, 21].

### DNA extraction

DNA was extracted from either whole blood or separated buffy coat fractions. Spleen tissue was first vigorously homogenized using a plastic shafted rayon swab (Copan Italia), which was then added to 800 µl of PBS (137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$, 2 mM $KH_2PO_4$, pH 7.4) and vortexed briefly. DNA was extracted from 200 µl of the blood or spleen samples using an X-tractor robot (Qiagen), utilizing the Qiaxtractor VX extraction kit as per manufacturer's instructions. Each bulk extraction contained both positive and negative extraction controls. The positive control was a diluted liquid culture of *E. coli* containing a portion of *pol* inserted into pGEM-T (Promega) plasmid, and the negative control was sterile PBS.

### Detection of KoRV provirus and *env* genotypes

Extracted DNA was tested for KoRV provirus using the KoRV *pol* qPCR primers described previously [5], but replacing the probe with Platinum SYBR Green qPCR SuperMix-UDG (Invitrogen). The housekeeping gene $\beta$-actin was utilized to standardize the genome copy numbers detected in each sample, as described previously [22]. Results were reported as the number of KoRV *pol* copies per $\beta$-actin copies in the extracted liquid sample.

Differentiation of KoRV-A and KoRV-B was conducted using genotype-specific primers as described previously [6], but replacing the genotype-specific probes with Platinum SYBR Green qPCR SuperMix-UDG. Our protocol also used ROX as a reference dye. Reactions were performed using an MX3000P real-time PCR machine (Stratagene). A melt curve with a 0.3 °C resolution from 75 to 95 °C was used to determine the amplification of the correct product, with comparison to positive controls.

A standard curve was employed for each qPCR consisting of 10-fold dilutions, in triplicate, of purified plasmid containing one of either KoRV *pol*, $\beta$-actin, KoRV-A *env* or KoRV-B *env* target genes from $10^7$ to $10^1$ copies per reaction. Copy numbers were calculated using a Qubit 3.0 fluorometer (Invitrogen) and dilutions prepared using a QIAgility robot (Qiagen).

### Statistical analyses

Univariable and multivariable statistical analyses to estimate associations with KoRV detection were conducted using Stata 14.1 (StataCorp) software. Explanatory variables included sex, age (based on tooth wear), BCS (1 to 5), reproductive and urinary tract disease (presence/absence), fecundity in females (based on the presence/absence of back or pouch young), wet bottom presence/absence, lymph node enlargement, *C. pecorum* presence/absence and geographic region of sampled koalas. Not all parameters were available for all animals. Koalas were classified as young (tooth wear<class III), mature adult (tooth wear between classes III andV, inclusive) or old adult (tooth wear>class V) [16]. Specific location data was pooled into seven broader regional locations (Fig. 1). As *C. pecorum* genotypes in Victoria are largely homogeneous and have no known pathogenic variation [21], only the presence or absence of the organism was considered. Univariable logistic regression was performed to estimate the association of each explanatory variable with an outcome variable representing the presence or absence of KoRV. Multivariable logistic regression models were fitted to the data using a manual backwards stepwise approach [23]. Variables with a *P* value$\leq$0.25 in univariable analysis were considered for the multivariable model as well as potential confounders. At each step, all eligible candidate variables were individually tested and removed from the model if they had a *P* value>0.10 based on the likelihood ratio test. First-order biologically plausible interaction terms were added to the final main-effects model and tested for statistical

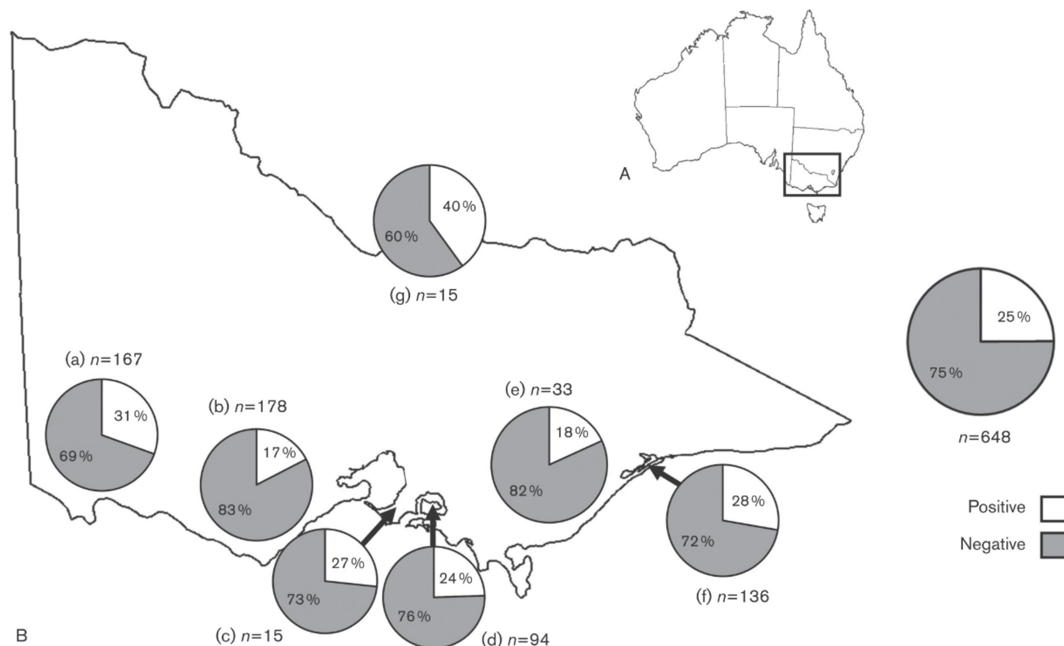



**Fig. 1.** Prevalence of koala retrovirus (KoRV) in sampled Victorian koala populations. (A) Map of Australia highlighting the state of Victoria. (B) Regional prevalence of KoRV detected in Victorian koala blood and spleen samples. Regions: (a) Far Western Victoria, (b) South West Coast, (c) Mornington Peninsula, (d) French Island, (e) Greater Gippsland, (f) Raymond Island, (g) Far North Victoria. Map not to scale. Regions not recorded for 10 koalas.

significance at $P<0.05$. Standard regression diagnostics were performed, including assessment of goodness-of-fit using the Hosmer–Lemeshow technique [23]. To assess the effects of clustering of observations from the same region, we refitted the final model using multilevel mixed-effects logistic regression, with a regional cluster-level random effects term.

Standardized KoRV copy numbers were logarithmically transformed but remained severely skewed, so comparison of KoRV provirus copies in each sample in relation to associated factors (sex, region, BCS, wet bottom presence/absence, urinary and reproductive tract disease presence/absence, lymph node enlargement and *C. pecorum* detection) was undertaken using non-parametric (Kruskal–Wallis) statistical tests. Outliers (determined using Tukey's outlier test) were removed from the dataset.

## RESULTS

### Prevalence of KoRV in Victorian koala populations

The prevalence of KoRV in each specific Victorian region ranged from 17 to 40 % of tested koalas, with an overall prevalence of KoRV in Victorian koalas of 24.7 % (95 % CI: 21.4, 28.2 %) (160/648) (Fig. 1). KoRV was significantly more likely to be detected in samples collected from koalas in Raymond Island, Far Western and Far Northern Victoria than in samples from koalas in the South West Coast region (Table 1). The South West Coast region was selected for use as the reference region as it had the largest number of samples available for comparison. The proportion of spleen samples from which KoRV was detected (36/135, 26.7 %) was comparable to the prevalence in blood samples (127/530, 24.0 %). KoRV-B was not detected in any of the 160 KoRV-positive koalas, whereas KoRV-A was detected in 141 koalas. In 19 *pol*-positive cases, *env* could not be amplified using either the KoRV-A or KoRV-B specific primers, and attempts to amplify the entire gene using previously published primers [24] were unsuccessful.

The prevalence of KoRV in koalas sampled from four regional populations could be compared across two time points, separated by either 2 years (South West Coast) or 3 years (French Island, Raymond Island and Far Western Victoria). In all four cases, there was no clear trend in the proportion of KoRV detections over time (Table 2).

### Relationship between KoRV infection and clinical observations

Univariable analyses revealed that koalas with wet bottom had 2.20 (95 % CI: 1.48, 3.27) times the likelihood of being

**Table 1.** Outputs of univariable logistic regression assessing demographic variables for associations with the presence of KoRV DNA in 648 koalas from Victoria, Australia, sampled between 2010 and 2016

| Variable | KoRV positive/*n* | Prevalence (%) | Odds ratio* | 95 % CI | *P* value† |
|---|---|---|---|---|---|
| Sex | | | | | **0.19** |
| Male | 37/174 | 21.3 | 0.76 | 0.50, 1.15 | |
| Female | 121/460 | 26.3 | 1.00 | – | |
| Not recorded | 2/14 | – | – | – | |
| Age | | | | | 0.67 |
| Young | 30/115 | 26.1 | 1.14 | 0.71, 1.83 | |
| Mature adult | 100/424 | 23.6 | 1.00 | – | |
| Old adult | 13/45 | 28.9 | 1.32 | 0.67, 2.60 | |
| Not recorded | 17/64 | – | – | – | |
| Back/pouch young or lactation (females only) | | | | | 0.28 |
| Yes | 37/160 | 23.1 | 0.77 | 0.47, 1.25 | |
| No | 53/188 | 28.2 | 1.00 | – | |
| Not recorded | 31/112 | – | – | – | |
| Region (west to east) | | | | | **0.070** |
| Far Western Victoria | 51/167 | 30.5 | 2.08 | 1.25, 3.47 | |
| South West Coast | 31/178 | 17.4 | 1.00 | – | |
| Far North Victoria | 6/15 | 40.0 | 3.16 | 1.05, 9.53 | |
| Mornington Peninsula | 4/15 | 26.7 | 1.72 | 0.52, 5.77 | |
| French Island | 23/94 | 24.5 | 1.54 | 0.84, 2.83 | |
| Greater Gippsland | 6/33 | 18.2 | 1.05 | 0.40, 2.77 | |
| Raymond Island | 38/136 | 27.9 | 1.84 | 1.07, 3.15 | |
| Other/not recorded | 1/10 | – | – | – | |

*Reference levels are indicated by odds ratios of 1.0. An interpretation of the odds ratio here would be that the odds of male koalas testing positive for KoRV DNA were 24 % less than those of female koalas. CI, confidence interval; –, not measured.

†*P* values estimated using the likelihood ratio test statistic. Results highlighted in bold ($P\leq0.25$) represent variables included in the initial multivariable model.

**Table 2.** Comparison of prevalence of KoRV within regional populations over time

| Region | Year sampled | KoRV positive/*n* | Prevalence (%) | 95 % CI | *P* value* |
|---|---|---|---|---|---|
| Far Western Victoria | 2010 | 39/120 | 32.5 | 24.2, 41.7 | 0.45 |
| | 2013 | 11/44 | 25.0 | 13.2, 40.3 | |
| South West Coast | 2013 | 11/56 | 19.6 | 10.2, 32.4 | 0.51 |
| | 2015 | 15/99 | 15.2 | 8.7, 23.8 | |
| French Island | 2010 | 15/63 | 23.8 | 14.0, 36.2 | 0.80 |
| | 2013 | 8/30 | 26.7 | 12.3, 45.9 | |
| Raymond Island | 2010 | 25/88 | 28.4 | 19.3, 39.0 | 0.83 |
| | 2013 | 11/36 | 30.6 | 16.3, 48.1 | |

CI, confidence interval.

**P* value estimated using Fisher's exact method (significant where $P<0.05$).

KoRV positive (Table 3). Koalas with the lowest BCS (a score of 1 out of 5) were 7.04 (95 % CI 1.98–25.00) times more likely to be KoRV positive than the reference population (a BCS of 3) (Table 3). We did not detect an association between KoRV positivity and the detection of *C. pecorum* ($P=0.84$). In the final multivariable logistic regression model (Table 4), the variables significantly associated with KoRV detection included the presence/absence of wet bottom, low BCS (<2) and region.

Repeating the multivariable analysis treating urogenital tract disease (commonly associated with *C. pecorum*) in female koalas as the outcome variable showed that KoRV was not statistically significantly associated with urogenital tract disease on its own ($P=0.21$). Although not statistically significant in this analysis, there was a clinically important increase in the odds of urogenital tract disease when *C. pecorum* and KoRV were both detected (odds ratio=20.5; 95 % CI: 0.97, 433.7) compared to the odds of urogenital tract disease when only *C. pecorum* was detected (odds ratio=5.71; 95 % CI: 1.04, 31.3); see Table S1 (available in the online Supplementary Material).

Kruskal–Wallis tests comparing log-transformed KoRV *pol* copies per $\beta$-actin copies found that median normalized proviral copy numbers were highly comparable in most categories tested (sex, region, BCS, wet bottom presence/absence, *C. pecorum* presence/absence, lymph node enlargement); however, a statistically significant increase in detected copy numbers was identified for female koalas with urogenital tract disease (Table 5). The median normalized KoRV proviral copy number of all samples was 10 KoRV copies per 1000 $\beta$-actin copies (range: 0.1 to 398 KoRV copies per 1000 $\beta$-actin copies).

## DISCUSSION

The detection of KoRV in a relatively low proportion of Victorian koalas is of significance for the future conservation of the koala across Australia. Previous research suggested that the prevalence in Victorian populations such as in Gippsland and the Strzelecki Ranges was as high as 63 % (29/46), while an assortment of samples from mainland Victoria had a detection rate of 84 % (36/43) [11]. Our research, which incorporates a larger number of test samples than in previous studies, from multiple regions across the state, suggests that the overall prevalence (for our sampled regions) is much lower than previously suspected. The sampling methodology utilized may have played a role in our reduced prevalence finding, with random sampling of free-ranging individuals less likely to be biased towards diseased individuals. The median normalized KoRV provirus detected fell within the range seen in a previous study of Victorian koalas that compared tissue and faecal samples [12].

Despite the current hypothesis of a north-to-south transmission of infection [1], there is no strong evidence in regard to the history of the virus in Victorian populations aside from its presence on French Island, as also detected previously [11]. French Island is a closed population that was established around 1900 with very few individuals [25]. Based on detailed records, the island has not had new individuals added since this initial translocation [26]. Therefore, the presence of KoRV within the French Island koala population suggests the virus was present in the founding animals and yet it has not managed to reach a high prevalence in this small, closed population. Historical museum specimens have shown that KoRV could not be detected in an individual NSW koala from 1899 (a similar time point to when the French Island population was established) but was detected in an individual NSW koala from 1971 [9]. We have shown, on a much shorter time scale, that there is no significant change in the prevalence of the infection when populations were sampled up to 3 years apart. While establishing correct information from small sample sizes (in the historical cases) or narrow time lines (in our case) has limitations, it may inform future modelling of viral spread in southern populations. Longer time-course studies of select populations and assessment of historical samples from Victorian origin could help to establish a trend of KoRV prevalence over time, to determine whether the proportion of infected koalas is increasing, decreasing or static. Such studies could be aided by recent findings that KoRV can be detected in faecal samples [12], making it easier to

**Table 3.** Outputs of univariable logistic regression assessing clinical variables for associations with the presence of KoRV DNA in 648 koalas from Victoria, Australia, sampled between 2010 and 2016

| Variable | KoRV positive/*n* | Prevalence (%) | Odds ratio* | 95 % CI | *P* value† |
|---|---|---|---|---|---|
| Body condition score (BCS) | | | | | **0.0082** |
| BCS 1 – Emaciated | 7/11 | 63.6 | 7.04 | 1.98, 25.0 | |
| BCS 2 | 16/75 | 21.3 | 1.09 | 0.58, 2.06 | |
| BCS 3 – Healthy | 49/246 | 19.9 | 1.00 | – | |
| BCS 4 | 52/186 | 28.0 | 1.56 | 1.00, 2.44 | |
| BCS 5 – Obese | 5/12 | 41.7 | 2.87 | 0.87, 9.44 | |
| Not recorded | 31/118 | – | – | – | |
| Wet bottom | | | | | **<0.001** |
| Present | 56/153 | 36.6 | 2.20 | 1.48, 3.27 | |
| Absent | 97/466 | 20.8 | 1.00 | – | |
| Not recorded | 7/29 | – | – | – | |
| Urinary tract disease (females only) | | | | | 0.91 |
| Present | 7/21 | 33.3 | 1.06 | 0.38, 2.97 | |
| Absent | 24/75 | 32.0 | 1.00 | – | |
| Not recorded | 90/364 | – | – | – | |
| Reproductive tract disease (females only) | | | | | 0.12 |
| Present | 9/19 | 47.4 | 2.25 | 0.81, 6.29 | |
| Absent | 22/77 | 28.6 | 1.00 | – | |
| Not recorded | 90/364 | – | – | – | |
| *Chlamydia pecorum* | | | | | 0.84 |
| Present | 29/115 | 25.2 | 1.05 | 0.66, 1.68 | |
| Absent | 123/506 | 24.3 | 1.00 | – | |
| Not recorded | 8/27 | – | – | – | |
| Lymph node enlargement | | | | | 0.25‡ |
| Present | 7/33 | 21.2 | 0.61 | 0.25, 1.47 | |
| Absent | 65/212 | 30.7 | 1.00 | – | |
| Not recorded | 88/403 | – | – | – | |

CI, confidence interval; –, not measured.

*Reference levels are indicated by odds ratios of 1.0.

†*P* values estimated using the likelihood ratio test statistic. Results highlighted in bold ($P \leq 0.25$) represent variables included in the initial multivariable model.

‡Excluded from multivariable model due to >20 % missing data.

accumulate the sample sizes required to assess a population. This finding would be particularly beneficial in regions of northeastern Victoria and southern NSW, where KoRV testing has been limited.

The absence of KoRV in a large proportion of koalas in Victoria also allows us to investigate links between clinical disease and KoRV presence, which is not possible in northern koalas where KoRV prevalence is 100 %. Interestingly, we found that koalas with KoRV were almost twice as likely to be recorded as having wet bottom, which is commonly associated with *C. pecorum* infection. Recently, we showed that a large number of Victorian koalas exhibiting mild wet bottom signs were negative for *C. pecorum* [16], and our subsequent studies demonstrated that the association between wet bottom and *C. pecorum* detection was only significant in male koalas [21]. KoRV may be directly contributing to wet bottom through a yet unknown mechanism or indirectly by causing immunosuppression that allows opportunistic pathogens to cause urogenital infections.

Our results did not show a significant link between the presence of *C. pecorum* and KoRV in Victorian koalas, nor did the inclusion of KoRV improve a multivariable model assessing associations between urogenital tract disease, koala age, wet bottom and *C. pecorum* in female koalas. However, there was a near significant interaction between infection with both KoRV and *C. pecorum,* and the detection of urogenital tract disease. Our findings disagree with a hypothesis that immunosuppression due to KoRV infection would increase the likelihood of subsequent infection with *C. pecorum* [27]. However, collection of more data could more accurately determine whether there is an increase in the development of the clinical signs associated with *C. pecorum* as a result of concurrent KoRV infection [28].

**Table 4.** Outputs of multivariable logistic regression assessing clinical and demographic variables for associations with the presence of KoRV DNA in 521 koalas from Victoria, Australia, sampled between 2010 and 2016

| Variable | OR* | SE (OR) | 95 % CI | *P* value† |
|---|---|---|---|---|
| Wet bottom | | | | |
| Present | 1.90 | 0.44 | 1.21, 2.98 | **0.005** |
| Absent | 1.00 | – | – | – |
| Body condition score (BCS) | | | | |
| BCS 1 – Emaciated | 8.06 | 7.02 | 1.46, 44.4 | **0.017** |
| BCS 2 | 1.39 | 0.47 | 0.71, 2.70 | 0.34 |
| BCS 3 – Healthy | 1.00 | – | – | – |
| BCS 4 | 1.20 | 0.31 | 0.72, 2.00 | 0.49 |
| BCS 5 – Obese | 2.51 | 1.57 | 0.73, 8.58 | 0.14 |
| Region (west to east) | | | | |
| Far Western Victoria | 2.16 | 0.70 | 1.14, 4.09 | **0.018** |
| South West Coast | 1.00 | – | – | – |
| French Island | 1.79 | 0.62 | 0.91, 3.52 | 0.090 |
| Raymond Island | 1.53 | 0.52 | 0.79, 2.98 | 0.21 |
| | | – | | |
| Constant | 0.14 | 0.62 | 0.08, 0.23 | <0.001 |

*n*=521; log likelihood=−274.91906; *P*<0.001; d.f.=9; Akaike information criterion=567.838; Hosmer–Lemeshow goodness-of-fit test (*P*=0.83).

*OR, odds ratio; SE, standard error; –, not measured. Reference levels are indicated by odds ratios of 1.0.

†*P* values estimated using the likelihood ratio test statistic. Results highlighted in bold (*P*<0.05).

The absence of KoRV-B in our samples is significant given recent suggestions that KoRV-B may be associated with the development of neoplasia. With a sample size of 160 KoRV *pol* koalas, our study would have been expected, with 95 % confidence, to have detected KoRV-B if it was present at a rate of 2 % or higher [29]. However, the methodology used to detect KoRV-B in this project may have not detected variants within this genotype should the variants contain nucleotide changes in the primer binding region, nor can it determine nucleotide changes in regions of the KoRV *env* gene outside the qPCR target, as has previously been demonstrated [10]. A recent publication used deep sequencing on 18 individual koalas from South East Queensland and noted a large number of variants, which were grouped into nine different genotypes of KoRV (A to I) [30]. Further analysis utilizing similar deep-sequencing approaches should be considered in future studies of Victorian koalas. This would allow the detection of all genotypes that may be present (and variants within those genotypes) both in the individuals where KoRV-A was detected as well as in the cases from which *pol* was amplified successfully but *env* was not.

While the potentially clinically important KoRV-B has now been detected in wild northern koalas [30], a prevalence survey has not been undertaken. In the initial findings in Californian zoos, KoRV-B was detected in 46 % (6/13) of captive koalas at Los Angeles Zoo, all of which were koalas originating from northern Australia [6]. In that study, the authors determined that KoRV-B was transmitted from dam to offspring. It is possible that the high percentage of koalas found to have KoRV-B in this population was influenced by captive breeding, resulting in amplification of a relatively rare genotype. This may also explain the high rates of neoplasia seen in captive koalas in zoos in northern Australia, with 60 to 80 % of deaths in captivity attributed to leukaemia and lymphoma [3, 5], while these diseases were detected in only 2 % of free-ranging koalas admitted to one Queensland animal hospital [28]. There was no association in our study between lymph node enlargement, a potential sign of lymphoid leukaemia [28], and the detection of KoRV provirus. Other pathogens may also be influencing the prevalence of neoplasia detected in some koala populations. For example, gammaherpesviruses, two of which have been detected in koalas [31, 32], can cause lymphoma in other species including mice [33] and humans [34]. In Victoria, herpesviruses were detected in 33/99 koalas sampled in one study, and a significant association was found between koalas with herpesvirus and those detected with *C. pecorum* [35]. A broader epidemiological assessment of herpesvirus in koalas Australia-wide would be worthwhile.

The presence of KoRV in only a quarter of koalas tested, and the apparent absence of the neoplasia-associated KoRV-B genotype in sampled Victorian koalas (based on methodology used), is of importance to koala conservation efforts. Despite lower genetic diversity in Victorian koalas compared to northern koalas [36, 37] and current research suggesting that the genetic diversity of koalas across Australia is impacted by their historical geographic barriers, koalas still belong to a single evolutionary significant unit [38]. It, therefore, follows that individuals could be translocated between populations without mixing of subspecies. A lower risk of the presence of KoRV and *Chlamydia* infection [21] may make Victorian koalas good candidates for future translocation programmes to establish or re-establish populations in northern regions of Australia, particularly in

**Table 5.** Comparisons of median KoRV provirus copies per 1000 β-actin copies within variable groups for up to 146 koalas

| Variable | Samples | Median | Interquartile range (Q1, Q3) | *P* value* |
|---|---|---|---|---|
| Sex | | | | 0.60 |
| Male | 31 | 8.63 | 2.74, 78.0 | |
| Female | 113 | 10.7 | 5.72, 23.9 | |
| Not recorded | 2 | – | – | |
| Region (west to east) | | | | |
| Far Western Victoria | 51 | 12.8 | 7.48, 21.7 | 0.18 |
| South West Coast | 24 | 6.60 | 2.80, 65.0 | |
| Far North Victoria | 5 | 66.3 | 5.10, 236.3 | |
| Mornington Peninsula | 3 | 4.61 | 2.82, 5.87 | |
| French Island | 22 | 12.2 | 4.73, 31.8 | |
| Greater Gippsland | 6 | 6.80 | 4.00, 51.8 | |
| Raymond Island | 34 | 8.80 | 4.65, 56.8 | |
| Other/not recorded | 1 | – | – | |
| Body condition score (BCS) | | | | 0.61 |
| BCS 1 – Emaciated | 6 | 12.7 | 8.10, 50.5 | |
| BCS 2 | 13 | 4.16 | 2.85, 55.5 | |
| BCS 3 – Healthy | 48 | 12.6 | 3.99, 31.9 | |
| BCS 4 | 50 | 11.4 | 7.19, 27.4 | |
| BCS 5 – Obese | 5 | 15.3 | 4.39, 24.0 | |
| Not recorded | 24 | – | – | |
| Wet bottom | | | | 0.85 |
| Present | 53 | 8.97 | 4.72, 45.6 | |
| Absent | 92 | 11.4 | 5.21, 31.9 | |
| Not recorded | 1 | – | – | |
| Urinary tract disease (females only) | | | | **0.033** |
| Present | 7 | 21.7 | 13.4, 110.8 | |
| Absent | 23 | 7.37 | 4.16, 13.6 | |
| Not recorded | 83 | – | – | |
| Reproductive tract disease (females only) | | | | **0.040** |
| Present | 9 | 15.3 | 4.00, 17.1 | |
| Absent | 21 | 6.90 | 10.7, 22.2 | |
| Not recorded | 83 | – | – | |
| *Chlamydia pecorum* | | | | |
| Present | 27 | 11.1 | 6.03, 22.6 | 0.73 |
| Absent | 114 | 10.5 | 4.59, 31.8 | |
| Not recorded | 5 | – | – | |
| Lymph node enlargement | | | | 0.70 |
| Present | 7 | 9.0 | 4.20, 134.1 | |
| Absent | 63 | 10.7 | 5.09, 27.1 | |
| Not recorded | 76 | – | – | |

**P* values estimated using the Kruskal–Wallis test for the equality of medians. Significant differences ($P<0.05$) are in bold.

NSW. Furthermore, it indicates they are a potentially more suitable population from which to derive animals for captive populations in Australia and internationally. A key barrier to this translocation of Victorian koalas is the current lack of accurate, rapid diagnostic tests for the detection of KoRV and *Chlamydia*, without which animals would be required to undergo lengthy holding periods before introduction into their new environment. Future research to develop such tests is warranted.

Funding information
Funding was provided for this research by the Holsworth Wildlife Research Endowment – Equity Trustees Charitable Foundation, the Vizard Foundation and the Albert George and Nancy Caroline Youngman Trust.

Acknowledgements
We gratefully acknowledge all of the assistance we have received with collection of samples and clinical information by staff and students of The University of Melbourne, Wildlife Health Surveillance Victoria, Zoos Victoria, Victorian Department of Environment, Land, Water and

Planning, and Parks Victoria. We thank Damien Higgins, Iona Maher and Andrea Casteriano of the Koala Health Hub at The University of Sydney for assistance with the KoRV-*env* qPCRs. Alistair Legione is supported by an Australian Postgraduate Award.

**Conflicts of interest**
The authors declare that there are no conflicts of interest.

**References**


1. **Tarlinton RE, Meers J, Young PR.** Retroviral invasion of the koala genome. *Nature* 2006;442:79–81.
2. **Canfield PJ, Sabine JM, Love DN.** Virus particles associated with leukaemia in a koala. *Aust Vet J* 1988;65:327–328.
3. **Hanger JJ, Bromham LD, Mckee JJ, O'Brien TM, Robinson WF.** The nucleotide sequence of koala (*Phascolarctos cinereus*) retrovirus: a novel type C endogenous virus related to Gibbon ape leukemia virus. *J Virol* 2000;74:4264–4272.
4. **O'Brien TM, Hanger JJ, McKee JJ, Robinson WF.** The isolation, characterisation and partial gene sequence of a retrovirus from koalas. In: *Conference on the Status of the Koala*. Soldiers Point, Port Stephens, NSW: Australian Koala Foundation; 1997. pp. 18–20.
5. **Tarlinton R, Meers J, Hanger J, Young P.** Real-time reverse transcriptase PCR for the endogenous koala retrovirus reveals an association between plasma viral load and neoplastic disease in koalas. *J Gen Virol* 2005;86:783–787.
6. **Xu W, Stadler CK, Gorman K, Jensen N, Kim D *et al.*** An exogenous retrovirus isolated from koalas with malignant neoplasias in a US zoo. *Proc Natl Acad Sci USA* 2013;110:11547–11552.
7. **Shimode S, Nakaoka R, Hoshino S, Abe M, Shogen H *et al.*** Identification of cellular factors required for the budding of koala retrovirus. *Microbiol Immunol* 2013;57:543–546.
8. **Xu W, Gorman K, Santiago J, Kluska K, Eiden M.** Genetic diversity of koala retroviral envelopes. *Viruses* 2015;7:1258–1270.
9. **Ávila-Arcos MC, Ho SYW, Ishida Y, Nikolaidis N, Tsangaras K *et al.*** One hundred twenty years of koala retrovirus evolution determined from museum skins. *Mol Biol Evol* 2013;30:299–304.
10. **Ishida Y, McCallister C, Nikolaidis N, Tsangaras K, Helgen KM *et al.*** Sequence variation of koala retrovirus transmembrane protein p15E among koalas from different geographic regions. *Virology* 2015;475:28–36.
11. **Simmons GS, Young PR, Hanger JJ, Jones K, Clarke D *et al.*** Prevalence of koala retrovirus in geographically diverse populations in Australia. *Aust Vet J* 2012;90:404–409.
12. **Wedrowicz F, Saxton T, Mosse J, Wright W, Hogan FE.** A non-invasive tool for assessing pathogen prevalence in koala (*Phascolarctos cinereus*) populations: detection of *Chlamydia pecorum* and koala retrovirus (KoRV) DNA in genetic material sourced from scats. *Conservation Genet Resour* 2016:1–11.
13. **Tsangaras K, Siracusa MC, Nikolaidis N, Ishida Y, Cui P *et al.*** Hybridization capture reveals evolution and conservation across the entire koala retrovirus genome. *PLoS One* 2014;9:e95633.
14. **Hobbs M, Pavasovic A, King A, Prentis P, Eldridge M *et al.*** A transcriptome resource for the koala (*Phascolarctos cinereus*): insights into koala retrovirus transcription and sequence diversity. *BMC Genomics* 2014;15:786.
15. **Meers J, Simmons G, Jones K, Clarke DTW, Young PR.** Koala retrovirus in free-ranging populations – prevalence. *Tech Rep Aust Mus* (online) 2014;24:15–17.
16. **Patterson JLS, Lynch M, Anderson GA, Noormohammadi AH, Legione AR *et al.*** The prevalence and clinical significance of *Chlamydia* infection in island and mainland populations of Victorian koalas (*Phascolarctos cinereus*). *J Wildl Dis* 2015;51:309–317.
17. **Griffith JE.** Studies into the diagnosis, treatment and management of chlamydiosis in koalas. PhD Thesis, The University of Sydney; 2010.
18. **Marschner C, Flanagan C, Higgins DP, Krockenberger MB.** Validation of ultrasonography in detecting structural disease of the urogenital tract of the koala, *Phascolarctos cinereus*. *Aust Vet J* 2014;92:177–178.
19. **Dickens RK.** The koala in health and disease. In: *Refresher Course for Veterinarians Proceedings No. 29*. Sydney, NSW: The University of Sydney; 1976. pp. 22–23.
20. **Martin RW.** Age-specific fertility in three populations of the koala, *Phascolarctos cinereus* Goldfuss, in Victoria. *Wildl Res* 1981;8:275–283.
21. **Legione AR, Patterson JLS, Whiteley PL, Amery-Gale J, Lynch M *et al.*** Identification of unusual *Chlamydia pecorum* genotypes in Victorian koalas (*Phascolarctos cinereus*) and clinical variables associated with infection. *J Med Microbiol* 2016;65:420–428.
22. **Shojima T, Yoshikawa R, Hoshino S, Shimode S, Nakagawa S *et al.*** Identification of a novel subgroup of koala retrovirus from koalas in Japanese zoos. *J Virol* 2013;87:9943–9948.
23. **Hosmer DW, Lemeshow S.** Assessing the fit of the model. In: Shewhart WA and Wilks SS (editors). *Applied Logistic Regression*, 2 ed. Hoboken, NJ: Wiley; 2005. pp. 143–202.
24. **Shojima T, Hoshino S, Abe M, Yasuda J, Shogen H *et al.*** Construction and characterization of an infectious molecular clone of koala retrovirus. *J Virol* 2013;87:5081–5088.
25. **Martin RW, Handasyde KA.** *The Koala: Natural History, Conservation and Management*, 2nd ed. Sydney, NSW: UNSW Press Ltd; 1999.
26. **Martin RW.** *Draft Management Plan for the Conservation of the Koala (Phascolarctos Cinereus) in Victoria*. Melbourne, Vic: Arthur Rylah Institute for Environmental Research; 1989.
27. **Denner J.** Immunization with envelope proteins of the KoRV as a basis for a preventive vaccine. *Tech Rep Aust Mus* (online) 2014; 24:71–77.
28. **Hanger JJ, Loader J.** Disease in wild koalas (*Phascolarctos cinereus*) with possible koala retrovirus involvement. *Tech Rep Aust Mus* (online) 2014;24:19–29.
29. **Cannon R, Roe RT.** *Livestock Disease Surveys: A Field Manual for Veterinarians*. Canberra, ACT: Australian Government Publishing Service; 1982.
30. **Chappell KJ, Brealey JC, Amarilla AA, Watterson D, Hulse L *et al.*** Phylogenetic diversity of koala retrovirus within a wild koala population. *J Virol* 2016.
31. **Vaz P, Whiteley PL, Wilks CR, Duignan PJ, Ficorilli N *et al.*** Detection of a novel gammaherpesvirus in koalas (*Phascolarctos cinereus*). *J Wildl Dis* 2011;47:787–791.
32. **Vaz P, Whiteley PL, Wilks CR, Browning GF, Gilkerson JR *et al.*** Detection of a second novel gammaherpesvirus in a free-ranging koala (*Phascolarctos cinereus*). *J Wildl Dis* 2012;48:226–229.
33. **Sunil-Chandra NP, Arno J, Fazakerley J, Nash AA.** Lymphoproliferative disease in mice infected with murine gammaherpesvirus 68. *Am J Pathol* 1994;145:818–826.
34. **Hausen HZ, Schulte-Holthausen H, Klein G, Henle W, Henle G *et al.*** Epstein–Barr virus in Burkitt's lymphoma and nasopharyngeal carcinoma: EBV DNA in biopsies of Burkitt tumours and anaplastic carcinomas of the nasopharynx. *Nature* 1970;228:1056–1058.
35. **Stalder K, Vaz PK, Gilkerson JR, Baker R, Whiteley P *et al.*** Prevalence and clinical significance of herpesvirus infection in populations of Australian marsupials. *PLoS One* 2015;10:e0133807.
36. **Lau Q, Jaratlerdsiri W, Griffith JE, Gongora J, Higgins DP.** MHC class II diversity of koala (*Phascolarctos cinereus*) populations across their range. *Heredity* 2014;113:287–296.
37. **Houlden BA, England PR, Taylor AC, Greville WD, Sherwin WB.** Low genetic variability of the koala *Phascolarctos cinereus* in south-eastern Australia following a severe population bottleneck. *Mol Ecol* 1996;5:269–281.
38. **Neaves LE, Frankham GJ, Dennison S, Fitzgibbon S, Flannagan C *et al.*** Phylogeography of the koala, (*Phascolarctos cinereus*), and harmonising data to inform conservation. *PLoS One* 2016;11: e0162207.

# 6. Variation in the microbiome of the urogenital tract of female koalas with and without 'wet bottom'

## 6.1 Introduction

A significant pathogen of koalas, *C. pecorum*, has been a main focus of koala infectious disease investigations since its discovery. *C. pecorum* has been commonly described as the causative agent of the clinical sign known as 'wet bottom' (Dickens 1976; Brown & Grice 1984; Markey *et al.* 2007; Marschner *et al.* 2014). This staining or scalding of the rump is associated with cystitis due to *C. pecorum* infection in some populations (Brown *et al.* 1987), but recently samples from a large number of koalas from Victorian populations with mild wet bottom were negative via qPCR for *C. pecorum* (Patterson *et al.* 2015). In particular, koalas on French Island, a population considered at the time to be free of *C. pecorum* (Martin & Handasyde 1999) had a similar prevalence and severity of wet bottom to populations where *C. pecorum* occurred in more than 35% of koalas tested. Results in chapter 4 demonstrated that whilst wet bottom could be significantly linked to the detection of *C. pecorum* infection in male Victorian koalas, this relationship was not significant in females in that study. It may be that an unidentified organism is causing these mild clinical signs of disease in koalas. To date there has not been extensive research to determine the normal flora of the koala urogenital tract, making it difficult to use traditional microbiological techniques to determine species of interest. Modern sequencing technology, specifically 16S rRNA biodiversity profiling, can be used to improve our understanding of the microbiome of the urogenital tract of koalas, and allow preliminary comparisons of the microbiome of the urogenital tract of female koalas with and without mild wet bottom.

## 6.2 Methods

### 6.2.1 Sample Collection and initial screening

Samples used in this chapter were a subset of urogenital swabs, from female koalas, used in chapter 4. Koala samples were collected as described in section 2.1. The presence and severity of wet bottom was assessed as described in section 2.1.1. Briefly, wet bottom scores are used to grade the severity of clinical findings relating to wet bottom from 0 (absent) to 10 (most severe). For the purpose of this study, koalas were reclassified as either wet bottom 'absent' or wet bottom 'present', based on a score of

either 0, or greater than 0, respectively. After screening all samples for *Chlamydiaceae* using a previously described qPCR (section 2.3.2), ten samples were selected from female koalas where no *Chlamydiaceae* was detected. Five samples collected from koalas showing no clinical signs of urogenital disease manifesting as wet bottom (wet bottom absent) and five samples collected from koalas that showed clinical signs of wet bottom (wet bottom present) were utilised (Table 6.1).

### 6.2.2 Amplification and sequencing

DNA extraction and amplification from the swab samples was performed commercially by The Australian Genome Research Facility (Australia). Variable regions three and four of bacterial 16S rRNA were amplified using primers 341F (5' CCTAYGGGRBGCASCAG 3') and 806R (5' GGACTACNNGGGTATCTAAT 3'). Sequencing was performed on the Illumina MiSeq platform to produce paired end reads of 300 bp (2 × 300 bp).

### 6.2.3 Quality filtering and OTU assignment

Quality filtering and OTU assignment was undertaken using a mixture of scripts and algorithms available in the programs USEARCH 8.1 (Edgar 2010) and QIIME 1.9.1 (Quantitative Insights Into Microbial Ecology) (Caporaso *et al.* 2010b). Unless otherwise stated, default settings were used for all scripts. Read processing to reduce errors was undertaken as described by Edgar and Flyvbjerg (2015). The forward and reverse 300 bp paired-end reads for each swab sample were merged using the USEARCH script **fastq_mergepairs**. In this process, the Phred score of overlapping bases is recalculated to improve error calling. Bases with the same nucleotide called in both the forward and reverse reads have an increased recalculated score, and those with disagreements are reduced. This increases confidence in the calculated error probability of the merged reads. Primers were then trimmed from the 5' and 3' ends of the merged reads using seqtk (github.com/lh3/seqtk). Trimmed reads were filtered for quality using the USEARCH script **fastq_filter**. This script filters reads using the maximum expected errors per merged read. The number of expected errors is obtained by the sum of the Phred derived error probability. If the expected number of errors is less than one, then the most probable number of errors is zero (Edgar & Flyvbjerg 2015). A maximum expected error threshold of 1 was utilised, resulting in reads with an error probability of 1 or greater being removed. In addition to using the number of expected errors for

filtering, trimmed reads shorter than 400 bp were discarded. Unique reads within the entire sample set were assigned OTUs using the USEARCH algorithms **derep_fulllength** and **cluster_otus** (Edgar 2013), with a minimum identity of 97% for clustering, or a cluster radius of 3.0. Chimeras are filtered from the sample set within the **cluster_otus** command using the UPARSE-REF maximum parsimony algorithm (Edgar 2013). Singletons were excluded from OTU clustering due to the high likelihood that they contain errors (Edgar 2013; Edgar & Flyvbjerg 2015). The merged/trimmed reads from each swab sample, including the previously excluded singletons and merged reads shorter than 400 bp were matched with the clustered OTUs using USEARCH script **usearch_global**, with a threshold of 97% identity to group a read into a specific OTU. The taxonomy of each OTU was determined by using the QIIME script **assign_taxonomy.py** in conjunction with the Greengenes taxonomy database (version 13_5, 97% clustered OTUs) (DeSantis *et al.* 2006). This script utilises the UCLUST algorithm (Edgar 2010) to identify a consensus taxonomy of the reads within an OTU against the curated database, based on a similarity of 90% and a minimum consensus fraction of 0.51. Chloroplast and mitochondrial OTUs were removed from the dataset using the QIIME script **filter_taxa_from_otu_table.py**.

### 6.2.4 Read normalisation and analysis

Read data were assessed using three different methods. Relative abundance was utilised to compare basic phylum presence in each sample. Rarefaction of reads was undertaken, using **multiple_rarefactions.py** QIIME script, to assess alpha and beta diversity at a set read level. Negative-binomial normalisation of reads, using DESeq2 (Love *et al.* 2014) as recommended by McMurdie and Holmes (2014), was performed using the QIIME script **normalize_table.py**. For rarefactions, reads within each sample are subsampled (without replacement) every 5000 reads, from 5000 to 250,000 reads. This represented the maximum number of reads present in the sample with the most reads (rounded down to the nearest value divisible by 5,000). At each step, 100 permutations were undertaken. Alpha-diversity metrics (including species richness, Chao1 (Chao 1984), phylogenetic diversity (Faith 1992), and Shannon's diversity (Shannon 1948)) were generated for each step. Comparisons of these values were undertaken using values obtained after subsampling to a depth of 160,000. This equalled the sample with the fewest reads (rounded down to the nearest value divisible by 5,000). Non-parametric comparisons of mean alpha diversity metrics between the two sample groups (wet

bottom present or absent) were undertaken with the **compare_alpha_diversity.py** QIIME script. This script utilised a non-parametric two sample t-test with 10,000 Monte Carlo permutations to determine whether the alpha diversity was significantly different between the two groups (wet bottom present/absent) at a depth of 160,000 reads. Beta-diversity was assessed at the same depth as above (160,000 reads) using the **beta_diversity_through_plots.py** QIIME script, in which both unweighted and weighted UniFrac distances (Lozupone & Knight 2005) were assessed. Bray-Curtis dissimilarity (Bray & Curtis 1957) between samples was also assessed. The analysis of beta-diversity required a phylogenetic tree. For this, an alignment of representative sequences of each OTU was created with PyNAST (Caporaso *et al.* 2010a) using the **align_seqs.py** QIIME script. A tree was produced from this alignment using FastTree (Price *et al.* 2009), and used as input for beta-diversity analysis. **beta_diversity_through_plots.py** produced distance matrices for each of the tests (UniFrac and Bray-Curtis), from which principal coordinates and eigen values could be calculated. PCoA plots using the 2 or 3 most influential principal coordinates were drawn from the resulting distance matrices using either the **make_2d_plots.py** QIIME script, or within the **beta_diversity_through_plots.py** script using EMPeror 9.51 software (Vázquez-Baeza *et al.* 2013), respectively. Distance and dissimilarity metrics were used to compare the microbial communities between the two groups by utilising the permutational ANOVA (PERMANOVA) method within the **compare_categories.py** QIIME script, with 10,000 permutations. Statistical comparisons of the differential abundance of OTUs between koalas with and without wet bottom utilised DESeq2 within the QIIME script **differential_abundance.py**. These comparisons aimed to determine OTUs which were over-represented in either group. Statistically significant results, from the negative binomial Wald test within DESeq2, were based on $P$-values < 0.05, and were adjusted for false discovery within the script, using the method described by Benjamini and Hochberg (1995).

The NCBI nucleotide database (Clark *et al.* 2016) was utilised to search for species level matches of significantly differentially abundant OTUs. This was conducted using the representative sequence of the significant OTU and the MegaBLAST algorithm (Morgulis *et al.* 2008), excluding uncultured sample sequences.

All reads used in the project are available through the NCBI BioProject ID: PRJNA359726. Illumina reads for each sample are available from the NCBI Short Read archive (Accession numbers SRX2464137 – SRX246146).

## 6.3 Results

### 6.3.1 Clinical status of koalas

Urogenital samples previously collected from ten koalas as a component of population health monitoring were selected from an archive of samples (Legione *et al.* 2016a; Legione *et al.* 2016b). Samples were selected from the APCAH archive for analysis based on their storage in RLT buffer from time of collection, which is appropriate for amplification and sequencing of environmental bacteria. Five koalas with wet bottom were in that cohort, and the swab samples from those individuals were utilised. An additional five samples, taken from koalas with no clinical signs of disease, were selected from the same population. Of the five koalas with wet bottom, the median wet bottom clinical score was 3 (range: 2 – 4). The five clinically healthy animals all had wet bottom clinical scores of 0. All koalas were negative for *Chlamydiaceae* using a pan-*Chlamydiaceae* qPCR.

### 6.3.2 Analysis and processing of sequencing data

A total of 2,295,607 paired reads were obtained across the ten samples, ranging between 189,315 to 312,131 reads per sample. The GC content of the reads was 51.8%. Merging paired reads, trimming 5' and 3' ends, quality filtering to remove errors and discarding merged sequences shorter than 400 bp resulted in a total of 1,347,512 reads. Dereplication resulted in 275,642 unique reads for clustering into OTUs. Through the clustering process, it was determined that 3953 unique reads were chimeric, representing 24,376 filtered reads. The non-chimeric unique reads were clustered into 261 OTUs, 7 of which were either chloroplasts or mitochondria and were subsequently removed from the analysis. In total 1,946,587 reads, from 2,221,529 merged reads (87.6%) were matched to the clustered OTUs. Within samples, this ranged from 162,343 (82% of available reads) to 254,327 (92.1% of available reads) (Table 6.1). For comparison, the same filtering and clustering methodology was run without the removal of singletons, which resulted in the clustering of reads into 592 OTUs.

### 6.3.3 Phylum presence and relative abundance

In total, 13 phyla were detected in the ten samples (Table 6.1), with *Firmicutes* occurring at the highest relative abundance (77.61%). Just over a third of the OTUs were classified as *Firmicutes* (95/254), followed by *Proteobacteria* (59/254) and the *Bacteroidetes* (35/254). When samples were split into the two groups, koalas without wet bottom had 89.3% of reads classified as *Firmicutes*, followed by OTUs which could not be assigned using the 90% similarity threshold (5.2%) and *Actinobacteria* (3.5%). Koalas with wet bottom had 68.2% reads assigned to OTUs classified as *Firmicutes*. The next two most prevalent phyla were *Proteobacteria* (12.5%) and *Bacteroidetes* (12.2%), however these phyla were over-represented in two samples, biasing the total relative values. *Deferribacteres* were detected in only one sample (Koala 70, wet bottom present) and *Acidobacteria* were only detected in two (one clinically normal koala and one displaying wet bottom). *Armatimonadetes* was detected in three koalas without wet bottom, but in none of the five diseased koalas. These three phyla were detected at the lowest relative abundance across the ten samples. Data for relative read abundance for OTUs that could be taxonomically assigned to a genus level and occurred at a percentage of 0.01% or more in either group can be found in Table 6.2. This shows that the order *Lactobacillales*, and within that the genus *Aerococcus*, had the highest proportion of relative reads.

### 6.3.4 Richness and diversity

Species richness within each sample is described in Table 6.1. The mean species richness and Chao1 from 100 iterations of subsampling every 5000 reads is shown in Figure 6.1. After 100 iterations of rarefaction to a depth of 160,000 reads per sample, the mean number of OTUs in the two groups was 80.0 (standard deviation (SD) ± 9.6) and 75.9 (SD ± 24.6) for koalas with wet bottom and without wet bottom, respectively. All alpha diversity metrics compared between samples from koalas with or without wet bottom were not significantly different. This included observed OTUs ($t = -0.31$, $P = 0.81$), Chao1 (with wet bottom group (WB) mean = 90.7, without wet bottom group (NWB) mean = 88.4, $t = -0.20$, $P = 0.83$), phylogenetic diversity (WB mean = 7.8, NWB mean = 8.1, $t = -0.39$, $P = 0.71$) and Shannon's diversity (WB mean = 2.4, NWB mean = 2.5, $t = -0.15$, $P = 0.86$) (see Table 6.3 for individual alpha diversity values and standard deviations). Results detailing abundance for all OTUs detected in koala urogenital samples are recorded in Appendix 6.

Fewer than half of the OTUs detected across the two sample groups were shared between them (112/254) (Figure 6.2). At a read depth of 160,000 there was a significant difference between the microbial communities in koalas with wet bottom compared to those without, based on the results of a 10,000 permutation PERMANOVA using Bray-Curtis dissimilarity (F = 4.92, $P = 0.019$) and unweighted (qualitative) UniFrac distances (F = 1.62, $P = 0.031$). There was no significant difference detected when using weighted (quantitative) UniFrac distances (F = 1.51, $P = 0.061$). 2D and 3D principal coordinate analysis (PCoA) graphs comparing koalas with and without wet bottom are shown in Figure 6.3. These identify two outliers in the wet bottom present group, koalas 49 and 70.

### 6.3.5 Comparisons between samples using DESeq2 normalised reads

Negative binomial normalisation of reads from each sample using DESeq2 still resulted in *Firmicutes* as the most dominant phylum across all samples. This was followed by *Proteobacteria* and *Bacteroidetes* (Figure 6.4). Overall there were 25 OTUs with significant (Benjamini and Hochberg (1995) adjusted $P < 0.05$) over-representation or under-representation in wet bottom affected koalas, in comparison to clinically normal koalas, based on these normalised read counts (Table 6.4). Of those OTUs, when assessing absolute read count, six occurred only in koalas with wet bottom, whilst eight occurred only in koalas without wet bottom (Table 6.4). Normalised read values for all OTUs can be found in Appendix 7, and statistical comparisons of normalised reads for all OTUs in relation to wet bottom presence or absence are in Appendix 8.

## 6.4 Discussion

Previous assessment of the koala microbiome has focused on the digestive system of koalas comparing either two free ranging animals from northern populations (Barker *et al.* 2013) or two captive koalas in Europe (Alfano *et al.* 2015), from which the ocular microbiome was also assessed. This study is the first investigation of the microbiome of the urogenital tract of the female koala using modern high-throughput techniques, and only the second to assess the urogenital tract of a marsupial, with the tammar wallaby (*Macropus eugenii*) investigated previously using terminal restriction fragment length polymorphism analysis (Chhour *et al.* 2008). The majority of reads in the sample set were classified in the order *Lactobacillales* (72.1%). This dominance of *Firmicutes* mirrors what has been seen in the human vaginal microbiome (Zhou *et al.* 2007). In

humans, the acidic pH of the genital tract is maintained by these lactic acid producing bacteria, which in turn is thought to play a role in preventing pathogenic infection (Boskey *et al.* 1999). It appears from the sample set that koalas have a different family within the *Lactobacillales*, possibly performing a similar role. The most common family within the classified OTUs, in terms of either relative or normalised read abundance, was *Aerococcaceae*, whilst in humans the *Lactobacilli* dominate the reproductive tract. Within the *Aerococcaceae*, the genera *Aerococcus* and *Facklamia* were both represented in the top four most abundant OTUs. For all four significantly differentially abundant *Aerococcus* spp. OTUs, the same OTU could be detected in at least 4/5 (80%) of the converse sample group in absolute reads. For example, OTU 4, an *Aerococcus* sp. occurred in all ten koala samples, but was present in significantly higher quantities in clinically normal koalas after normalisation ($P = 0.004$). Whether specific *Aerococcus* spp. that are over or under-represented are an important factor in terms of disease presence requires further investigation. The production of hydrogen peroxide by commensal *Lactobacillus* species is thought to play a role in reducing the successful establishment of sexually transmitted diseases in humans (Klebanoff & Coombs 1991; Martin *et al.* 1999a), and it has been shown that *Aerococcus* spp. are involved in hydrogen peroxide production (Kontchou & Blondeau 1990; Streitenberger *et al.* 2001). In humans *Aerococcus* spp. have also been associated with disease, such as *Aerococcus urinae*, which can cause urinary tract infections (Zhang *et al.* 2000) and septicaemia (de Jong *et al.* 2010). Investigations into the urinary microbiome of women with and without 'urgency urinary incontinence' found that *Aerococcus* spp. were detected more frequently in cases where disease was present (Pearce *et al.* 2014). In this study, the four *Aerococcus* spp. OTUs that had significantly different normalised abundance were evenly split, with two having higher abundance in koalas with wet bottom and two in koalas without wet bottom. The role of organisms within this family as opportunistic pathogens in koalas cannot be ruled out and further work is required to more fully characterise these bacteria.

The *Aerococcus* were the most common genus amongst those OTUs with significant differential abundance after normalisation using DESeq2. The representative sequences of these four OTUs did not match known species within the *Aerococcus* genus, using the Greengenes database, with an identity greater than 90%, suggesting that these may represent novel species. This is not unexpected, as the culture of organisms from the

koala urogenital tract has been limited to only a small number of studies, with the majority focused on diagnosing what was later deemed to be chlamydial infection (McKenzie 1981; Obendorf 1983; Brown & Grice 1984). Efforts in culturing novel bacteria from koalas have focused primarily on its gut microbiome (Osawa *et al.* 1993b), of interest due to the koala's unusual diet of *Eucalyptus* leaves, as well as the microbial flora in the pouch (Osawa *et al.* 1992). Now that some organisms of interest in the female koala urogenital microbiome have been identified (to the genus level), it would be beneficial to use traditional microbiology techniques to further study these organisms. The other family of interest are the *Tissierellaceae*, within the order *Clostridiales*. The four *Tissierellaceae* OTUs with a significant differential abundance, all occurred in higher normalised quantities in koalas with wet bottom present. Three of these OTUs were in the genus *Peptoniphilus.* Interestingly, only one of these four OTUs was detected at all in the group of koalas without wet bottom, and only from the reads of one koala within this group. The *Peptoniphilus*, previously part of the genus *Peptostreptococcus* (Ezaki *et al.* 2001) within the family *Peptostreptococcaceae* (also in the order *Clostridiales*), have been associated with inflammatory diseases in other species. This includes mastitis in cattle (Madsen *et al.* 1990) and pelvic inflammatory disease in humans (Cunningham *et al.* 1978). Organisms in this genus are obligate anaerobes (Ezaki *et al.* 2001) and therefore potentially overlooked in culture based methods of investigating urogenital tract pathogens.

The average number of OTUs detected in the samples used here is difficult to compare to other publications investigating koala microbiomes. This is both due to the impact that sample site differences would have on OTUs present, as well as the method of OTU classification used. For instance, previous research on the koala intestinal microbiome used QIIME for analysis of 454 pyrosequencing reads (Barker *et al.* 2013) and detected 1855 OTUs, after removal of chimeras and singletons, from caecum, colon, and faecal samples. Similarly, an Illumina based study of microbiomes from ocular, oral, rectal and faecal samples from two captive koalas found OTU numbers ranging between 597 to 3,592, with a median of 1,456 (Alfano *et al.* 2015). The average raw read numbers per sample assessed in these projects ranged from 12,831 (454 pyrosequencing) to 323,030 (Illumina). The average raw reads per sample in this study were within that range (229,561), suggesting that the OTU differences between the studies are either associated with the sample site (urogenital versus digestive tract) or clustering

methodology used. UPARSE was used in this study due to its demonstrated ability to correctly identify OTUs in a mock community and minimise spurious OTUs (Edgar 2013). Whilst there did not appear to be any strong clustering on the 2D or 3D PCoA plots, comparisons of the beta-diversity between groups highlighted that the makeup of the communities was significantly different when assessing both Bray-Curtis dissimilarity and unweighted UniFrac distances. These metrics assess presence/absence of OTUs between groups, with UniFrac also considering phylogenetic distance between OTUs present. Weighted UniFrac distances, which considers the abundance of individual OTUs, were not significantly different between groups. Therefore, koalas with and without wet bottom appear to have a significant difference in which OTUs are present in the samples, but not necessarily the abundance of OTUs between samples. Two samples had widely different OTU profiles (koala 49 and 70). This finding may support the hypothesis that wet bottom in female koalas without *C. pecorum* may be caused by more than one aetiological agent (Patterson *et al.* 2015; Legione *et al.* 2017). Further investigations to examine this hypothesis are indicated but require access to a large number of appropriately collected and stored samples. Such sample sets are currently not available for this species.

It could be argued that the skewed relative abundance of *Proteobacteria* and *Bacteroidetes* in the samples from koala 49 and 70, respectively, could be a result of swab contamination with faecal material, which would impact diversity inferences. The human microbiome project identified that reads from stool samples were predominately from the *Bacteroidetes* phylum (Human Microbiome Project Consortium 2012), and the most recent assessment of the koala rectal microbiome found these two phyla to be the most abundant in samples taken from both koalas assessed (Alfano *et al.* 2015). In koalas, the urogenital tract is accessed through the cloaca, which also contains the rectal opening. This makes faecal contamination difficult to avoid during sample collection. Future studies of the urogenital tract microbiome would benefit from either taking control samples from the rectum of the koala being sampled, or inverting the cloaca so that the urogenital opening is more easily accessible, as described previously for the tammar wallaby (Chhour *et al.* 2008). In that study, approximately a quarter of phylotypes (26/96) were detected in both the urogenital and rectal samples, suggesting that bacteria being detected at multiple sites in marsupials is not unusual.

The sample size utilised in this study is larger than previous studies of koala microbiomes, which have incorporated at most two individuals, yet it is substantially smaller than many studies in human medicine which include hundreds of samples (Ravel *et al.* 2011). The samples utilised were opportunistically collected during population management exercises, and chosen from the available sample archive due to the absence of *C. pecorum* from the French Island koala population at the time of testing (Martin & Handasyde 1999). Whilst *C. pecorum* was subsequently determined to be present in this population (Legione *et al.* 2016a), no koalas used in this project were positive via a *Chlamydiaceae* PCR. Importantly, no koalas used in this study were found to have reads classified within the *Chlamydiae* phylum after taxonomic assignment of OTUs, which supports the use of the 16SG PCR as a sensitive screening technique to detect *Chlamydiaceae* in clinical samples.

## 6.5 Concluding remarks

Disturbance of the normal vaginal flora in humans, such as in cases of bacterial vaginosis, is a risk factor associated with infection by retroviruses (such as human immunodeficiency virus (HIV)) and *C. trachomatis* (Wiesenfeld *et al.* 2003). This study provides useful data as to what bacteria could be expected in the urogenital tract of a clinically normal koala. This will allow for broader, more detailed studies on the impact that infection with *C. pecorum* has on the koala urogenital microbiome, and vice versa. The prevalence of koala retrovirus was unknown in theses samples, as no blood was collected from individuals, and therefore its impact on the urogenital microbiome could not be determined. The validity of examining swab samples for KoRV prevalence has not be established, and is unlikely to be reliable in southern koalas were KoRV has not entered the germline. Future studies would benefit from the knowledge of KoRV status in individuals, as well as a greater sample size and a more diverse array of sampled regions both within a single state, and across the country. Further research covering a wider diversity of kingdoms, including viral and fungal diseases, would be valuable. However, the required enrichment and sequencing methodology to undertake this work is significantly more costly than the methods used here, which may represent a barrier to its undertaking. It would be interesting to follow the same individuals over time to determine if mating and breeding impact the microbiome of the urogenital tract, as occurs in humans (Aagaard *et al.* 2012). However, animal welfare issues regarding recapturing wild koalas multiple times may make this unfeasible. Additionally, as this

study focused solely on female koalas, a follow up survey of the microbiome of the male urogenital tract would be enlightening. Finally, targeted studies assessing the prevalence of organisms associated with wet bottom would increase our understanding of organisms potentially impacting koala populations and could in turn assist with conservation of this iconic species.

**Table 6.1.** Koala wet bottom score, read metrics and relative abundance data from ten samples submitted for 16S rRNA amplicon sequencing. All koalas were female and sampled from French Island, Victoria, Australia in 2011.

| Koala/Sample name | K1 | K2 | K3 | K4 | K5 | K31 | K49 | K55 | K59 | K70 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Wet bottom score**[#] | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 3 | 4 | 3 |
| **Merged reads** | 253256 | 211620 | 186912 | 220410 | 185592 | 183126 | 199985 | 263685 | 216495 | 300448 |
| **Reads after filtering** | 156100 | 134940 | 118418 | 132125 | 112823 | 110292 | 116321 | 160328 | 136996 | 169169 |
| **Reads clustered to OTUs** | 225868 | 178678 | 169576 | 203062 | 166906 | 162343 | 177452 | 216270 | 192105 | 254327 |
| **Absolute OTUs** | 93 | 66 | 86 | 89 | 74 | 55 | 61 | 74 | 76 | 126 |
| **Standardised OTUs**[^] **± SD** | 88.8 ± 1.7 | 64.1 ± 1.2 | 85.4 ± 0.7 | 88 ± 0.9 | 73.7 ± 0.6 | 54.9 ± 0.3 | 59.2 ± 1.4 | 69.2 ± 1.9 | 72.9 ± 1.5 | 123.4 ± 1.3 |
| **Phyla**[&] | | | | | | | | | | |
| ***Acidobacteria*** | - | - | - | - | < 0.01% | - | - | - | - | 0.01% |
| ***Actinobacteria*** | 5.47% | 9.06% | 2.92% | 0.17% | 0.03% | 3.27% | 0.66% | 1.50% | 0.30% | 0.19% |
| ***Armatimonadetes*** | < 0.01% | < 0.01% | - | - | < 0.01% | - | - | - | - | - |
| ***Bacteroidetes*** | 0.57% | 0.05% | 2.14% | 1.72% | 0.21% | 0.33% | 0.05% | 9.05% | 1.00% | 50.53% |
| ***Cyanobacteria*** | < 0.01% | - | < 0.01% | - | - | - | - | - | - | 0.02% |
| ***Deferribacteres*** | - | - | - | - | - | - | - | - | - | < 0.01% |
| ***Firmicutes*** | 92.92% | 89.57% | 85.67% | 79.17% | 98.92% | 80.35% | 40.92% | 84.88% | 95.65% | 39.09% |
| ***Fusobacteria*** | 0.02% | < 0.01% | < 0.01% | 0.07% | < 0.01% | < 0.01% | - | < 0.01% | 0.02% | 1.09% |
| ***Planctomycetes*** | - | - | < 0.01% | - | 0.01% | - | - | - | < 0.01% | 0.80% |
| ***Proteobacteria*** | 0.24% | 0.15% | 1.66% | 1.51% | 0.45% | 0.23% | 56.90% | 0.19% | 2.37% | 2.70% |
| ***Synergistetes*** | 0.08% | 0.02% | 0.30% | 0.31% | 0.01% | - | - | < 0.01% | 0.02% | 4.35% |
| **TM7** | 0.02% | 0.50% | 0.21% | - | < 0.01% | 1.38% | 0.05% | 2.86% | < 0.01% | 0.02% |
| ***Verrucomicrobia*** | < 0.01% | < 0.01% | < 0.01% | - | 0.02% | < 0.01% | - | - | 0.01% | 0.69% |
| **Unassigned** | 0.69% | 0.65% | 7.07% | 17.04% | 0.34% | 14.44% | 1.42% | 1.52% | 0.61% | 0.52% |

[#] Wet bottom score ranges from 0 (absent) to 10 (most severe) (Griffith 2010)

[^] The average number of OTUs detected in 100 iterations of subsampling to a depth of 160,000 reads

[&] Phyla assigned using QIIME (Caporaso *et al.* 2010b) script **assign_taxonomy.py** utilising Greengenes (DeSantis *et al.* 2006) curated 16S rRNA library

**Table 6.2.** Relative abundance of OTUs with taxonomic classification to a genus level, in female koalas with and without wet bottom. Only OTUs with relative abundance greater than 0.01% in at least one group are shown.

| Phylum | Class | Order | Family | Genus | OTUs | WB absent | WB present | Combined |
|---|---|---|---|---|---|---|---|---|
| *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Actinomycetaceae* | *Mobiluncus* | 1 | Nil[^] | 0.05% | 0.03% |
| | | | *Corynebacteriaceae* | *Corynebacterium* | 6 | 0.68% | 0.60% | 0.64% |
| *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | 14 | 0.03% | 0.54% | 0.29% |
| | | | *Porphyromonadaceae* | *Dysgonomonas* | 1 | <0.01%[+] | 0.18% | 0.09% |
| | | | | *Parabacteroides* | 7 | 0.89% | 9.55% | 5.22% |
| | | | | *Porphyromonas* | 2 | <0.01% | 1.88% | 0.94% |
| | | | *Prevotellaceae* | *Prevotella* | 2 | <0.01% | 0.02% | 0.01% |
| *Firmicutes* | *Bacilli* | *Bacillales* | *Staphylococcaceae* | *Staphylococcus* | 1 | 0.02% | <0.01% | 0.01% |
| | | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | 6 | 77.45% | 54.74% | 66.10% |
| | | | *Aerococcaceae* | *Facklamia* | 1 | 6.55% | 5.43% | 5.99% |
| | | | *Carnobacteriaceae* | *Trichococcus* | 1 | 0.02% | 0.05% | 0.04% |
| | | | *Streptococcaceae* | *Streptococcus* | 2 | 0.03% | <0.01% | 0.02% |
| | *Clostridia* | *Clostridiales* | *Tissierellaceae* | *Gallicola* | 1 | <0.01% | 0.27% | 0.14% |
| | | | | *Peptoniphilus* | 4 | <0.01% | 0.53% | 0.27% |
| | | | | *ph2* | 3 | Nil | 0.10% | 0.05% |
| | | | *Clostridiaceae* | *Clostridium* | 8 | 4.48% | 1.87% | 3.18% |
| | | | *Peptococcaceae* | *Peptococcus* | 1 | Nil | 0.23% | 0.11% |
| | | | *Ruminococcaceae* | *Ruminococcus* | 2 | 0.07% | 0.10% | 0.08% |
| | | | *Veillonellaceae* | *Dialister* | 1 | Nil | 0.04% | 0.02% |
| | | | | *Phascolarctobacterium* | 1 | 0.04% | 1.03% | 0.54% |
| *Fusobacteria* | *Fusobacteriia* | *Fusobacteriales* | *Fusobacteriaceae* | *Fusobacterium* | 2 | 0.02% | 0.22% | 0.12% |
| *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylobacteriaceae* | *Methylobacterium* | 2 | 0.31% | 0.06% | 0.19% |
| | *Betaproteobacteria* | *Burkholderiales* | *Alcaligenaceae* | *Sutterella* | 1 | <0.01% | 0.05% | 0.02% |
| | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | *Desulfovibrio* | 2 | 0.06% | 0.12% | 0.09% |
| | *Gammaproteobacteria* | *Pasteurellales* | *Pasteurellaceae* | *Lonepinella* | 1 | 0.06% | 0.25% | 0.15% |
| | | *Pseudomonadales* | *Moraxellaceae* | *Acinetobacter* | 4 | 0.01% | 0.02% | 0.01% |

| Phylum | Class | Order | Family | Genus | OTUs | WB absent | WB present | Combined |
|---|---|---|---|---|---|---|---|---|
| | | | *Pseudomonadaceae* | *Pseudomonas* | 2 | 0.01% | <0.01% | 0.01% |
| *Synergistetes* | *Synergistia* | *Synergistales* | *Synergistaceae* | *vadinCA02* | 1 | Nil | 0.04% | 0.02% |
| *Verrucomicrobia* | *Verrucomicrobiae* | *Verrucomicrobiales* | *Verrucomicrobiaceae* | *Akkermansia* | 1 | <0.01% | 0.14% | 0.07% |

^ No reads clustering with OTUs that were assigned this genus were present in any of the 5 koalas within this group

+ Less than 0.01% of reads were clustered to OTUs within this genus, but are included in this table due to the converse group having greater than 0.01% of reads clustered to OTUs within this genus.

**Table 6.3.** Alpha diversity metrics for microbial communities in the urogenital tract of female koalas with and without wet bottom. All metrics assessed based on OTU values after subsampling to a depth of 160,000 reads, with 100 permutations. *P* values are non-parametric t-tests using 10,000 Monte Carlo permutations.

| | Richness (OTUs) | Shannon's diversity | Chao1 | Phylogenetic diversity |
|---|---|---|---|---|
| **Wet bottom absent** | | | | |
| Koala 1 | 88.8 (± 1.7) [#] | 2.6 (± <0.01) | 97.1 (± 5.9) | 9.1 (± 0.2) |
| Koala 2 | 64.1 (± 1.2) | 2.7 (± <0.01) | 84.9 (± 7.4) | 7.0 (± 0.1) |
| Koala 3 | 85.4 (± 0.7) | 3.0 (±<0.01) | 91.5 (± 2.7) | 8.9 (± 0.1) |
| Koala 4 | 88 (± 0.9) | 3.1 (± <0.01) | 92.5 (± 3.7) | 7.7 (± 0.1) |
| Koala 5 | 73.7 (± 0.6) | 1.1 (± <0.01) | 87.6 (± 4.9) | 7.9 (± 0.1) |
| Mean | 80.0 (± 9.6) | 2.5 (± 0.7) | 90.7 (± 4.2) | 8.1 (± 0.8) |
| **Wet bottom present** | | | | |
| Koala 31 | 54.9 (± 0.3) | 2.4 (± <0.01) | 58.7 (± 0.8) | 6.5 (± 0.0) |
| Koala 49 | 59.2 (± 1.4) | 1.4 (± <0.01) | 76.4 (± 7.2) | 6.5 (± 0.2) |
| Koala 55 | 69.2 (± 1.9) | 2.3 (± <0.01) | 91.5 (± 13.5) | 7.8 (± 0.2) |
| Koala 59 | 72.9 (± 1.5) | 1.8 (± <0.01) | 87.4 (± 7.1) | 7.8 (± 0.1) |
| Koala 70 | 123.4 (± 1.3) | 4.1 (± <0.01) | 127.9 (± 5.9) | 10.4 (± 0.1) |
| Mean | 75.9 (± 24.6) | 2.4 (± 0.9) | 88.4 (± 22.8) | 7.8 (± 1.4) |
| ***t* stat** | -0.31 | -0.15 | -0.20 | -0.39 |
| ***P* value** | 0.81 | 0.86 | 0.83 | 0.71 |

[#] All ± values are standard deviation from the mean

**Table 6.4.** Significant operational taxonomic units (OTU) assessed using DESeq2 (Love *et al.* 2014), ordered from lowest to highest adjusted *P* value. Representative sequences were compared to NCBI nucleotide database using MegaBLAST (Morgulis *et al.* 2008), excluding 'uncultured organisms'

| OTU ID | Adjusted *P* value * | Higher abundance group # | OTU present in samples/n | | NCBI MegaBLAST best hit^ | | |
|---|---|---|---|---|---|---|---|
| | | | WB absent | WB present | Organism | Nucleotide Identity (%) | Accession number |
| 38 | < 0.001 | WB present | 0/5 | 5/5 | *Peptoniphilus indolicus* | 96.8 | NR_117566 |
| 21 | < 0.001 | WB present | 1/5 | 5/5 | *Peptoniphilus asaccharolyticus* | 100 | KP944181 |
| 47 | < 0.001 | WB present | 0/5 | 3/5 | *Levyella massiliensis* | 100 | NR_133039 |
| 51 | < 0.001 | WB present | 0/5 | 3/5 | *Peptoniphilus lacrimalis* | 100 | KM624632 |
| 65 | 0.001 | WB present | 1/5 | 2/5 | *Sutterellaceae bacterium* | 99.5 | LK054638 |
| 86 | 0.003 | WB absent | 3/5 | 0/5 | *Bacteroides thetaiotaomicron* | 100 | KU234409 |
| 75 | 0.004 | WB absent | 2/5 | 0/5 | *Clostridium* sp. | 96.5 | AB622820 |
| 4 | 0.004 | WB absent | 5/5 | 5/5 | *Lactobacillales bacterium* | 92.8 | HQ115584 |
| 70 | 0.005 | WB absent | 2/5 | 0/5 | *Clostridium neopropionicum* | 94.6 | JQ897394 |
| 73 | 0.005 | WB present | 0/5 | 2/5 | *Alistipes onderdonkii* | 93.6 | NR_113151 |
| 69 | 0.005 | WB absent | 2/5 | 0/5 | *Lachnospiraceae bacterium* | 95.3 | EU728729 |
| 2 | 0.006 | WB absent | 5/5 | 5/5 | *Trichococcus* sp. | 94.2 | KU533824 |
| 94 | 0.007 | WB absent | 2/5 | 1/5 | *Rhizobiales* sp. | 100 | KJ016001 |
| 95 | 0.013 | WB absent | 2/5 | 0/5 | *Rhizobium leguminosarum* | 100 | KX346599 |
| 103 | 0.019 | WB absent | 2/5 | 0/5 | *Piscinibacter aquaticus* | 88.6 | NR_114061 |
| 106 | 0.019 | WB absent | 3/5 | 0/5 | *Burkholderia cenocepacia* | 100 | KU749979 |
| 109 | 0.019 | WB present | 0/5 | 2/5 | *Peptostreptococcus anaerobius* | 94.1 | NR_042847 |
| 148 | 0.019 | WB present | 0/5 | 2/5 | *Trichococcus* sp. | 87.5 | KU533824 |
| 159 | 0.019 | WB present | 2/5 | 4/5 | *Abiotrophia defectiva* | 87.9 | JF803600 |
| 114 | 0.019 | WB absent | 2/5 | 1/5 | *Massilia* sp. | 99.8 | JF279920 |
| 113 | 0.019 | WB absent | 3/5 | 0/5 | *Agrobacterium tumefaciens* | 100 | KU955329 |
| 1 | 0.030 | WB present | 5/5 | 5/5 | *Aerococcus viridans* | 95.1 | KC699123 |
| 105 | 0.035 | WB present | 4/5 | 5/5 | *Aerococcus sanguinicola* | 93.0 | LC145565 |

| OTU ID | Adjusted *P* value * | Higher abundance group # | OTU present in samples/n | | NCBI MegaBLAST best hit^ | | |
|---|---|---|---|---|---|---|---|
| | | | WB absent | WB present | Organism | Nucleotide Identity (%) | Accession number |
| 250 | 0.038 | WB present | 1/5 | 2/5 | *Hippea* sp. | 79.5 | FR754504 |
| 90 | 0.038 | WB present | 1/5 | 2/5 | *Olsenella scatoligenes* | 97.8 | NR_134781 |

* *P* value are from negative binomial Wald test, adjusted using the false discovery rate calculation described by Benjamini and Hochberg (1995)

\# OTU was detected with significantly higher normalised read counts in koalas with (WB present) or without (WB absent) wet bottom

^ Organism with the lowest e-value detected using a MegaBLAST search of the NCBI nucleotide database, the nucleotide identity compared to the representative sequence, and the accession number of the hit

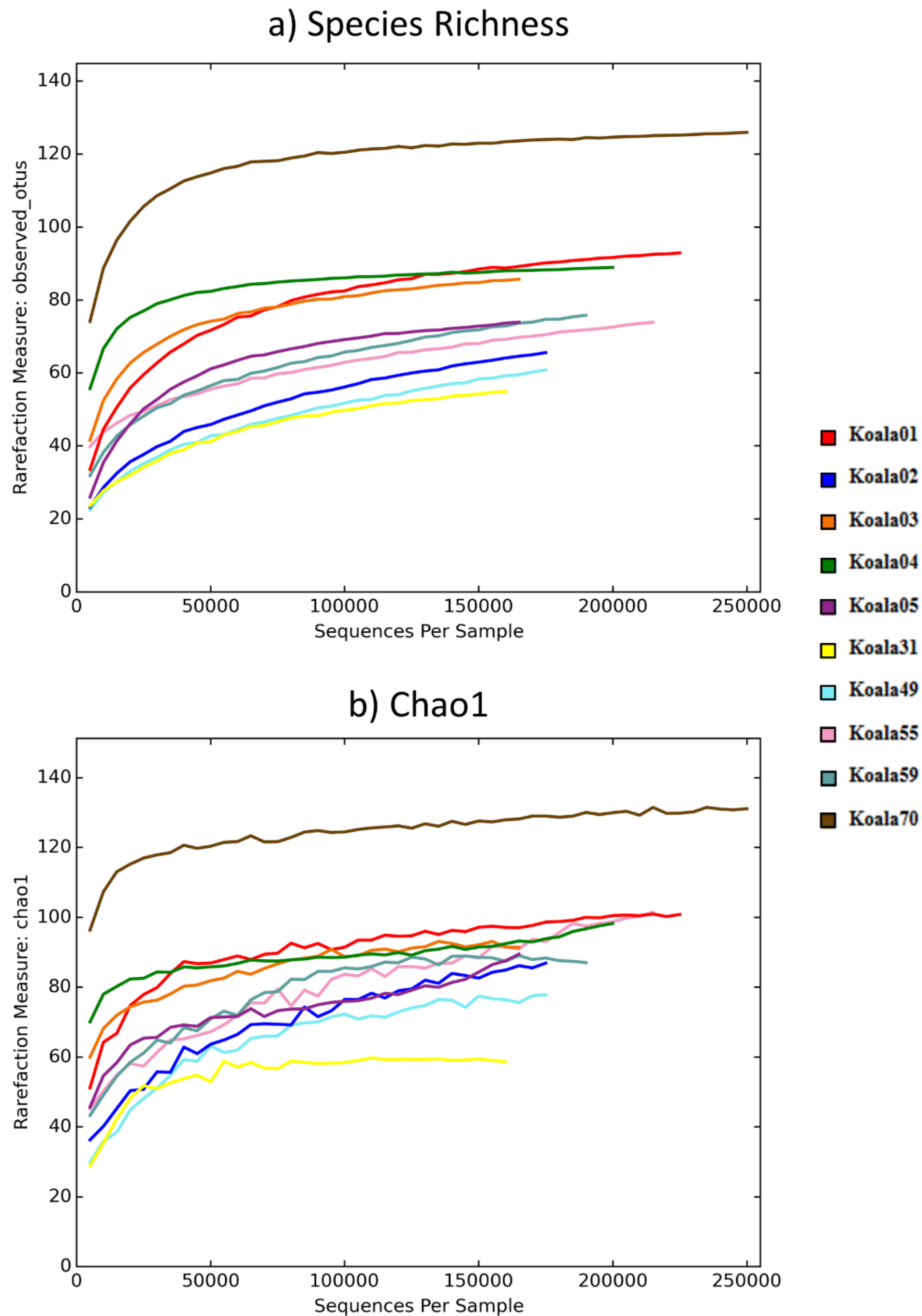



**Figure 6.1.** Rarefaction plots showing **a)** species richness (OTU abundance) and **b)** Chao1. OTUs were subsampled every 5000 reads, with 100 iterations, with the mean result of these iterations forming the plots. Koalas 1 – 5 were clinically normal (wet bottom absent), whilst koalas 31 – 70 had wet bottom.

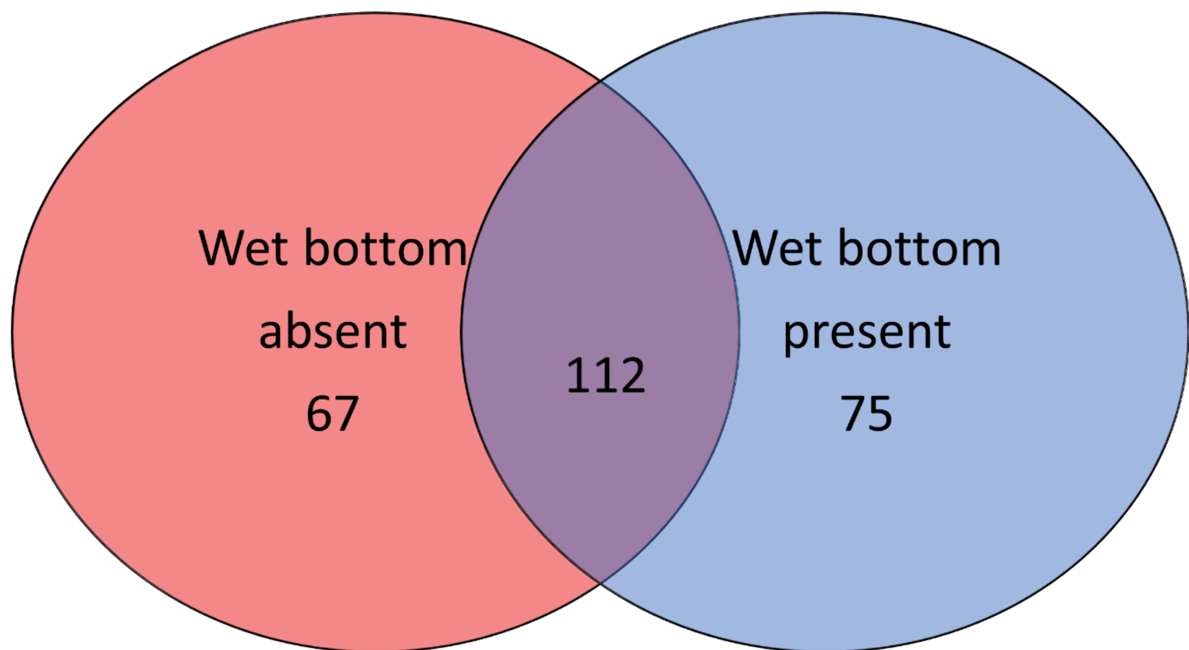



**Figure 6.2.** Venn diagram of the total operational taxonomic units (OTUs) detected in koalas with or without wet bottom. Overlap does not scale with OTU number.

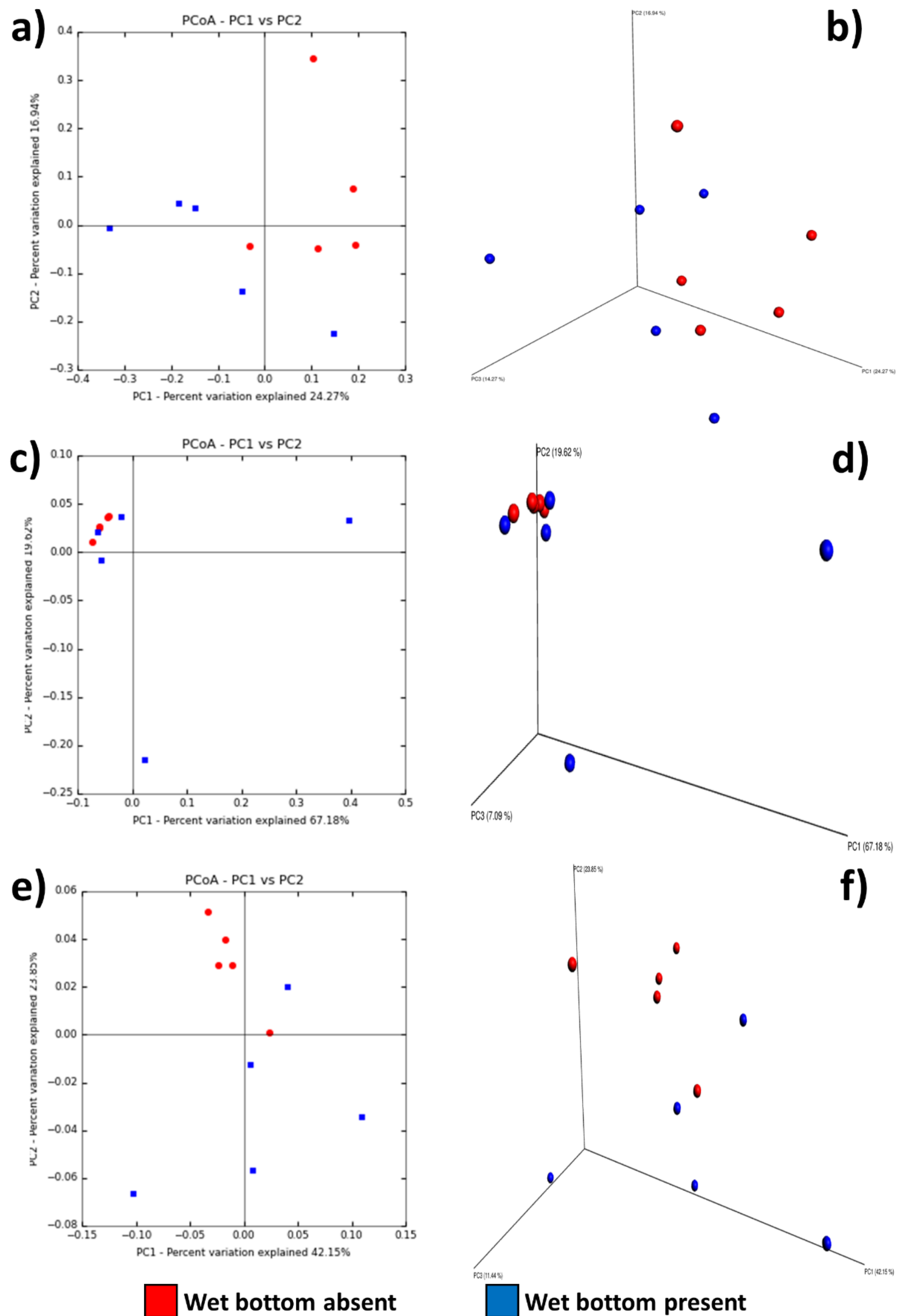



**Figure 6.3.** 2D and 3D PCoA plots of koala samples, with and without wet bottom, using **a/b)** unweighted UniFrac distances of OTUs at a depth of 160,000 reads, **c/d)** weighted UniFrac distances of OTUs at a depth of 160,000, **e/f)** weighted UniFrac distances of normalised reads

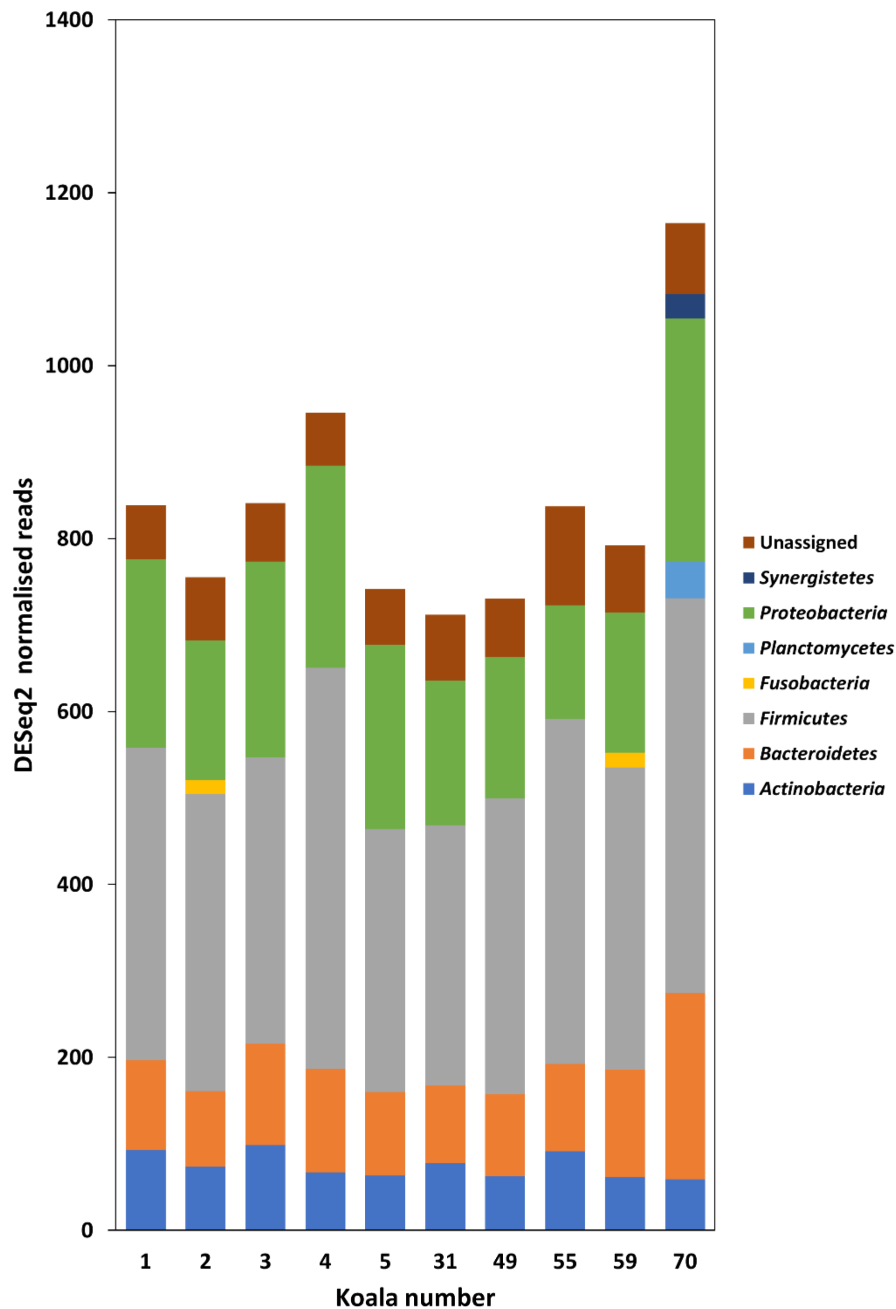



**Figure 6.4.** DESeq2 normalised read counts of phyla detected in koala urogenital swab samples. Phyla with fewer than 2% relative reads within each sample have been excluded for clarity. Reads were characterised into taxanomic groups using QIIME (Caporaso *et al.* 2010b), utilising Greengenes (DeSantis *et al.* 2006) as a reference database. Koalas 1 – 5 were clinically normal (wet bottom absent), whilst koalas 31 – 70 had wet bottom.

# 7. Genomic comparisons of *Chlamydia pecorum* infecting koalas across Australia

## 7.1 Introduction

In northern and southern koalas, the clinical signs of disease associated with *C. pecorum* are variable. Victorian koalas rarely suffer from ocular infection (Patterson *et al.* 2015; Legione *et al.* 2016a), which has been identified in both northern populations and (Polkinghorne *et al.* 2013) in South Australian koalas (Speight *et al.* 2016). Furthermore, *C. pecorum ompA* genotypes vary considerably across the different states of Australia (Kollipara *et al.* 2013b; Legione *et al.* 2016b). However, single gene analysis is not considered robust for elucidating true differences between genomes, and may not reflect accurate evolutionary history (Marsh *et al.* 2011). Additionally, recombination across the genome cannot be identified through single gene comparisons, as demonstrated previously for the human pathogen *C. trachomatis* (Harris *et al.* 2012)

Robust phylogenetic analysis of *C. pecorum* utilising MLST, which assesses seven housekeeping genes, has been used to investigate infection in livestock and its association with clinical disease (Jelocnik *et al.* 2013). Sheep presenting with particular clinical signs of disease, such as polyarthritis or conjunctivitis, were more likely to be infected with specific STs of *C. pecorum*. The presence or absence of the *C. pecorum* plasmid may also play a role in virulence (Jelocnik *et al.* 2016), as in *C. trachomatis* a similar plasmid is a transcriptional regulator. (Carlson *et al.* 2008). Individual genes have also been identified as potential virulence factors, with research in livestock identifying repeat regions in the genes encoding inclusion membrane protein A (*incA*) and the hypothetical protein ORF663 as key targets. Both a decrease in the number of clustered tandem repeats within ORF663 and certain 3 – 5 amino acid motifs in *incA* correlate with increased pathogenesis in livestock (Yousef Mohamad *et al.* 2014). However, it is not known if these factors play a role in disease in koalas.

Full genome sequencing can help to solve questions regarding pathogenicity of prokaryotic organisms, and several *C. pecorum* isolates from livestock have been sequenced, including strains from cattle (Mojica *et al.* 2011; Sait *et al.* 2014), sheep (Bachmann *et al.* 2014; Sait *et al.* 2014) and pigs (Jelocnik *et al.* 2015). Standard isolate shotgun sequencing has also been used to obtain near complete genomes of *C. pecorum*

from koalas (Bachmann *et al.* 2014). However, *C. pecorum* is an obligate intracellular parasite, which makes isolation of the organism difficult, as either cell culture or chicken embryo yolk sac inoculation is required (Sait *et al.* 2014). Recent advances in probe based sequencing technology have rapidly advanced our understanding of *C. pecorum* phylogenetics, enabling sequencing of entire genomes from single swabs and bypassing the tedious cell culture process (Bachmann *et al.* 2015). This methodology has both increased our basic understanding of the *C. pecorum* genome in different species, whilst having the benefit of avoiding the introduction of mutations that may arise through multiple passages *in vitro*. Initial analysis comparing the genomes of *C. pecorum* from koalas with those detected in livestock utilised two koala samples, one from South Australia and another from New South Wales (Bachmann *et al.* 2015). Whilst this provided new and useful information regarding *C. pecorum* genomes, a broader investigation comparing *C. pecorum* from a wide range of koala populations may provide new insights into this debilitating infection.

This chapter utilises the method described by Bachmann *et al.* (2015) to investigate genomic differences between *C. pecorum* by sequencing and analysing complete or near complete genomes of *C. pecorum* detected in koalas that had ocular and/or urogenital infections from four Australian states, namely Queensland, New South Wales, South Australia and Victoria.

## 7.2 Methodology

### 7.2.1 Swab sample collection and selection

Samples were collected from Victorian koalas as previously described (Section 2.1.2). These samples are referred throughout as APCAH samples. In total, 41 urogenital (UGT) swabs from Victorian koalas that were identified as *C. pecorum*-positive using a 16S rRNA specific PCR (chapters 3 and 4) were selected for full genome sequencing (Table 7.1). Samples were selected based on *C. pecorum* genome copy numbers, the availability of clinical data and, in an effort to increase the diversity of the target sequences, the geographical origin of the sample and *ompA* genotype were also considered. Samples with higher concentrations of genome copies were predicted to have a more successful sequencing outcome. An additional 69 samples (53 UGT and 16 ocular swabs), some of which were duplicates, were selected from the koala *C. pecorum* sample archive at the University of the Sunshine Coast (USC) (Appendix 11.), based on

*C. pecorum* genome copy numbers and geographical origin of the sample. A summary of the samples available for sequence analysis is available in Table 7.2.

### 7.2.2 DNA extraction and target enrichment

DNA was extracted from the APCAH samples, as previously described (section 2.3.1). The DNA concentration of extracted samples was measured using a Nanodrop 1000 spectrophotometer and *C. pecorum* presence confirmed using qPCR as previously described (section 2.3.2). Genomic copy numbers of *C. pecorum* were calculated by utilising a standard curve in the qPCR as previously described (section 2.3.2). Aliquots (75 µL) of extracted DNA from each sample were transported to the USC, on dry ice, and then further transported to the Sanger Institute, UK. DNA samples then underwent library preparation, target DNA enrichment and sequencing.

Target enrichment and sequencing followed methods previously described (Bachmann *et al.* 2015). In brief, target enrichment utilised SureSelect (Agilent) DNA hybridisation kits. These were used to bind sheared genomic DNA to SureSelect biotinylated RNA library baits. These baits were 120-mer probes, designed using the E58 reference genome of *C. pecorum* (accession number: NC_015408), as previously described (Bachmann *et al.* 2015). The DNA/probe mix was then bound to streptavidin coated magnetic beads, which were captured through a magnetic column. Beads and digested RNA were then removed from the captured DNA, after which it underwent a standard amplification and sequencing protocol. Paired end read sequencing was undertaken by the Sanger Institute, UK, using the Illumina HiSeq 2000 platform with 2 × 75 bp reads in most cases. The Illumina MiSeq platform (2 × 150 bp) and Illumina HiSeq 2500 platform (2 × 125 bp) were utilised eleven times each, dependant on the timing of the sample submission.

### 7.2.3 Analysis

Illumina reads were assessed for quality using FastQC version 0.11.4 (www.bioinformatics.babraham.ac.uk/projects/fastqc/) and subsequently trimmed to improve quality using Trimmomatic version 0.36 (Bolger *et al.* 2014) using the following parameters: ILLUMINACLIP to remove Truseq 3.0 adapters, removal of leading and trailing low quality bases, and trimming when the average quality of a four base sliding window fell below a Phred quality score of 15. After this process, reads shorter than 36 bp were removed from the sample set, as were reads whose pair did not

meet the quality threshold. Reads were assessed post-trimming using FastQC to confirm removal of adapters and assess post-processing quality. In addition, samples that had greater than 20,000,000 read pairs were subsampled to this depth to allow for downstream processing using the available computing resources.

After trimming, paired reads were aligned to E58 using the Burrows-Wheeler Aligner (BWA) algorithm 'bwa mem' version 0.7.12 (Li & Durbin 2010). The completed Sequence Alignment/Map (SAM) file was converted to a sorted Binary Alignment/Map (BAM) format using SAMtools version 1.3.1 (Li *et al.* 2009) and coverage across the reference genome was summarised using BEDTools version 2.26.0 (Quinlan & Hall 2010). Genomes that had a depth of coverage of greater than 10 across greater than 90% of the reference genome underwent *de novo* assembly with SPAdes 3.9.0 (Bankevich *et al.* 2012), using built in error correction. Resulting contigs were matched to a custom BLAST database containing all reference genomes of the *Chlamydiaceae* family, in addition to plasmid sequence from *C. pecorum* L1 (Genbank accession number: KT223773). Assembled contigs were assessed with QUAST 4.4.1 (Gurevich *et al.* 2013) to assess assembly quality and identify potential assembly errors in relation to the reference genome (*C. pecorum* E58). Contigs were then visualised against the reference genome using Geneious 10.0.9 (Biomatters Ltd). Where appropriate assemblies were manually curated to fix assembly errors in contigs as identified in QUAST, remove contigs with a 'longest k-mer' coverage of less than one (determined by SPAdes), and to merge overlapping contigs. Regions with no coverage were left as gaps for downstream analysis. In addition to the genomes generated in this project, three other published *C. pecorum* draft genomes of koala origin were included in further analysis: MC/MarsBar (Genbank accession number: AZBC01000000), IpTaLE (AZBE01000000) and DBDeUg (AZBB01000000) (Bachmann *et al.* 2014). Two previously described koala origin *C. pecorum* genomes were also included, but their genomes were reassembled from short reads: Gun/Koa1/Ure (Genbank accession number: SRR1693763) and SA/k2/UGT (SRR1693792) (Bachmann *et al.* 2015).

Genomes were annotated with Prokka version 1.12 (Seemann 2014), which produces annotations for protein-coding features (CDS) with Prodigal version 2.6 (Hyatt *et al.* 2010), transfer RNA (tRNA) features with Aragon version 1.2 (Laslett & Canback 2004) and ribosomal RNA (rRNA) features with Barrnap version 0.7 (github.com/tseemann/barrnap). Prokka then utilises a combination of BLAST+ version

2.5 (Camacho *et al.* 2009) and HMMER version 3.1 (Finn *et al.* 2011) to elucidate possible protein functions of annotated regions based on the UniProt knowledge database (The UniProt Consortium 2017). Signal peptides and non-coding RNAs are detected within Prokka using SignalP (Petersen *et al.* 2011) and Infernal (Kolbe & Eddy 2011), respectively.

The resulting annotated genomes were then used in pangenome analysis using Roary version 3.7.0 (Page *et al.* 2015) to determine the core genome of *C. pecorum,* and to align individual genes from multiple genomes. The protein BLAST identity threshold was set at 75% for clustering in Roary. This threshold was utilised so that Roary would not consider highly polymorphic coding regions, such as *pmp* genes, as separate protein clusters, as it would if the default value (95%) was used. The gene scoring tool Scoary version 1.6.9 (Brynildsrud *et al.* 2016) was used to statistically compare gene differences between *C. pecorum* genomes. Within Scoary, a Fisher's exact test is used to compare gene presence/absence with an associated binary variable, with adjustments for false discovery rate using the method described by Benjamini and Hochberg (1995). Scoary was used to compare *C. pecorum* from different geographical regions (northern and southern koalas) to assess any regional differences, and *C. pecorum* detected from either urogenital or ocular swabs, in an effort to elucidate genes associated with tissue tropism. Scoary uses a pairwise comparison algorithm, in conjunction with a binomial test, to correct for population structure. Snippy version 3.1 (github.com/tseemann/snippy) was used to determine genomic variants such as SNPs and insertions and deletions (indels) between E58 and each draft genome. Core SNPs, that is SNP bases that are present in each draft genome (but not necessarily different to the reference in all genomes), were then determined by Snippy and a core SNP genome was produced for each sample using the function snippy-core with default parameters.

Alignments of nucleotide sequences for whole genomes, plasmids, concatenated genes, and individual genes (within Roary) all utilised MAFFT version 7.305b (Katoh et al. 2002). For whole genome comparisons, poorly aligned sections were removed with Gblocks version 0.91b (Castresana 2000), as previously described (Jelocnik et al. 2015). Phylogenetic trees of alignments of *C. pecorum* genomes (treated with Gblocks), SNP core genomes (produced by Snippy), concatenated MLST genes, *ompA* genes, and plasmids, were built using MrBayes version 3.2.6 (Huelsenbeck & Ronquist 2001). Nucleotide substitution rate was determined for each tree using the Akaike Information

Criterion (AIC) (Akaike 1987) within jModelTest version 2.1.10 (Darriba et al. 2012), which in turn utilises the maximum likelihood tool PHYML (Guindon & Gascuel 2003). Each MAFFT alignment, in FASTA format, was converted to the NEXUS format using EMBOSS' seqret (Rice et al. 2000), and a 'Bayes block' containing options and priors for MrBayes was appended to the NEXUS file for each gene. An example Bayesian priors block can be found in Appendix 9. Each Bayesian MCMC tree construction was undertaken with 2 runs, each of 4 chains, and 1,000,000 permutations, with a 25% burn-in. Each consensus tree was built using a 'half majority rule'. Summary statistics to assess convergence also utilised this 25% burn in. Phylogenetic trees for individual core gene alignments were built in the same manner as above, however permutations were stopped if the average standard deviation of the split frequencies fell below 0.01, and the 'potential scale reduction factor' statistic equalled $1.00 \pm 0.10$.

To assess potential regions of recombination, the Gblocks treated genomes were assessed with Gubbins (Croucher *et al.* 2015) using default settings (i.e. 5 iterations of RAxML version 8.2.10 (Stamatakis 2014)). Gubbins uses spatial scanning of the genome to identify regions with elevated numbers of SNPs suggestive of horizontal gene transfer. The resulting output was visualised in Phandango (Hadfield *et al.* 2017).

The 'compare' tool within the python package ETE3 (Environment for Tree Exploration) (Huerta-Cepas *et al.* 2016) was used to compare whole, SNP, and plasmid genomes, as well as trees built for genes of interest. This was to evaluate the similarity of branching distances between trees, and to identify trees with a phylogeny that accurately represented genomic level phylogenetics clustering. To investigate potential positive selection within individual genes, the codeml program within PAML 4.9 (Yang *et al.* 2000; Yang 2007) was utilised. The MAFFT alignment and MrBayes phylogenetic tree for each gene was used as input, and three site model calculations performed (one ratio, nearly neutral selection, and positive selection) (Goldman & Yang 1994; Nielsen & Yang 1998; Yang & Nielsen 1998; Yang *et al.* 2005). Within the 'one ratio' model, codeml produced omega values (dN/dS) for each gene. This is an assessment of synonymous (dS) and non-synonymous (dN) nucleotide changes within a genome alignment. Omega values less than 1 are considered to be evidence of purifying selection, those equal to 1 are considered neutral, whilst a ratio greater than 1 is considered evidence of Darwinian or positive selection. An example codeml control file

can be found in Appendix 10. The comparison of the 'one ratio' or 'null' model (M0) log likelihood ratio with those of 'nearly neutral' (M1a) and positive selection (M2a) models was used to investigate omega values affected by sites within a gene and branches within a phylogeny, as described by Jeffares *et al.* (2015). Briefly, the log likelihood ratio of M0 is compared to the lowest log likelihood ratio of M1a and M2a using a likelihood ratio test (LRT), with a critical $\chi^2$ value for 1 degree of freedom and an alpha of 0.001 ($\chi^2 = 10.83$). If the LRT statistic is greater than the critical $\chi^2$ then the null model, and the presence of sites where omega is 1, can be considered valid. The log likelihood ratios of the M1a and M2a model were then compared to each other, using the critical $\chi^2$ value for 2 degrees of freedom and an alpha of 0.01 ($\chi^2 = 9.21$). If the LRT statistic is greater than the critical $\chi^2$ then the M2a model is statistically valid (i.e. a gene is undergoing positive codon selection). Other metrics assessing the individual gene evolution that were calculated include nucleotide diversity, the number of segregation sites (that is, polymorphisms between strains), the number of haplotypes and Tajima's D, which is a measure of 'random' selection. Nucleotide diversity values of greater than 0.02 were considered of interest based on prior research focusing on *C. pecorum* (Marsh *et al.* 2011). Each of these was assessed using functions within the R (R Core Team 2016) packages APE (Paradis *et al.* 2004) and pegas (Paradis 2010).

In addition to the core gene phylogenetic analysis described above, analysis was conducted using *C. pecorum* MLST genes (Jelocnik *et al.* 2013), putative virulence genes and potential epidemiological markers (*ompA*, ORF663, and *incA*) (Marsh *et al.* 2011; Yousef Mohamad *et al.* 2014), as well as the full genome of the *C. pecorum* plasmid (Jelocnik *et al.* 2015; Jelocnik *et al.* 2016). Allele and sequence types for the *Chlamydiales* MLST (Pannekoek *et al.* 2008) were compared to those held in the PubMLST database (https://pubmlst.org/ chlamydiales/) sited at the University of Oxford (Jolley & Maiden 2010), as described in section 2.3.6. Non-parametric statistical comparisons between genomic features and geographical regions, site of infection, or disease status, were undertaken with Minitab 17 (Minitab Inc). Summaries of data determined to be non-parametric, based on a normality test using the Anderson-Darling statistic, are reported as medians.

## 7.3 Results

### 7.3.1 Genome assemblies

The median DNA concentration of the 41 koala urogenital samples submitted for full genome sequencing from the APCAH sample archive was 70.5 ng/µL (range: 20.9 – 266.2 ng/µL). The qPCR determined that the median genomic copy numbers in the extracted DNA was 11,526 copies/µL of eluted DNA (range: 83 – 705,600 copies/µL of eluted DNA). Details for individual samples are in Table 7.1 (APCAH) and Appendix 11 (USC).

The median number of paired reads obtained from Illumina sequencing, combining data from APCAH samples in addition to those from the USC, was $2.34 \times 10^6$ (range: $8.00 \times 10^4 – 1.22 \times 10^8$). After trimming for quality, the median number of reads obtained was $2.23 \times 10^6$ (range: $7.32 \times 10^4 – 1.18 \times 10^8$), with a median quality score of 35.1 (range: 34.2 – 37.3) and GC content of 43% (range: 41 – 49%). Details for all reads before and after trimming are in Appendix 12. The number of reads in six samples were greater than the available computing resources could process, and were subsampled, without replacement, to a depth of 20,000,000 reads prior to downstream analysis.

A summary of the depth of coverage for each genome compared to E58 is available in Appendix 13, with visual plots for a subset of genomes in Figure 7.1. In brief, the median value of the average depth for each sample was 43.5 (range: 0.09 – 2777.1). The median coverage of each sample's reads to E58 was 98.6% (range: 7.3 – 100%), whilst the median coverage where depth was greater than 10 reads was 97.2% (range: 0.01 – 99.96%). This represents a median of 15,929 bp with zero coverage across the 1,106,197 bp reference genome, or 31,582 bp where depth was 10 reads or fewer. Plotting the coverage depths highlighted that there was an overrepresentation of reads mapping to the region coinciding with the 16S and 23S rRNA genes, whereas an underrepresentation was seen in both the central polymorphic membrane protein (Pmp) coding region and the PZ (Figure 7.1).

Of the 110 samples initially selected for sequencing, 61 met the criteria for continuing with *de novo* assembly (i.e. greater than 90% coverage of E58 with a depth of at least ten reads) (Appendix 13). One sample, USC_No_Heri_Eye, could not be assembled in SPAdes due to unresolvable segmentation faults within the software. *De novo* assembly produced a median of 47 contigs per sample (range: 4 – 12566) with sufficient identity

to the custom *Chlamydiaceae* database based on BLAST results. Focusing on contigs over 1000 bp, the median number of contigs per sample was 15.5 (range: 2 – 269), with an average nucleotide identity to the *Chlamydiaceae* database of 93.2% (SD ± 4.3%) across 60 samples. The median value for the average length of contigs greater than 1000 bp in each sample was 72,492 bp (range: 4092 – 559,032 bp). Values for average nucleotide identity and average contig length for each individual sample are in Appendix 14. Analysis with QUAST identified that contigs successfully mapped to a median of 95.5% of the reference genome (range: 92.5 – 99.3%), and detected a median of 0 assembly errors per genome (range: 0 – 8) (Appendix 14). Two genomes, USC_RayIs_13-14_UGT and USC_RayIs_6-14_UGT failed to produce any contigs larger than 50,000, and consistently had the lowest qualitative metrics. To eliminate the risk of producing erroneous genomes, these samples were not included in downstream analysis.

### 7.3.2 Genomic differences between *C. pecorum* infecting koalas

After contig mapping to E58, manual curation, and the concatenation of contigs, draft *C. pecorum* genomes from 57 koala samples were produced. In addition to the two genomes listed in the previous section, the large number of small contigs produced for 102_RayIs_3D2_B_UGT made resolution of the draft genome impracticable. All downstream analysis utilised the previously published draft *C. pecorum* genomes DBDeUG, IpTaLE, and Mc_Marsbar, allowing for a total of 60 genomes to be assessed. A breakdown of the region of origin and sample site of these swabs can be found in Table 7.2**Table 7.2.**. The median genome length of all samples was 1.11 Mb (range: 1.09 – 1.11 Mb). Alignment of the raw genomes showed a nucleotide identity of 92.5% across all samples. This improved to 97.8% after processing through Gblocks to remove poorly aligned sections, with an alignment length of 1.07 Mb. The GC content in both instances was 41% across the 60 genomes included in the alignment. A phylogenetic tree built from full genomes revealed that *C. pecorum* from southern koalas generally fell into a separate clade to those from northern koalas (Figure 7.2). The exception to this were two subclades, one containing a single genome (USC_SA_12-220_LE), and another containing three genomes (123_MoPen_4G5_C_UGT, 127_Gipps_4C10_C_UGT, 134_MoPen_15B5_C_UGT). In all genomic trees (full genome, MLST, plasmid, SNP, Gubbins), these two subclades, containing southern koala *C. pecorum* genomes, were clustered with, but were distinct from, clades

containing northern koala *C. pecorum* genomes (Figure 7.2, Figure 7.3, Figure 7.4, Figure 7.5, Figure 7.6). Additionally, strains from Gippsland, Victoria (101_Gipps_2G12_F_UGT, 124_Gipps_4E6_C_UGT, and 125_Gipps_4H6_F_UGT) and one strain from Cape Otway, Victoria (116_SWCoast_3H9_L_UGT) generally formed their own separate clades. Recombination, as derived by Gubbins' algorithm, was abundant across the genomes, as highlighted in Figure 7.3. The total length of the filtered polymorphic bases produced by Gubbins was 8074 bp.

### 7.3.3 Draft genome variant analysis

The median number of SNPs in coding regions between each genome and E58, as generated by Snippy, was 4389.5 (range: 2908 – 5368), with the majority of SNPs occurring within these regions rather than intergenic regions (median: 91.7%, range: 91.0 - 92.8%). The median percentage of SNPs between each genome and E58 that were non-synonymous was 46.5% (range: 41.6 – 52.7%). A smaller number of indels were detected, with a median of 211.5 (range: 138 – 251). A majority of these occurred outside of the predicted coding regions (median: 62.2%, range: 55.4 – 67.0%). A core SNP alignment of 17,138 SNPs was generated using snippy-core. A total of 15,703 (91.6%) of these SNPs occurred in predicted coding regions of E58 (based on annotation with Prokka). A phylogenetic tree was generated from this core SNP alignment (Figure 7.4).

### 7.3.4 Novel MLST genes in *C. pecorum*

For 44/57 of the draft genomes, *oppA,* one of the seven housekeeping genes used for *C. pecorum* MLST analysis, was not completely assembled. The full-length sequence of *oppA* from all 22 genomes from the APCAH sample archive were confirmed by Sanger di-deoxy sequencing as previously described (sections 2.2.3, 2.3.4, and 2.3.6 ). For consistency, when determining MLST sequence type for the USC samples, *oppA* type 12 was used as a placeholder allele type when full length *oppA* could not be elucidated, as prior research identified it as the most common *oppA* allele type in *C. pecorum* (119/132 isolates previously submitted to the PubMLST database). This approach was supported by the observation that, for each of these genomes in which *oppA* was only partially sequenced, this partial sequence was identical to *oppA* type 12. In total, 18 novel sequence types (NSTs) were detected across the koala origin *C. pecorum* genomes (Table 7.3.), with 12/18 occurring just once. The most common sequence type

was 'NST 2', of which 18 genomes belonged. This ST occurred exclusively in southern koalas. NSTs had SNPs in at least one of the *enoA*, *fumC*, *gidA*, or *gatA* genes compared to previously published allele sequences. A phylogenetic tree was generated from concatenated MLST sequences, omitting *oppA* from all sequences (Figure 7.5).

### 7.3.5 Plasmid variation

Full-length, or near full-length, sequence of the *C. pecorum* plasmid was obtained from *de novo* assemblies in all but two samples (2/61). This included full length or near full-length plasmid sequence from the three samples where contigs were not analysed beyond initial assembly (102_RayIs_3D2_B_UGT, USC_RayIs_13-14_UGT, and USC_RayIs_6-14_UGT). Full-length was considered approximately 7549 bp, based on the length of previously published *C. pecorum* plasmids (Jelocnik *et al.* 2015). Two plasmid genomes were near full-length: 203_Bella_290513_UGT, and USC_SA_12-220_LE, with lengths of 7462 and 7385 bp, respectively. The two instances where no plasmid sequence was generated were the previously published genomes USC_Gun_koa1_UGT and USC_SA_K2_UGT. After processing a MAFFT alignment of the 61 plasmid genomes (including plasmids from the previously published genomes MC/MarsBar, IpTaLE, and DBDeUg) through Gblocks, the alignment was 7255 bp long with 7161 (98.7%) identical sites, and a phylogenetic tree was generated (Figure 7.6).

### 7.3.6 Gene variation between koala-derived *C. pecorum*

The average number of genes across the genomes included in pangenome analysis, based on annotation with Prokka, was 991 (range: 983 – 1003) (Appendix 15). Using Roary it was determined that the genome of *C. pecorum* detected in koalas contained 915 core genes (those occurring in at least 99% of genomes) and 8 soft core genes (between 95 and 98%, inclusive) across the 60 genomes. Analysis with Scoary identified 8 genes that were significantly more likely to occur in *C. pecorum* genomes from either northern (5/8) or southern (3/8) koalas (BH-adjusted Fisher's exact test $P < 0.05$) (Appendix 16). Six genes were significantly more likely to occur in *C. pecorum* genomes from either ocular (4/6) or urogenital (2/6) swab samples (BH-adjusted Fisher's exact test $P < 0.05$) (Appendix 16). Further analysis identified that in four cases (assessed by Scoary as significant for both region and tissue tropism), the 'missing' genes were in fact present, but with amino acid homology lower than the 75% threshold

utilised in Roary for identifying proteins (Appendix 16). No genes potentially associated with tissue tropism that were identified as significant using a Fisher's exact test had a 'best case' pairwise comparison binomial test *P* value less than 0.05, suggesting a strong population structure bias (Brynildsrud *et al.* 2016). One significant difference of note (BH-adjusted Fisher's exact test $P < 0.001$) was that 22/27 (81%) *C. pecorum* genomes from southern koalas had a non-synonymous SNP in the tRNA-Cys(aca) region, preventing the tRNA from being predicted by Prokka, whilst only 8/33 (24%) *C. pecorum* genomes from northern koalas had the same SNP. This SNP was present in 26/44 (59%) *C. pecorum* genomes of urogenital tract origin and 4/16 (25%) *C. pecorum* genomes of ocular origin, however this discrepancy was not significant (BH-adjusted Fisher's exact test $P = 0.21$).

Analysis of complete PZ regions found that coding region structure mostly concurred with previously described *C. pecorum* genomes, containing the MACPF gene, purine biosynthesis genes, two cytotoxin coding regions, acetyl-CoA-carboxylase genes, and phospholipase D (PLD) coding regions (Mojica *et al.* 2011; Sait *et al.* 2014). The two notable differences when comparing the PZ between *C. pecorum* genomes were the variable number of PLD genes (Table 7.4), as well as the presence or absence of an open reading frame (ORF) encoding a hypothetical protein towards the 5' end of the PZ ('hORF-PZ') (Figure 7.8). A large proportion (mean: 68% per genome, SD ± 20%) of the annotated *pld* genes within each genome were predicted to encode truncated proteins, resulting from frameshifts created by variable numbers of repeats in homopolymer tracts at the 5' end of the individual coding regions. A similar phenomenon occurred in the majority of genomes with full length *toxB* coding regions present (76.9% 40/52). In those instances, the coding region was present, but a frameshift at the 5' end reduced the annotated gene length by ~200 bp due to the insertion of a premature stop codon (Figure 7.8). 'hORF-PZ' was identified in 27/33 (82%) *C. pecorum* genomes from northern koalas and 21/27 (78%) *C. pecorum* genomes from southern koalas, and was not significantly different between regions based on analysis by Scoary (BH-adjusted Fisher's exact test $P = 0.91$). Transmembrane Hidden Markov models within the Geneious plugin Transmembrane Prediction Tool version 0.9 were used to predict that 'hORF-PZ' had an extracellular and transmembrane domain (Figure 7.9). A translated BLAST (blastx) search using the NCBI non-redundant protein sequence database identified that it is also present in *C.*

*pecorum* strains L1 and L17. Orthologues with 50% amino acid similarity, but no predicted extracellular or transmembrane domains, are also present in the PZs of *C. trachomatis* and the porcine pathogen *C. suis*. Typically, 'hORF-PZ' was flanked on either side by regions predicted to code for PLDs.

In total, 852/962 of the protein coding regions predicted by Prokka occurred in complete form in all 60 genomes. Analysis of the nucleotide diversity of each gene revealed that *ompA* had the highest diversity (0.091). Details for the 19 coding regions with a nucleotide diversity greater than 0.02 are in Table 7.5. Of note, four encoded Pmps, two (*incA* and *incC*) coded for inclusion membrane proteins, and five were consecutive open reading frames. These five coding sequences had an average haplotypic diversity of 8.2 (range: 3 – 11), and encompassed the *tyrS* gene (which encodes Tyrosine-tRNA ligase), one ORF encoding a hypothetical protein upstream of *tyrS* (between *tyrS* and *sigD*), and three downstream genes including *gnd* (6-phosphogluconate dehydrogenase), *lepA* (elongation factor 4) and another ORF encoding a hypothetical protein. The number of segregation sites in *tyrS* was 660, more than double those identified in *ompA* (312).

After assessment of omega values in codeml, genes were categorised as either being under neutral (less than 1) or positive selection (greater than 1). Based on these criteria, a total of 73 genes were undergoing positive selection and 721 genes were under neutral selection. There were 92 genes with either no synonymous or non-synonymous differences, and therefore the omega value could not be calculated. Using the likelihood ratio tests between the two codeml models (for nearly neutral and adaptive evolution), it was found that 42 coding regions had codons under significant positive/adaptive selection (critical $\chi^2 = 9.21$, degrees of freedom (df) = 1, $P < 0.01$) (Table 7.6).

### 7.3.7 Other genomic features

A summary of the findings from potential virulence genes examined in previous research is shown in Table 7.4. As has been detected in previously sequenced *C. pecorum*, there were 15 *pmps*. These were found across four regions of the genome, with one cluster containing a series of eleven *pmp* genes, as has previously been described (Sait *et al.* 2014). Four of the *pmp* regions (annotated as *pmp15/E*, *pmp13/G*, *pmp16/E* and *pmp14/H*) had greater than 0.02 nucleotide diversity and more than 300

segregation sites. Analysis with PAML identified that two of the *pmp* regions (predicted to encode Pmp21/D and Pmp14/H) had significant ($P < 0.01$) positive codon selection.

Phylogenetic analysis based on *ompA* types (Figure 7.7) was similar to the full genome phylogenetic tree. The exception to this were the genotype F strains, which clustered with northern origin strains of the same genotype, whereas in the full genome phylogenetic trees these strains were more closely related to other Victorian origin strains. Additionally, 124_Gipps_4E6_C_UGT clustered with other genotype C strains from Gippsland and Mornington Peninsula amongst northern strains, compared to full genome analysis where this genome clustered with other southern koala origin *C. pecorum* genomes. Mixed infection could be identified via contigs with different *ompA* genotypes being assembled from the same samples. This occurred in significantly fewer samples from southern koalas (14.8%; 4/27) compared to northern koalas (70%; 21/30), based on a two-tailed Fisher's exact test (Difference = 55.2%; 95% confidence interval (CI) = 34.0, 76.4%; Z = 5.11; $P < 0.001$). Four of the 25 instances of mixed infections appeared to contain more than two distinct *ompA* genotypes, however in all cases, the minor infection (i.e. the *ompA* contig with the lower read coverage) was incomplete, covering 1 – 3 of the 4 variable domains (Table 7.4). The read depth across these minor strain *ompA* contigs was less than five in the majority of samples, and several fold lower than the dominant strain in all cases, making mixed strain contigs unlikely.

A type III secretion system effector protein, IncA, contains a variable number of APA motifs, which are hypothesised to be associated with virulence of *C. pecorum* (Yousef Mohamad *et al.* 2014). The median number of APA motifs encoded by *incA* in the *C. pecorum* genomes was 5 (range: 0 – 12) (Table 7.4). A Kruskal-Wallis test demonstrated that there was a significant difference (H = 13.36, df = 1, $P < 0.001$) between the median number of APA motifs in *C. pecorum* genomes from southern koalas (median: 3, range: 0 – 9) and northern koalas (median: 6, range: 1 – 12). Only four *C. pecorum* samples lacked the APA motif entirely. Kruskal-Wallis tests found no significant difference between the number of APA motifs and site of infection (UGT n = 44, median = 4.0, range = 0 – 10; ocular n = 16, median = 5.5, range = 1 – 12; H = 2.44, df = 1, $P = 0.118$), nor the presence of chlamydial disease (present n = 37, median = 4.0, range = 0 – 12; absent n = 16, median = 4.5, range = 0 – 9; H = 0.00, df = 1, $P = 0.969$). All samples lacking the APA motif were from the southern koala group, from

koalas sampled within either the Gippsland or Mornington Peninsula regions of Victoria. The second most common variant detected after APA was the APE motif.

Yousef Mohamad *et al.* (2014) also noted an association between virulence and the number of 15-nt coding tandem repeats (CTR) in ORF663, which encodes a hypothetical protein. A Kruskall-Wallis test found that the median number of CTRs was significantly lower in *C. pecorum* from southern koalas (n = 20, median = 6, range: 2 - 10) than from northern koalas (n = 24, median = 10, range: 3 - 30) (H = 7.93, df = 1, *P* = 0.005). However, there was no significant difference between the median number of repeats in the genomes from ocular infections (n = 11, median = 7, range: 3 - 22) and urogenital infections (n = 33, median = 6, range: 2 - 30) (H = 0.54, df = 1, *P* = 0.464). Nor was there a significant difference in the number of CTRs in ORF663 between koalas with clinical disease present (n = 28, median = 7, range: 3 – 30) or absent (n = 13, median = 6, range: 2 – 13) (H = 1.29, df = 1, *P* = 0.257). SNPs resulting in truncations due to frame shifts in this hypothetical coding region meant that 19 genomes were not included in the analysis (Table 7.4).

### 7.3.8 Tree distances and selection pressure on genes

Tree comparison analysis identified similarities between trees built using alignments resulting from full genomes processed through Gblocks and through Gubbins, as well as alignments of MLST genes, SNPs, plasmids, and *ompA*. These comparisons are summarised in Table 7.7. The tree comparison with the lowest normalised Robinson-Foulds metric, representing the most similar trees, was the comparison between the full genome tree and the SNP tree.

## 7.4 Discussion

*C. pecorum* infects koalas throughout Australia, including northern koala populations in Queensland and New South Wales, and southern koalas in Victoria and South Australia. The work presented in this chapter is the most comprehensive comparative study of *C. pecorum* infecting these different populations of animals. Previously the largest study of this kind had included only 3 *C. pecorum* genomes, all from Queensland koalas (Bachmann *et al.* 2014).

Hybridization techniques, by allowing sequencing of full or nearly full genomes from limited sample amounts, without the need for culture of the organism, have unlocked the possibility of sequencing genomes from larger sample sets than previously possible.

This approach however has some limitations. In particular, the design of probes for *C. pecorum* in koalas utilised the reference genome for *C. pecorum*, E58, which was originally isolated from a case of sporadic bovine encephalomyelitis in the USA (McNutt & Waller 1940). The divergence of this E58 reference genome from that found in koalas could explain why some regions in the target genomes have a low depth of coverage when the reads are mapped to the reference sequence. This primarily occurred in hypervariable regions, particularly the chlamydial PZ and the *pmp* genes. The completion and validation of reference genomes for *C. pecorum* from koalas, such as MC/Marsbar, which was sequenced by shotgun sequencing and is published in draft form (Bachmann *et al.* 2014), may allow hybridisation probes specifically for koala origin *C. pecorum* to be designed and utilised. The size of the PZ (~42 kb in E58, up to ~48 kb in the genomes assessed in this study) makes long range PCR and conventional di-deoxy sequencing of PCR products for all genomes produced in this study unfeasible. The development of long read sequencing technology, such as those offered by Oxford Nanopore and Pacific Biosystems, could be used in future studies to improve the assembly process to allow complete genome construction. Such analysis would be particularly useful in resolving contig gaps which arise in the *de novo* assembly of Illumina short reads, as well as accurately assembling repeat and hypervariable regions.

### 7.4.1 Genome assembly comparisons

More than half the samples utilised in this study had a depth of coverage sufficient for producing genomes of a similar length to those previously published after concatenation of contigs (Bachmann *et al.* 2014; Sait *et al.* 2014; Bachmann *et al.* 2015; Jelocnik *et al.* 2015). Previous research on koala-associated *C. pecorum* using whole shotgun sequencing approaches produced genomes of length 1.09 Mb, with N50 values between 277,726 and 587,274 (Bachmann *et al.* 2014). The genomes presented here have an average length of 1.10 Mb and median N50 value of 630,015, for a much larger sample set and utilising only swab samples, highlighting the value of the hybridisation method. N50 values measure the minimum contig length required to cover 50% of the total length of the assembly, and are a mathematical means of highlighting the strength of a *de novo* assembly. The near complete genomes produced also had a similar GC content compared to previously sequenced *C. pecorum* genomes, with the average of 41% being comparable to both livestock and koala *C. pecorum* strains previously assessed (Bachmann *et al.* 2014; Sait *et al.* 2014; Bachmann *et al.* 2015; Jelocnik *et al.* 2015).

### 7.4.2 Phylogenetics of *C. pecorum*

The phylogenetic tree produced by aligned genomes, processed through Gblocks, and those produced by plasmid alignment, MLST, and SNPs were relatively congruent. Similar congruence has previously been identified between the *C. pecorum* plasmid and MLST phylogenetic trees in a smaller dataset containing samples from livestock and koalas (Jelocnik *et al.* 2016). The phylogenetic separation of northern and southern origin *C. pecorum* strains, aside from a few examples in specific regions, may reflect koala population structures historically. This was mirrored in the detected *ompA* genotypes, in that genotype B was prominent in Victorian samples, whilst northern koalas were infected with a more diverse range, including genotypes A, F, G, and H. These regional *ompA* genotypes have been highlighted previously (Kollipara *et al.* 2013b; Legione *et al.* 2016b). The history of koala populations in southern Australia includes a population bottleneck caused by a decline in the species followed by a large repopulation effort in Victoria. This repopulation resulted in the majority of koala populations around the state being founded through translocations from populations on Phillip Island and French Island (Martin 1989). Such events would have likely created genomic bottlenecks for *C. pecorum*, where infected koalas would have most likely been taken from Phillip Island only, as French Island has historically been considered either free of *C. pecorum* (Martin & Handasyde 1999; Patterson et al. 2015) or as having only a very low prevalence of *C. pecorum* infection (McColl *et al.* 1984; Legione *et al.* 2016a). The separate clade of Victorian *C. pecorum*, encompassing samples from the Mornington Peninsula and Gippsland, both on mainland south-eastern Victoria, may represent historical remnant populations of koalas that survived the population decline in southern Australia. The unique *C. pecorum* identified in Cape Otway, Victoria (116_SWCoast_3H9_L_UGT), in conjunction with the low prevalence of *C. pecorum* in this population (Legione *et al.* 2016b), may be indicative of a recent cross species transmission from an as yet unidentified species. Another possibility, raised by mitochondrial DNA sequencing, is that this part of this koala population may have been established through undocumented translocations of koalas from New South Wales (Neaves *et al.* 2016). In turn, it could be hypothesised that the *C. pecorum* that affects koalas in this region shares this northern origin. Similarly, koalas in South Australia originated through recorded translocations from Victoria, and possibly unrecorded translocations from New South Wales, as evidenced by shared

mitochondrial haplotypes (Neaves *et al.* 2016). The translocation of individuals infected with *C. pecorum* from both northern and southern koala populations may explain the mixed clustering of *C. pecorum* genomes found in koalas from South Australia.

### 7.4.3 Tissue tropism and strain type

This study was unable to elucidate any genomic differences between *C. pecorum* infecting different anatomical sites (either ocular or urogenital sites). In cases where *C. pecorum* genomes were identified in both ocular and urogenital samples from the same individual, the genomes of the *C. pecorum* infecting the two sites were highly similar, with pairwise identity ranging from 99.3% to 99.9% across the full genomes in the four individuals (USC_NSW_Chingee_Eye/UGT, USC_Haz_Bo_Eye/ USC_QLD_Bobby_UGT, USC_QLD_Chestnut_LE/RE/UGT, and USC_QLD_Travis_LE/UGT). This increased to greater than 99.9% for all four individuals when using Gblocks processed genomes. In all but one case, the strains infecting different sites in the same individual had the same MLST profile. The exception was USC_QLD_Travis_LE/UGT that had different allele profiles in the *gidA* region, with 3 SNPs. Previous research investigating pathogenicity and tissue tropism of *C. pecorum* in livestock, utilising MLST genes, found that particular sequence types were more likely to cause either polyarthritis or ocular infection, compared to those which did not elicit clinical signs of disease (Jelocnik *et al.* 2014b). Phylogenetic analysis of concatenated MLST genes from the genomic samples did not reveal similar findings in koalas, nor did any other phylogenetic analysis. These results are likely to be confounded by geographical clustering of the ocular samples, with all but two being from northern koalas. Therefore, it is difficult to tease out phylogenetic differences that result in tissue tropism, rather than those that are a result of natural strain divergence due to regional differences.

### 7.4.4 Genome annotation comparisons

The 'core genome' of *C. pecorum*, being genes that occurred in all genomes in the dataset, included 95% (915/962) of the protein coding regions identified by Roary and Prokka. This reflects the syntenic nature of the *C. pecorum* genome, as previously described (Mojica *et al.* 2011; Sait *et al.* 2014), where the majority of genomic variation is concentrated in the PZ. Of the non-core genes (as determined by Roary), 17 were annotated in only a single genome and likely represent assembly related artefacts, or

pseudogenes arising from truncations. A further ten genes not in the core genome were absent from fewer than five genomes, which may also be due to similar limitations with genome assembly.

### 7.4.5 Variation within the plasticity zone

The PZ is known to be a region of high variation in chlamydial genomes, and the dataset in this study had similar findings. The majority of genomes (48/55, 87%) assembled here encoded 'hORF-PZ', a hypothetical protein within the PZ, with 50% amino acid similarity to those encoded within the PZ of *C. trachomatis* and *C. suis*, and the same hypothetical protein can be identified in two previously published *C. pecorum* genomes of porcine origin (L1 and L17) (Jelocnik *et al.* 2015). The function of this hypothetical protein requires further investigation. A variable number of PLD coding regions were found in the PZ in this study, ranging from 2 to 7, and occurring upstream, downstream, and between the two cytotoxin genes in the PZ. Variation in PLD copy number has been identified previously in *C. pecorum* genomes of both livestock (Sait *et al.* 2014) and koala origin (Bachmann *et al.* 2014). It has been suggested that the PZ-PLD genes in *C. trachomatis* and *C. muridarum*, the only other *Chlamydia* spp. with PZ-PLD genes, are involved in species-specific adaptation, as phylogenetic evidence indicates gene duplication within each species occurred after the PZ-PLD gene diverged from a single common ancestor (Thomson *et al.* 2008). PLD genes are a major cytotoxin in other bacteria, suggesting they may be virulence factors of *Chlamydiaceae*. For example, in *Corynebacterium pseudotuberculosis*, a pathogen of sheep and goats, PLD is an exotoxin capable of inducing chemotaxis of neutrophils (McKean *et al.* 2007). Elucidating the primary function of PLD genes in *C. pecorum* may require the development of genetic techniques to create defined mutants and an appropriate animal model.

All genomes assembled had two cytotoxin genes (*toxA* and *toxB*) present in the PZ, but many had a poor level of read coverage in this region, as described earlier. This resulted in only 26 genomes where the full coding region of both cytotoxins was completed, and 26 genomes where only one of the two cytotoxins were completely assembled (23 of which were *toxB*). A third cytotoxin in the PZ previously identified in a porcine origin *C. pecorum* (Jelocnik *et al.* 2015) was not identified in any genomes in this study. The failure to fully assemble the cytotoxin coding regions in some *C. pecorum* strains may be due to their divergence from the reference probes used for DNA capture, and

targeted long read sequencing would be useful to investigate this further. Notably, the draft genomes of MC/MarsBar, IpTaLe, and DBDeUG also have missing portions, or truncations, of *toxA* and *toxB* (Bachmann *et al.* 2014). These genomes were sequenced using shotgun methodology on cultured isolates, rather than hybridisation methods, and therefore their divergence from E58 should not impact the reads available for assembly. In *C. trachomatis*, there are a range of deletions and truncations in the cytotoxins across strains, and the cytotoxic effect of the bacteria has been shown to be reduced or absent *in vitro* depending on which region of the gene is missing (Belland *et al.* 2001).

Polyguanine or polycytosine sequences were found within the *pld* (6 – 28 nt long) and *toxB* (9 – 18 nt long) genes in the PZ, as well as upstream of truncated versions of these genes. It is likely that this is evidence of phase-variation, a mechanism of immune evasion, in these *C. pecorum* genomes. Phase-variation can occur through slipped-strand mispairing of repeat regions during replication. Such mispairing can result in a frameshift, and a truncation of the ORF (Reviewed in van der Woude & Bäumler 2004). This has been previously described in the PLD regions of *C. pecorum* from livestock (Sait *et al.* 2014) and the *pmp* regions of other *Chlamydiaceae* (Viratyosin *et al.* 2002; Thomson *et al.* 2005). The nature of the short-read sequencing utilised, in conjunction with the low read depth in the PZ, may have given rise to these apparent variants, where single base deletions have caused predicted coding regions to be truncated. Illumina HiSeq is typically robust when encountering indels, but it has been shown that the error rate increases with an increased length of homopolymer (Minoche *et al.* 2011). The use of long read sequencing in conjunction with short read technology may help to elucidate the true nature of this apparent phase-variation.

### 7.4.6 Genomic variation and recombination

In this study *C. pecorum* encoded 15 Pmps in four clusters, similar to results in previous studies (Bachmann *et al.* 2014; Sait *et al.* 2014; Jelocnik *et al.* 2015). The positive codon selection in two of the *pmps* (*pmp21/D* and *pmp14/H*) suggests that mutations in these genes are likely to be advantageous to the bacterium. The largest of the *pmp* gene clusters, containing 11 *pmp* genes, had a high level of recombination, based on the number of recombination blocks assessed by Gubbins. In the past, genetic recombination of intracellular bacteria such as *C. pecorum* was assumed to be low, but recent *C. trachomatis* studies revealed that recombination is surprisingly common. Approximately 51% of the *C. trachomatis* genome contained recombination blocks,

determined using similar methods to those used in this present study (Harris *et al.* 2012). Recombination could occur as a result of co-infection of the same anatomical site with different *C. pecorum* strains, which was observed in this present study as well as previously (Bachmann *et al.* 2015). *In vitro* co-infection studies of *C. trachomatis* have shown that the same cell can be infected by multiple strains, and that separate inclusion vacuoles carrying different strains are capable of fusing, providing a locale for recombination to occur (Ridderhof & Barnes 1989). Whether the same mechanism occurs in *C. pecorum* is unknown.

One of the potential virulence genes, *ompA,* had the highest nucleotide diversity of the genes included in analysis. Despite this, the phylogenetic trees produced from the *ompA* gene alignment clustered in a similar manner to the phylogenetic trees produced from full genome sequences. This analysis used only sequences from which the full *ompA* sequence was available, which excluded the minor (less abundant) strains found in a number of samples where more than one *C. pecorum* strain was present. Dual infections consisting of more than one *ompA* genotype were more commonly detected in northern koalas than in southern koalas. In *C. trachomatis* the genomic region containing *ompA* has previously been identified as being prone to recombination (Gomes *et al.* 2007). If this is true for *C. pecorum ompA*, it is possible that *C. pecorum* recombination and genomic evolution is better facilitated in northern koala populations where there is a higher rate of co-infection with more genetically diverse strains, although this requires further investigation.

The APA virulence-associated motif of IncA occurred more frequently in northern koalas, but no clear association between disease and motif frequency could be determined. Similar findings occurred for the virulence-associated tandem repeat regions in ORF663, where there were significantly fewer repeats in genomes originating from southern koalas, compared to northern, but there were no significant differences between *C. pecorum* from different tissues nor any association with clinical disease. Interestingly *C. pecorum* from southern koalas had fewer repeats in this region (hypothesised to increase pathogenicity), but no APA motifs in IncA (conversely hypothesised to decrease pathogenicity). It may be that despite *C. pecorum* infecting multiple host species, the impact of specific genes on virulence during infection may not be conserved across these different host species.

### 7.4.7 Chlamydia plasmid

This plasmid, which has homologues in other chlamydial species such as *C. trachomatis*, is of interest due to its potential role in virulence. The plasmid is considered to mirror chromosomal evolution (Seth-Smith *et al.* 2009; Jelocnik *et al.* 2016), and has been associated with an increase in pathogenicity of *C. trachomatis*. The plasmid encodes virulence genes that impact the ability of the organism to colonise a host, with studies in mice identifying a 400-fold increase in the median infective dose of strains lacking plasmids (Carlson *et al.* 2008). In koalas, the presence of the *C. pecorum* plasmid has not been strongly correlated with virulence. Indeed, the koala population with the highest proportion of plasmid-positive *C. pecorum* were in Victoria (Jelocnik *et al.* 2016; Legione *et al.* 2016b), where clinical disease in koalas appears to be less severe than disease in South Australia, New South Wales, and Queensland. This present study identified full plasmid sequences from 58/60 *de novo* assemblies. This prevalence (96.7%) was higher than expected, as the combined results of previous studies, encompassing koalas from each state, found the *C. pecorum* plasmid present in 176/238 (73.9%) of strains (Jelocnik *et al.* 2016; Legione *et al.* 2016b). The increased prevalence in this present study may suggest that the plasmid occurs in strains infecting individual koalas more frequently than previously assessed. Another possibility is that the presence of the plasmid allows increased growth of *C. pecorum* in the host, and therefore the samples containing the plasmid in this study also provided the highest level of *C. pecorum* per sample, resulting in better sequencing and assembly. Due to the overwhelming majority of the samples containing this plasmid, it was not possible to identify an association between disease severity or tissue tropism and the presence or absence of the *C. pecorum* plasmid. However, as the plasmid was detected equally in koalas with and without clinical signs, this could suggest that it has a reduced role in the virulence of *C. pecorum*, at least in this host. Determining the impact of the *C. pecorum* plasmid on the pathogenicity of infection in koalas should remain a priority for researchers.

## 7.5 Concluding remarks

This study is the largest genomic comparison of *C. pecorum* infecting a single species to date, and demonstrates interesting differences between strains detected across different koala populations. Notably, phylogenetic trees of *C. pecorum* genomes showed that in most cases there was a clear divergence between strains detected in southern koalas,

compared to those detected in northern koalas. Definitive conclusions about the relationship between genetic differences and disease, or tissue tropism, were not possible. A larger study, incorporating not only koala origin *C. pecorum* but contemporary livestock origin samples, would be of great benefit in fully understanding the evolutionary history of *C. pecorum* in Australia. Determining whether *C. pecorum* infection in koalas predates the arrival of domestic livestock by European settlers needs further examination. Studies on chlamydial disease in indigenous Australians have suggested *C. pneumoniae* occurred in Australia, at least in humans, prior to European settlement (Roulis *et al.* 2015). It is possible that characterisation of modern livestock *C. pecorum* samples, as well as any historical samples present in archived material, may help to determine how an Australian marsupial contracted this significant pathogen.

**Table 7.1.** Samples selected for genome sequencing from the APCAH sample archive. Samples were selected based on a combination of geographical origin, copy numbers detected by previous qPCR screening, and *ompA* type (Chapter 4).

| Sample ID | Region | Population | Infection site | Platform | Accession | *ompA* type [#] | Genome copies/µL ^ | Clinical Signs |
|---|---|---|---|---|---|---|---|---|
| 101_Gipps_2G12_F_UGT | Gippsland, Vic | Southern | UGT | HiSeq 2000 | ERR1431905 | F | 998 | UGT abnormality* |
| 102_RayIs_3D2_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431906 | B | 12640 | None |
| 103_RayIs_3F3_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431907 | B | 14186 | Wet bottom† |
| 104_WestVic_3F4_B_UGT | Far West, Vic | Southern | UGT | HiSeq 2000 | ERR1431908 | B | 91360 | UGT abnormality, wet bottom |
| 105_WestVic_3C6_B_UGT | Far West, Vic | Southern | UGT | HiSeq 2000 | ERR1431909 | B | 16406 | None |
| 106_WestVic_3G6_B_UGT | Far West, Vic | Southern | UGT | HiSeq 2000 | ERR1431910 | B | 13434 | Wet bottom |
| 107_RayIs_3D7_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431911 | B | 19914 | Wet bottom |
| 108_RayIs_3E7_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431912 | B | 28860 | None |
| 109_RayIs_3C8_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431913 | B | 604000 | Wet bottom |
| 110_RayIs_3D8_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431914 | B | 5388 | Wet bottom |
| 111_RayIs_3E8_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431915 | B | 11526 | Wet bottom |
| 112_RayIs_3A9_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431916 | B | 705600 | Wet bottom |
| 113_RayIs_3C9_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431917 | B | 139540 | Wet bottom |
| 114_RayIs_3D9_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431918 | B | 6796 | Wet bottom |
| 115_RayIs_3E9_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431919 | B | 76040 | Wet bottom |
| 116_SWCoast_3H9_L_UGT | South Coast, Vic | Southern | UGT | HiSeq 2000 | ERR1431920 | L | 2716 | None |
| 117_RayIs_3A10_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431921 | B | 14466 | None |
| 118_RayIs_3G10_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431922 | B | 239200 | Wet bottom |
| 119_Gipps_3D11_C_UGT | Gippsland, Vic | Southern | UGT | HiSeq 2000 | ERR1431923 | C | 1046 | UGT abnormality |
| 120_Gipps_3A12_M_UGT | Gippsland, Vic | Southern | UGT | HiSeq 2000 | ERR1431924 | M | 2714 | Wet bottom |
| 121_MoPen_3F12_B_UGT | Mornington Peninsula, Vic | Southern | UGT | HiSeq 2000 | ERR1431925 | B | 9000 | None |
| 122_SWCoast_4B5_B_UGT | South Coast, Vic | Southern | UGT | HiSeq 2000 | ERR1431926 | B | 127180 | UGT abnormality, wet bottom |
| 123_MoPen_4G5_C_UGT | Mornington Peninsula, Vic | Southern | UGT | HiSeq 2000 | ERR1431927 | C | 22060 | None |
| 124_Gipps_4E6_C_UGT | Gippsland, Vic | Southern | UGT | HiSeq 2000 | ERR1431928 | C | 3062 | None |
| 125_Gipps_4H6_F_UGT | Gippsland, Vic | Southern | UGT | HiSeq 2000 | ERR1431929 | F | 4108 | UGT abnormality |
| 126_MoPen_4G8_C_UGT | Mornington Peninsula, Vic | Southern | UGT | HiSeq 2000 | ERR1431930 | C | 3480 | UGT abnormality |
| 127_Gipps_4C10_C_UGT | Gippsland, Vic | Southern | UGT | HiSeq 2000 | ERR1431931 | C | 85020 | None |
| 128_RayIs_5F5_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431932 | B | 12290 | None |
| 129_RayIs_5H6_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431933 | B | 181680 | None |
| 130_RayIs_7B7_B_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431934 | B | 88700 | UGT abnormality |

| Sample ID | Region | Population | Infection site | Platform | Accession | *ompA* type [#] | Genome copies/µL ^ | Clinical Signs |
|---|---|---|---|---|---|---|---|---|
| 131_WestMelb_7B8_C_UGT | Inner West, Vic | Southern | UGT | HiSeq 2000 | ERR1431935 | C | 516 | UGT abnormality, wet bottom |
| 132_Vic_7E9_B_UGT | Victoria | Southern | UGT | HiSeq 2000 | ERR1431936 | B | 2826 | ND |
| 133_SWCoast_7H12_L_UGT | South Coast, Vic | Southern | UGT | HiSeq 2000 | ERR1431937 | L | 3116 | None |
| 134_MoPen_15B5_C_UGT | Mornington Peninsula, Vic | Southern | UGT | HiSeq 2000 | ERR1431938 | C | 370200 | UGT abnormality |
| 135_Gipps_15C7_C_UGT | Gippsland, Vic | Southern | UGT | HiSeq 2000 | ERR1431939 | C | 83 | UGT abnormality, wet bottom |
| 136_FrIs_15H9_N_UGT | French Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431940 | N | 876 | Wet bottom |
| 137_FrIs_16H3_N_UGT | French Island, Vic | Southern | UGT | HiSeq 2000 | ERR1431941 | N | 142 | Wet bottom |
| 138_SWCoast_18H3_L_UGT | South Coast, Vic | Southern | UGT | HiSeq 2000 | ERR1431942 | L | 378 | None |
| 139_SWCoast_18C7_L_UGT | South Coast, Vic | Southern | UGT | HiSeq 2000 | ERR1431943 | L | 510 | None |
| 140_SWCoast_18D7_L_UGT | South Coast, Vic | Southern | UGT | HiSeq 2000 | ERR1431944 | L | 195 | None |
| 141_SWCoast_18C8_L_UGT | South Coast, Vic | Southern | UGT | HiSeq 2000 | ERR1431945 | L | 263 | None |

[#] *ompA* type determined by Sanger di-deoxy sequencing

^ Genome copies/µL determined with qPCR

[*] Urogenital abnormalities assessed via ultrasound or gross pathology at necropsy

[†] Wet bottom assessed in the field or at necropsy using previously described methods.

ND – Not determined, or information unavailable

**Table 7.2.** Summary of samples utilised in the study in relation to geographical region, body site sampled, and available record of clinical disease.

| | All samples | | | Clinical disease recorded (present/total) | | |
|---|---|---|---|---|---|---|
| **Swab samples** | **Conjunctival** | **Urogenital** | **Total** | **Conjunctival** | **Urogenital** | **Total** |
| Victoria | 1 | 47 | 48 | 0/1 | 26/46 | 26/47 |
| South Australia | 6 | 8 | 14 | 0/6 | 1/6 | 1/12 |
| **Southern** | 7 | 55 | 62 | 0/7 | 27/52 | 27/59 |
| New South Wales | 7 | 14 | 21 | 7/7 | 4/13 | 11/20 |
| Queensland | 7 | 7 | 14 | 5/5 | 5/5 | 10/10 |
| **Northern** | 18 * | 30 * | 48 * | 12/12 | 10/20 * | 22/32 * |
| **Total** | 25 | 85 | 110 | 12/19 | 37/72 | 49/91 |
| **Genomes analysed ^** | **Conjunctival** | **Urogenital** | **Total** | **Conjunctival** | **Urogenital** | **Total** |
| Victoria | 1 | 23 | 24 | 0/1 | 14/23 | 14/24 |
| South Australia | 1 | 2 | 3 | 0/1 | 1/2 | 1/3 |
| **Southern** | 2 | 25 | 27 | 0/2 | 15/25 | 15/27 |
| New South Wales | 7 | 6 | 13 | 7/7 | 2/6 | 9/13 |
| Queensland | 6 | 8 | 14 | 6/6 | 6/6 | 12/12 |
| **Northern** | 14 * | 19 * | 33 * | 13/13 | 9/13 * | 22/25 * |
| **Total** | 16 | 44 | 60 | 13/15 | 24/38 | 37/53 |

* Differences between state breakdowns and total sample numbers represent northern koalas where no clear state of origin was recorded, but population (northern) was known

^ Three published draft genomes were included in genome analysis, all of which have clinical signs of disease recorded

**Table 7.3.** Multilocus sequence typing result of *C. pecorum* genomes. Allele types and sequence types (STs) were based on sequences deposited at 'pubmlst.org/Chlamydiales' (Pannekoek *et al.* 2008; Jolley & Maiden 2010). Novel sequence types (NST) were denoted where no match to published sequence type occurred.

| Sample | ST | *enoA* | *fumC* | *gatA* | *gidA* | *hemN* | *hflX* | *oppA* |
|---|---|---|---|---|---|---|---|---|
| 101_Gipps_2G12_F_UGT | NST 1 | ~20 | 9 | ~30 | 23 | 8 | 10 | 12^ |
| 103_Rayls_3F3_B_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12^ |
| 104_WestVic_3F4_B_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12^ |
| 108_Rayls_3E7_B_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12^ |
| 109_Rayls_3C8_B_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12^ |
| 110_Rayls_3D8_B_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12^ |
| 112_Rayls_3A9_B_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12^ |
| 113_Rayls_3C9_B_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12^ |
| 114_Rayls_3D9_B_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12^ |
| 115_Rayls_3E9_B_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12^ |
| 116_SWCoast_3H9_L_UGT | NST 3 | 20 | 20 | ~27 | 23 | 8 | 10 | 21^ |
| 117_Rayls_3A10_B_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12^ |
| 118_Rayls_3G10_B_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12^ |
| 121_MoPen_3F12_B_UGT | NST 4 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12^ |
| 122_SWCoast_4B5_B_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12^ |
| 123_MoPen_4G5_C_UGT | NST 5 | 18 | 9 | 27 | ~23 | 8 | 10 | 12^ |
| 124_Gipps_4E6_C_UGT | NST 6 | ~30 | 9 | ~30 | 23 | 8 | 10 | 12^ |
| 125_Gipps_4H6_F_UGT | NST 7 | ~30 | 9 | 27 | 23 | 8 | 10 | 12^ |
| 127_Gipps_4C10_C_UGT | NST 8 | ~18 | 9 | 27 | 23 | 8 | 10 | 12^ |
| 129_Rayls_5H6_B_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12 |
| 130_Rayls_7B7_B_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12^ |
| 134_MoPen_15B5_C_UGT | NST 9 | ~18 | 9 | 27 | 23 | 8 | 10 | 12^ |
| 201_Belvedere_S1_UGT | NST 10 | 23 | 9 | 27 | 31 | 8 | 10 | 12 |
| 203_Bella_290513_UGT | NST 11 | 23 | ~9 | 21 | 23 | 8 | 10 | 12* |
| 204_R1_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12* |
| 207_Savannah_S2_UGT | NST 10 | 23 | 9 | 27 | 31 | 8 | 10 | 12 |
| 208_Buddy_S1_Ocular | NST 12 | 23 | 9 | 21 | 23 | 8 | 10 | 12* |
| DBDeUG | 69 | 23 | 9 | 27 | 23 | 8 | 10 | 12 |
| IpTaLE | NST 12 | 23 | 9 | 21 | 23 | 8 | 10 | 12 |
| Mc_Marsbar | 69 | 23 | 9 | 27 | 23 | 8 | 10 | 12 |
| USC_Eleena_UGT | NST 10 | 23 | 9 | 27 | 31 | 8 | 10 | 12* |
| USC_Gun_koa1_UGT | 73 | 23 | 9 | 30 | 23 | 8 | 10 | 12 |
| USC_Haz_Bo_Eye | NST 10 | 23 | 9 | 27 | 31 | 8 | 10 | 12 |
| USC_Max_S1_UGT | NST 10 | 23 | 9 | 27 | 31 | 8 | 10 | 12* |
| USC_NSW_Adelle_LE | NST 12 | 23 | 9 | 21 | 23 | 8 | 10 | 12* |
| USC_NSW_Chingee_Eye | 73 | 23 | 9 | 30 | 23 | 8 | 10 | 12* |
| USC_NSW_Chingee_UGT | 73 | 23 | 9 | 30 | 23 | 8 | 10 | 12* |
| USC_NSW_Elmo_Eye | NST 13 | 23 | ~9 | 21 | ~23 | 8 | 10 | 12* |
| USC_NSW_Knox_Eye | NST 12 | 23 | 9 | 21 | 23 | 8 | 10 | 12* |
| USC_PM_11_UGT2 | NST 14 | 23 | 9 | ~30 | 23 | 8 | 10 | 24 |
| USC_PM_13_UGT2 | NST 15 | 23 | 9 | ~30 | 23 | 8 | 10 | 12 |
| USC_PM_17_UGT2 | NST 15 | 23 | 9 | ~30 | 23 | 8 | 10 | 12* |
| USC_PM_3_UGT | NST 15 | 23 | 9 | ~30 | 23 | 8 | 10 | 12* |
| USC_Posh_S1_Eye | NST 16 | 23 | 9 | 31 | 23 | 8 | 10 | 12* |
| USC_QLD_Bobby_UGT | NST 10 | 23 | 9 | 27 | 31 | 8 | 10 | 12 |
| USC_QLD_Chestnut_LE | NST 17 | 23 | 9 | 27 | ~23 | 8 | 10 | 12* |
| USC_QLD_Chestnut_RE | NST 17 | 23 | 9 | 27 | ~23 | 8 | 10 | 12* |
| USC_QLD_Chestnut_UGT | NST 17 | 23 | 9 | 27 | ~23 | 8 | 10 | 12* |

| Sample | ST | *enoA* | *fumC* | *gatA* | *gidA* | *hemN* | *hflX* | *oppA* |
|---|---|---|---|---|---|---|---|---|
| USC_QLD_Helen_UGT | NST 10 | 23 | 9 | 27 | 31 | 8 | 10 | 12* |
| USC_QLD_Jasper_LE | NST 10 | 23 | 9 | 27 | 31 | 8 | 10 | 12* |
| USC_QLD_Talle_UGT | 69 | 23 | 9 | 27 | 23 | 8 | 10 | 12 |
| USC_QLD_Travis_LE | NST 18 | 23 | 9 | 27 | ~23 | 8 | 10 | 12* |
| USC_QLD_Travis_UGT | NST 17 | 23 | 9 | 27 | ~23 | 8 | 10 | 12 |
| USC_Rayls_7-14_Eye | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12* |
| USC_SA_12-220_LE | 48 | 18 | 9 | 20 | 22 | 17 | 20 | 12* |
| USC_SA_13-9_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12* |
| USC_SA_K2_UGT | NST 2 | ~30 | 20 | ~30 | 23 | 8 | 10 | 12 |
| USC_Ted_Hu_UGT | NST 11 | 23 | ~9 | 21 | 23 | 8 | 10 | 12 |
| USC_Tya_Butler_LE | NST 12 | 23 | 9 | 21 | 23 | 8 | 10 | 12 |
| USC_Tya_Mavis_LE | NST 12 | 23 | 9 | 21 | 23 | 8 | 10 | 12* |

^ Confirmed sequence with Sanger sequencing

* Incomplete sequence, nearest allele type listed

~ Single or poly-nucleotide differences to closest allele type listed

Novel sequence type nucleotide changes as listed, base number in square brackets:

NST 1 - *enoA*: [318] T→C *gatA*: [183] C→T; [381] G→A.

NST 2 - *enoA*: [180] G→A *gatA*: [183] C→T; [381] G→A.

NST 3 - *gatA*: [111] A→G.

NST 4 - *enoA*: [180] G→A; [322] C→A *gatA*: [93] G→C; [183] C→T; [381] G→A.

NST 5 - *gidA*: [127] C→T.

NST 6 - *enoA*: [180] G→A *gatA*: [183] C→T; [381] G→A.

NST 7 - *enoA*: [180] G→A.

NST 8 - *enoA*: [241] C→T.

NST 9 - *enoA*: [51] G→A.

NST 11 - *fumC*: [302] A→G.

NST 13 - *fumC*: [85] C→T *gidA*: [432] C→T; [450] G→A.

NST 14 - *gatA*: [411] C→T.

NST 15 - *gatA*: [411] C→T.

NST 17 - *gidA*: [129] C→T.

NST 18 - *gidA*: [113] G→A; [150] C→A.

**Table 7.4.** Details of regions of interest from *C. pecorum* genomes, including truncated and full-length coding Phospholipase D (PLD) genes, *ompA* genotypes, coding tandem repeats (CTR) in ORF663, and IncA repeat motifs. Mixed infections are highlighted by the presence of shortened *ompA* contigs (denoted as minor genotypes).

| Sample | PLDs (complete/total) | *ompA* genotype | | | ORF663 CTR | IncA motifs (APA/total) |
|---|---|---|---|---|---|---|
| | | Major | Minor | Minor length (bp) | | |
| 101_Gipps_2G12_F_UGT | 2/5 | F | - | - | 6 | 0/8 |
| 103_RayIs_3F3_B_UGT | 3/7 | B | - | - | - | 3/6 |
| 104_WestVic_3F4_B_UGT | 1/7 | B | - | - | 6 | 1/4 |
| 108_RayIs_3E7_B_UGT | 3/7 | B | - | - | - | 3/6 |
| 109_RayIs_3C8_B_UGT | 6/7 | B | - | - | 6 | 3/6 |
| 110_RayIs_3D8_B_UGT | 4/7 | B | - | - | - | 3/6 |
| 112_RayIs_3A9_B_UGT | 2/7 | B | - | - | 6 | 3/6 |
| 113_RayIs_3C9_B_UGT | 3/7 | B | - | - | 6 | 6/9 |
| 114_RayIs_3D9_B_UGT | 2/7 | B | - | - | 7 | 3/6 |
| 115_RayIs_3E9_B_UGT | 4/7 | B | - | - | 6 | 3/6 |
| 116_SWCoast_3H9_L_UGT | 1/5 | L | - | - | 2 | 6/7 |
| 117_RayIs_3A10_B_UGT | 1/7 | B | - | - | - | 3/6 |
| 118_RayIs_3G10_B_UGT | 3/7 | B | - | - | - | 4/7 |
| 121_MoPen_3F12_B_UGT | 2/4 | B* | - | - | 5 | 2/6 |
| 122_SWCoast_4B5_B_UGT | 5/7 | B | - | - | 6 | 4/7 |
| 123_MoPen_4G5_C_UGT | 1/4 | C* | B | 224† | 8 | 0/5 |
| 124_Gipps_4E6_C_UGT | 2/6 | C | - | - | 6 | 7/8 |
| 125_Gipps_4H6_F_UGT | 3/6 | F | - | - | 6 | 3/6 |
| 127_Gipps_4C10_C_UGT | 3/5 | C | - | - | 10 | 0/9 |
| 129_RayIs_5H6_B_UGT | 4/7 | B | - | - | 6 | 3/6 |
| 130_RayIs_7B7_B_UGT | 3/7 | B | - | - | 6 | 3/6 |
| 134_MoPen_15B5_C_UGT | 2/5 | C | B | 552† | 10 | 0/6 |
| 201_Belvedere_S1_UGT | 0/5 | F | - | - | - | 8/9 |
| 203_Bella_290513_UGT | 1/5 | F | - | - | 4 | 8/10 |
| 204_R1_UGT | 5/7 | B | - | - | 6 | 4/8 |
| 207_Savannah_S2_UGT | 2/5 | F | G* | 260† | - | 7/8 |
| 208_Buddy_S1_Ocular | 2/6 | F | A | 836† | 20 | 12/13 |
| DBDeUG | 1/3 | F | - | - | 30[&] | 10/11 |
| IpTaLE | 0/2 | A | - | - | 18[&] | 5/5 |
| Mc_Marsbar | 1/3 | G | - | - | 11[&] | 3/4 |
| USC_Eleena_UGT | 3/5 | F | G | 617 | 15 | 5/6 |
| USC_Gun_koa1_UGT | 2/4 | A | - | - | - | 6/7 |
| USC_Haz_Bo_Eye | 1/6 | F | - | - | 7 | 9/10 |
| USC_Max_S1_UGT | 0/5 | F | G | 642 | - | 9/10 |
| USC_NSW_Adelle_LE | 1/6 | H | L* | 226† | 3 | 1/2 |
| USC_NSW_Chingee_Eye | 1/5 | F | B** | 266† | 3 | 5/6 |
| USC_NSW_Chingee_UGT | 1/5 | F | - | - | 5 | 4/4 |
| USC_NSW_Elmo_Eye | 0/6 | F | A | 668 | 9 | 6/7 |
| USC_NSW_Knox_Eye | 1/6 | H | F | 674 | 12 | 2/3 |
| USC_PM_11_UGT2 | 3/7 | F | B** | 660† | 7 | 6/7 |
| USC_PM_13_UGT2 | 4/7 | F | G** | 657 | 7 | 5/6 |
| USC_PM_17_UGT2 | 3/6 | F | G** | 438† | 13 | 6/7 |
| USC_PM_3_UGT | 1/5 | F | B** | 606† | 6 | 5/6 |
| USC_Posh_S1_Eye | 0/5 | G* | B*, F | 356,393 | - | 9/10 |
| USC_QLD_Bobby_UGT | 2/6 | F | - | - | 7 | 9/10 |
| USC_QLD_Chestnut_LE | 0/5 | G | B** | 280† | - | 5/5 |
| USC_QLD_Chestnut_RE | 2/7 | G | B* | 454† | 6 | 5/6 |
| USC_QLD_Chestnut_UGT | 2/7 | G | L** | 732† | 12 | 5/6 |
| USC_QLD_Helen_UGT | 1/5 | F | G** | 535† | 11 | 5/6 |
| USC_QLD_Jasper_LE | 1/7 | F | B**, K** | 460†, 265† | - | 2/3 |
| USC_QLD_Talle_UGT | 0/7 | G | B* | 454† | 14 | 6/7 |
| USC_QLD_Travis_LE | 1/5 | A | - | - | 7 | 10/11 |
| USC_QLD_Travis_UGT | 2/6 | A | B**, F | 466, 275† | - | 1/2 |

| Sample | PLDs (complete/total) | *ompA* genotype | | | ORF663 CTR | IncA motifs (APA/total) |
|---|---|---|---|---|---|---|
| | | Major | Minor | Minor length (bp) | | |
| USC_Rayls_7-14_Eye | 0/5 | B* | F, G, J^, A^ | 1104, 687, 471^, 201^ | 6 | 3/6 |
| USC_SA_12-220_LE | 2/5 | B** | - | - | - | 9/10 |
| USC_SA_13-9_UGT | 2/7 | B | - | - | 7 | 7/10 |
| USC_SA_K2_UGT | 2/4 | B | G* | 635 | - | 6/9 |
| USC_Ted_Hu_UGT | 2/6 | F | - | - | 16 | 10/12 |
| USC_Tya_Butler_LE | 1/6 | F | A* | 287† | 22 | 6/7 |
| USC_Tya_Mavis_LE | 1/6 | F | - | - | - | 8/9 |

[*] Less than 99% nucleotide identity, the letter used represents the closest related koala associated *ompA* genotype

[**] Less than 95% nucleotide identity, the letter used represents the closest related koala associated *ompA* genotype

[^] Regions do not overlap and may be part of the same strain

[†] Coverage depth of less than 5 reads per base across the contig

[&] ORF663 CTR number from (Yousef Mohamad *et al.* 2014)

**Table 7.5.** Measurements of strain diversity based on nucleotide sequence alignments of individual genes from *C. pecorum* genomes. Gene clusters were produced with Roary (Page *et al.* 2015) and aligned with MAFFT (Katoh *et al.* 2002). Diversity metrics were calculated within the R package 'APE' (Paradis *et al.* 2004).

| Gene cluster | Product | Haplotypes | Segregation Sites | Nucleotide Diversity | Tajima's D[#] |
|---|---|---|---|---|---|
| *ompA* | Major outer membrane protein | 26 | 312 | 0.091 | 2.10* |
| *gnd* | 6-phosphogluconate dehydrogenase, decarboxylating | 8 | 252 | 0.086 | 4.66** |
| *tyrS* | Tyrosine-tRNA ligase | 9 | 660 | 0.077 | -2.24* |
| *pmp15/E* | Polymorphic membrane protein 15/E | 25 | 839 | 0.048 | -0.89 |
| group_36 | hypothetical protein | 19 | 229 | 0.044 | 0.37 |
| *pmp13/G* | Polymorphic membrane protein 13/G | 31 | 464 | 0.044 | 1.08 |
| group_26 | hypothetical protein | 19 | 98 | 0.042 | 0.06 |
| group_30 | hypothetical protein | 15 | 205 | 0.036 | -0.40 |
| *pmp16/E* | Polymorphic membrane protein 16/E | 31 | 876 | 0.034 | -1.87 |
| group_44 | hypothetical protein | 8 | 48 | 0.032 | 0.80 |
| *incA* | Inclusion membrane protein A | 34 | 353 | 0.028 | -2.41* |
| *incC* | Inclusion membrane protein C | 10 | 129 | 0.027 | -1.50 |
| group_60 | hypothetical protein | 24 | 300 | 0.027 | 0.36 |
| *pmp14/H* | Polymorphic membrane protein 14/H | 21 | 315 | 0.027 | 0.69 |
| group_228 | hypothetical protein | 8 | 65 | 0.027 | 2.50* |
| *lepA* | Elongation factor 4 | 11 | 91 | 0.022 | 3.66** |
| *guaB* | Inosine-5'-monophosphate dehydrogenase | 12 | 54 | 0.021 | 3.27** |

[#] Tajima's D is a measure of evolution. A significant D value less than 0 suggests rare alleles occur at a higher than expected frequency, and a significant D value greater than 0 suggests rare alleles occur at a lower than expected frequency (Tajima 1989)

[*] Significant Tajima's D ($P < 0.05$)

[**] Significant Tajima's D ($P < 0.01$)

**Table 7.6.** Proteins with significant (P < 0.01) positive codon selection based on the likelihood ratio test comparing log likelihoods from codeml model tests: null model (M0), nearly neutral (M1a) and positive selection (M2a) (Yang 2007; Jeffares *et al.* 2015).

| Gene cluster | Product * | Omega (dN/dS) | Model log likelihood | | | Null LRT # (M0 vs M1a) | Positive selection LRT ^ (M1a vs M2a) |
|---|---|---|---|---|---|---|---|
| | | | M0 | M1a | M2a | | |
| *menH*_1 | 2-succinyl-6-hydroxy-2, 4-cyclohexadiene-1-carboxylate synthase | 7.62 | -1325.2 | -1310.6 | -1194.6 | 29.09 | 232.09 |
| group_233 | **Putative transmembrane protein (*C. psittaci* 28.9%)** | 0.57 | -7085.5 | -7029.6 | -6961.3 | 111.61 | 136.61 |
| group_241 | hypothetical protein | 0.87 | -821.9 | -772.4 | -705.9 | 99.03 | 133.09 |
| *pth* | Peptidyl-tRNA hydrolase | 0.47 | -1060.3 | -1020.8 | -957.7 | 78.98 | 126.18 |
| *pkn1* | Serine/threonine-protein kinase pkn1 | 0.41 | -2812.0 | -2783.7 | -2741.3 | 56.68 | 84.7 |
| *def* | Peptide deformylase | 0.30 | -966.2 | -944.2 | -907.8 | 44.04 | 72.71 |
| group_36 | hypothetical protein | 1.14 | -3048.1 | -3022.3 | -2991.1 | 51.58 | 62.37 |
| group_376 | **Putative lysine decarboxylase family protein (*C. psittaci* 60.7%)** | 0.43 | -3174.9 | -3156.4 | -3125.8 | 37.06 | 61.25 |
| *xcpQ* | Type II secretion system protein D | 0.18 | -4355.0 | -4329.0 | -4300.1 | 52.1 | 57.66 |
| group_53 | hypothetical protein | 0.51 | -1658.7 | -1631.6 | -1605.5 | 54.19 | 52.33 |
| group_60 | hypothetical protein | 0.28 | -6011.0 | -5871.2 | -5848.1 | 279.6 | 46.18 |
| group_86 | **Putative inner membrane protein (*C. abortus* 48.4%)** | 0.57 | -1501.4 | -1488.1 | -1466.2 | 26.57 | 43.83 |
| *dxs* | 1-deoxy-D-xylulose-5-phosphate synthase | 0.32 | -2968.1 | -2952.2 | -2931.5 | 31.78 | 41.36 |
| *ribF* | Riboflavin biosynthesis protein RibF | 0.59 | -1317.3 | -1308.9 | -1289.1 | 16.7 | 39.64 |
| *ompA* | Major outer membrane protein | 0.22 | -4383.1 | -4181.4 | -4161.7 | 403.32 | 39.46 |
| group_30 | hypothetical protein | 0.76 | -2903.2 | -2854.7 | -2835.5 | 96.95 | 38.39 |
| *pmp21/D* | Polymorphic membrane protein 21/D | 0.31 | -6280.2 | -6262.1 | -6245.2 | 36.14 | 33.75 |
| group_38 | hypothetical protein | 0.63 | -1600.6 | -1584.0 | -1567.8 | 33.33 | 32.27 |
| *pheT* | Phenylalanine--tRNA ligase beta subunit | 0.40 | -3432.9 | -3426.5 | -3411.1 | 12.78 | 30.87 |
| group_193 | **Thymidylate kinase (*C. pecorum* 70.1%)** | 0.17 | -2743.4 | -2720.5 | -2705.2 | 45.63 | 30.7 |
| group_102 | **Phage tail fiber protein (*C. pecorum* 61.4%)** | 0.52 | -2128.5 | -2115.8 | -2101.0 | 25.47 | 29.54 |
| group_959 | **Effector from type III secretion system family protein (*C. psittaci* 45.9%)** | 1.01 | -3547.0 | -3480.4 | -3467.8 | 133.19 | 25.23 |
| *pyrH* | Uridylate kinase | 0.18 | -1186.9 | -1174.2 | -1162.5 | 25.33 | 23.29 |
| *pfp*_2 | Pyrophosphate--fructose 6-phosphate 1-phosphotransferase | 0.43 | -2441.0 | -2434.0 | -2422.7 | 14 | 22.57 |

| Gene cluster | Product * | Omega (dN/dS) | Model log likelihood | | | Null LRT # (M0 vs M1a) | Positive selection LRT ^ (M1a vs M2a) |
|---|---|---|---|---|---|---|---|
| | | | M0 | M1a | M2a | | |
| group_888 | **Competence protein family (*C. pecorum* 98.6%)** | 0.49 | -2214.6 | -2209.0 | -2198.7 | 11.14 | 20.6 |
| *guaA* | GMP synthase [glutamine-hydrolyzing] | 0.13 | -2623.2 | -2600.6 | -2590.4 | 45.24 | 20.23 |
| group_871 | **Lipoprotein (*C. pecorum* 99.4%)** | 0.31 | -4208.6 | -4197.8 | -4187.7 | 21.70 | 20.17 |
| group_28 | hypothetical protein | 0.39 | -2026.5 | -1983.2 | -1973.7 | 86.62 | 18.98 |
| *lplT* | Lysophospholipid transporter LplT | 0.20 | -2346.8 | -2339.0 | -2329.6 | 15.49 | 18.84 |
| group_220 | hypothetical protein | 0.20 | -1896.1 | -1881.2 | -1871.8 | 29.83 | 18.68 |
| *pmp14/H* | Polymorphic membrane protein 14/H | 0.22 | -6690.8 | -6616.5 | -6607.3 | 148.56 | 18.44 |
| *proS* | Proline--tRNA ligase | 0.27 | -2606.0 | -2598.1 | -2589.8 | 15.84 | 16.58 |
| *gatB* | Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferasesubunit B | 0.36 | -2190.7 | -2184.3 | -2176.0 | 12.88 | 16.49 |
| *sctD* | Type III secretion system structural protein | 0.11 | -4002.4 | -3983.9 | -3975.8 | 37.04 | 16.26 |
| *mnmG* | tRNA uridine 5-carboxymethylaminomethyl modification enzyme | 0.22 | -2650.1 | -2644.6 | -2636.9 | 10.84 | 15.44 |
| *leuS* | Leucine-tRNA ligase | 0.16 | -3725.3 | -3713.4 | -3705.9 | 23.85 | 14.86 |
| group_938 | **Nucleoside-diphosphate-sugar epimerase (*C. pneumoniae* 64.6%)** | 0.23 | -1983.3 | -1973.3 | -1966.6 | 19.91 | 13.41 |
| group_54 | hypothetical protein | 0.79 | -903.6 | -897.8 | -891.8 | 11.55 | 12.1 |
| *mqnE* | Aminodeoxyfutalosine synthase | 0.31 | -1702.3 | -1696.6 | -1690.9 | 11.45 | 11.43 |
| *infB* | Translation initiation factor IF-2 | 0.21 | -3329.1 | -3322.3 | -3316.6 | 13.62 | 11.34 |
| group_947 | **Secretion system effector C (SseC) like family protein (*C. psittaci*, 55.7%)** | 0.27 | -2089.9 | -2084.0 | -2078.9 | 11.80 | 10.17 |
| *sodA* | Superoxide dismutase [Mn/Fe] | 0.23 | -962.2 | -954.9 | -950.3 | 14.53 | 9.28 |

[*] Product determined by Prokka (Seemann 2014), or for hypothetical proteins in **bold**, the closest match from the *Chlamydiaceae* family derived from a BLAST search of the UniProtKB database (The UniProt Consortium 2017). Gene clusters listed as 'hypothetical proteins' were also uncharacterised *C. pecorum* proteins in the UniProtKB database.

[#] To validated the null model (M0), a likelihood ratio test between M0 and M1a/M2a used a critical $\chi^2$ value of 10.83 with 1 degree of freedom and an alpha of 0.001.

[^] Significant likelihood ratio tests between models M1a and M2a, with a critical $\chi^2$ value of 9.21, 2 degrees of freedom, and an alpha of 0.01.

**Table 7.7.** Measurements of tree similarity based on normalised Robinson-Foulds metric (Robinson & Foulds 1981).

| | MLST | Plasmid | SNP genome^ | *ompA* | Gubbins* |
|---|---|---|---|---|---|
| **Full genome** | 0.82# | 0.82 | 0.20 | 0.83 | 0.38 |
| **MLST** | | 0.71 | 0.78 | 0.80 | 0.80 |
| **Plasmid** | | | 0.82 | 0.85 | 0.77 |
| **SNP genome** | | | | 0.83 | 0.38 |
| ***ompA*** | | | | | 0.84 |

# All values are between 0 and 1, with higher numbers representing an increased distance between trees.

^ Single nucleotide polymorphism (SNP) core genome produced by Snippy from assembled genomes

* Polymorphic genome regions after recombination blocks removed

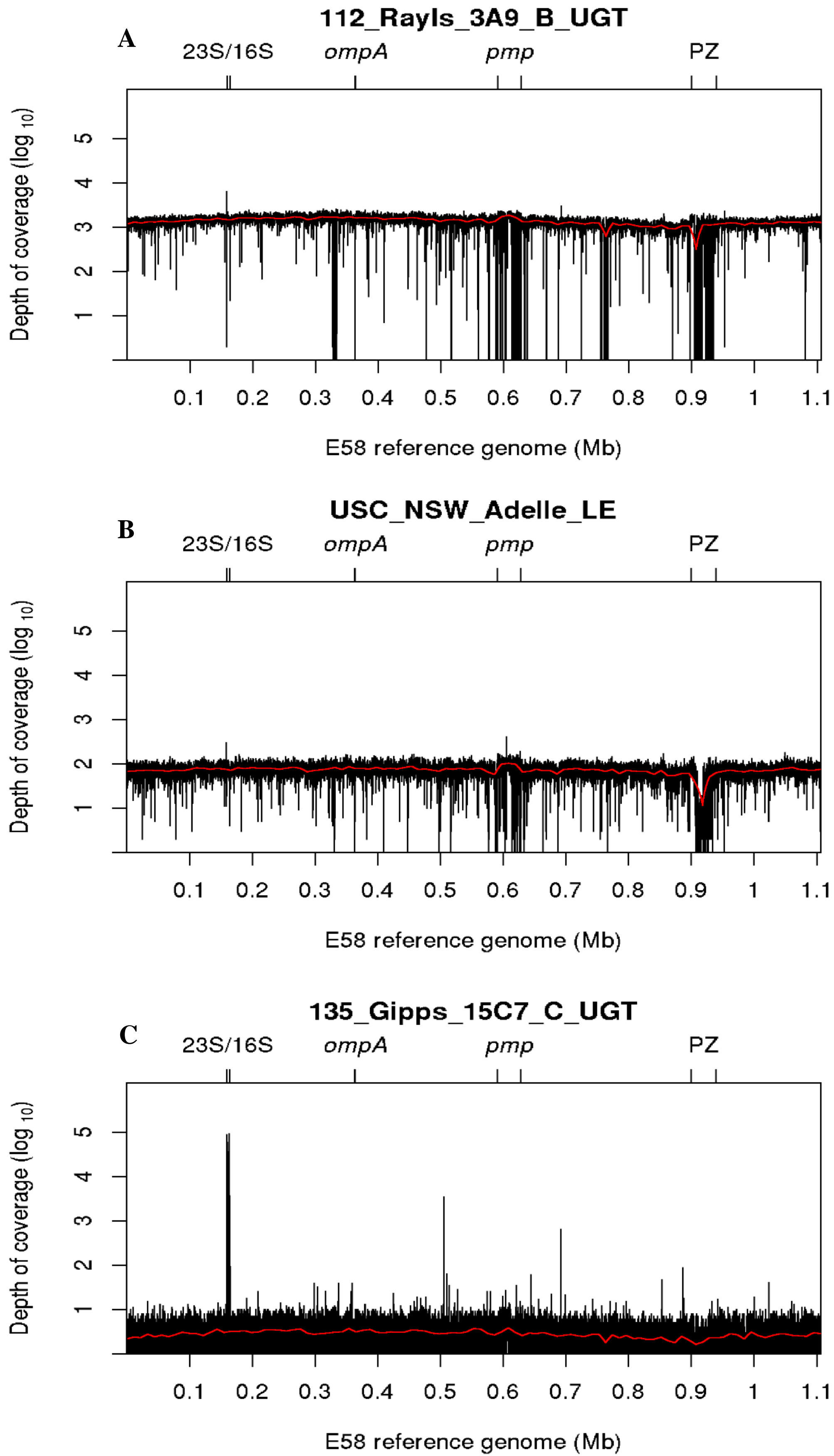

A
112_Rayls_3A9_B_UGT
23S/16S
ompA
pmp
PZ
Depth of coverage (log 10)
E58 reference genome (Mb)
B
USC_NSW_Adelle_LE
C
135_Gipps_15C7_C_UGT

**Figure 7.1.** Three example plots, illustrating the depth of coverage of short reads mapped to the *C. pecorum* type strain E58, as performed by 'bwa mem' (Li & Durbin 2010). The red line in each plot is a locally weighted scatterplot smoothing with a smoother span of 0.01. Genomic regions of interest are marked above each plot: 23S/16S ribosomal RNA coding region, *ompA*, polymorphic membrane proteins and the plasticity zone. Examples were chosen to demonstrate **(A)** high proportion genome coverage with high median depth, **(B)** high proportion genome coverage with sufficient median depth, and **(C)** low proportion genome coverage with poor median depth.

**Figure 7.2.** A phylogenetic tree representing the *C. pecorum* full genome alignment. A full genome alignment was generated with MAFFT (Katoh *et al.* 2002), and processed for poorly aligned regions with Gblocks (Castresana 2000). A phylogeny was produced using MrBayes (Huelsenbeck & Ronquist 2001), using the GTR+I+G model as determined in jModelTest2 (Darriba *et al.* 2012). Bayesian MCMC analysis utilised 4 chains, with a chain length of 1,000,000 and a relative burn in of 25% chain length. The scale bar represents the number of substitutions per site. Node labels represent posterior probability.

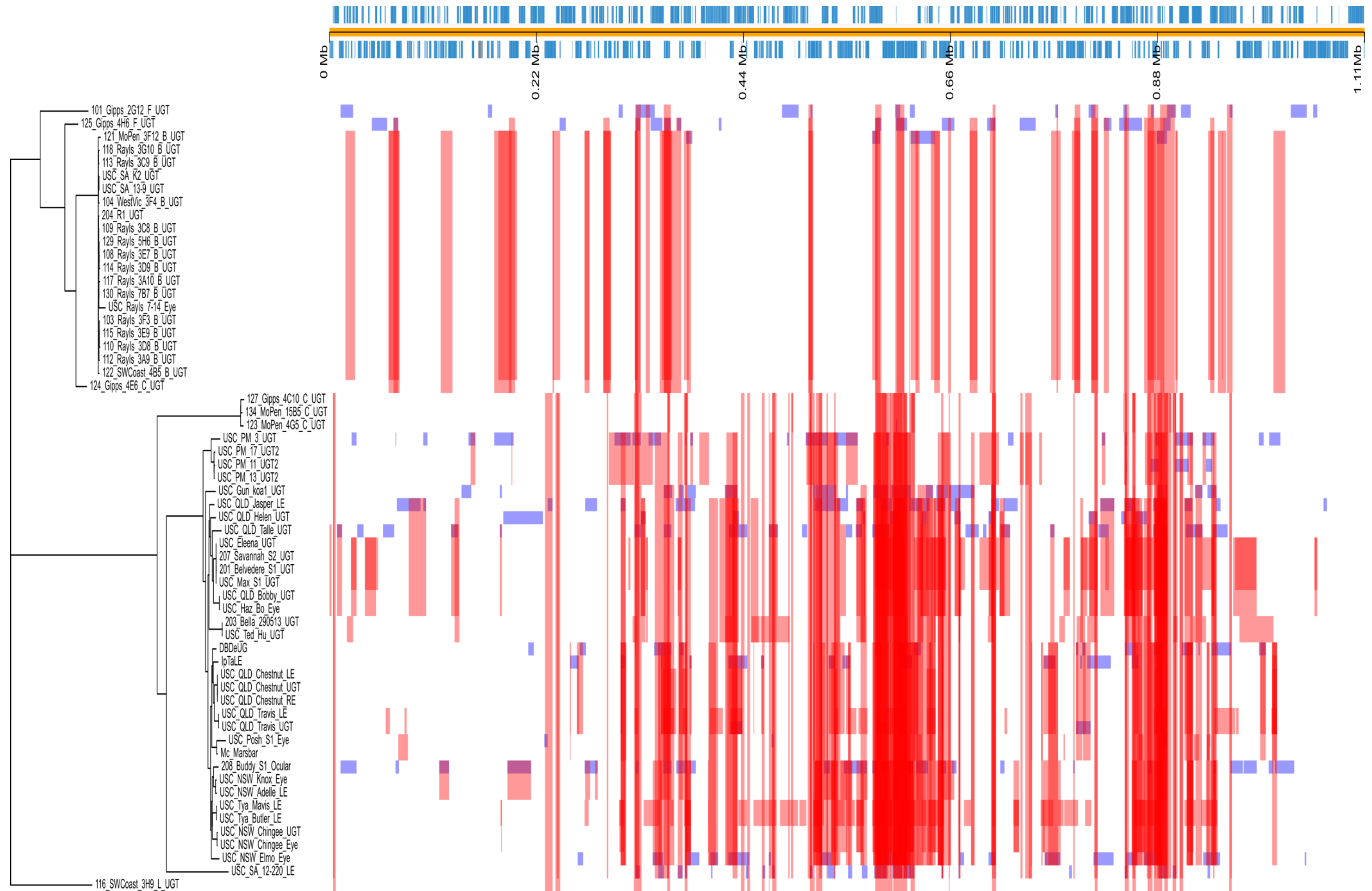

0 Mb
0.22 Mb
0.44 Mb
0.66 Mb
0.88 Mb
1.11Mb
101_Gipps_2G12_F_UGT
125_Gipps_4H6_F_UGT
121_MoPen_3F12_B_UGT
118_Rayls_3G10_B_UGT
113_Rayls_3C9_B_UGT
USC_SA_K2_UGT
USC_SA_13-9_UGT
104_WestVic_3F4_B_UGT
204_R1_UGT
109_Rayls_3C8_B_UGT
129_Rayls_5H6_B_UGT
108_Rayls_3E7_B_UGT
114_Rayls_3D9_B_UGT
117_Rayls_3A10_B_UGT
130_Rayls_7B7_B_UGT
USC_Rayls_7-14_Eye
103_Rayls_3F3_B_UGT
115_Rayls_3E9_B_UGT
110_Rayls_3D8_B_UGT
112_Rayls_3A9_B_UGT
122_SWCoast_4B5_B_UGT
124_Gipps_4E6_C_UGT
127_Gipps_4C10_C_UGT
134_MoPen_15B5_C_UGT
123_MoPen_4G5_C_UGT
USC_PM_3_UGT
USC_PM_17_UGT2
USC_PM_11_UGT2
USC_PM_13_UGT2
USC_Gun_koa1_UGT
USC_QLD_Jasper_LE
USC_QLD_Helen_UGT
USC_QLD_Talle_UGT
USC_Eleena_UGT
207_Savannah_S2_UGT
201_Belvedere_S1_UGT
USC_Max_S1_UGT
USC_QLD_Bobby_UGT
USC_Haz_Bo_Eye
203_Bella_290513_UGT
USC_Ted_Hu_UGT
DBDeUG
IpTaLE
USC_QLD_Chestnut_LE
USC_QLD_Chestnut_UGT
USC_QLD_Chestnut_RE
USC_QLD_Travis_LE
USC_QLD_Travis_UGT
USC_Posh_S1_Eye
Mc_Marsbar
208_Buddy_S1_Ocular
USC_NSW_Knox_Eye
USC_NSW_Adelle_LE
USC_Tya_Mavis_LE
USC_Tya_Butler_LE
USC_NSW_Chingee_UGT
USC_NSW_Chingee_Eye
USC_NSW_Elmo_Eye
USC_SA_12-220_LE
116_SWCoast_3H9_L_UGT

**Figure 7.3.** Whole genome recombination analysis output from Gubbins (Croucher *et al.* 2015). The tree represents the final tree output of Gubbins, which uses 5 iterations of RAxML (Stamatakis 2014) using an alignment of the polymorphic sites of the genomes with areas of recombination removed. The bar above the chart represents the genome of *C. pecorum* E58, with open reading frames in blue. In the chart area, red blocks are areas of genomic recombination present across samples, whilst blue blocks are areas of genomic recombination occurring only within a single sample.

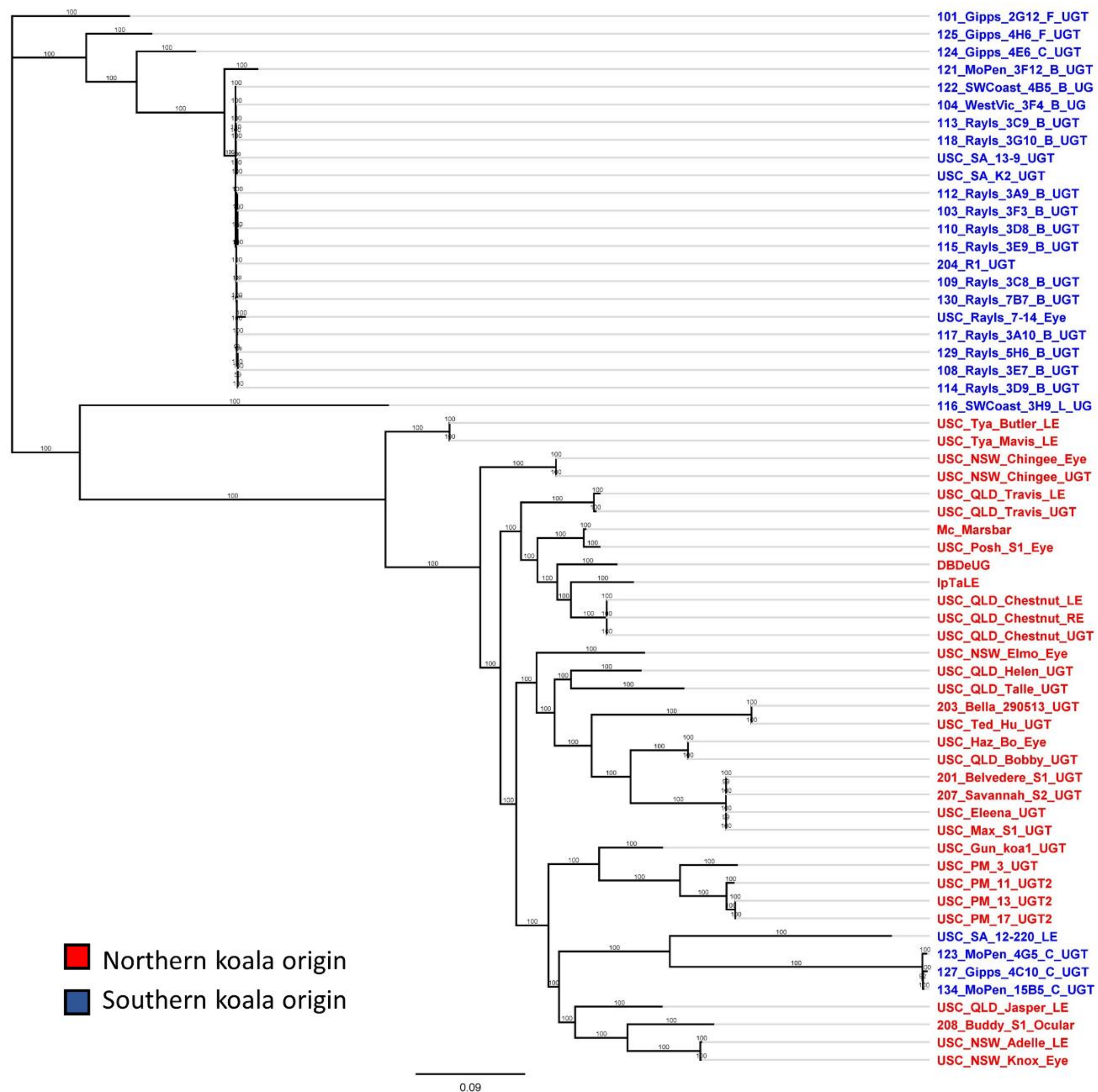



**Figure 7.4.** A phylogenetic tree representing the *C. pecorum* single nucleotide polymorphism (SNP) alignment. A SNP alignment was generated by Snippy, which compared genome contigs to the *C. pecorum* reference strain (E58). The core SNP phylogeny was produced using MrBayes (Huelsenbeck & Ronquist 2001), using the GTR+G model as determined in jModelTest2 (Darriba *et al.* 2012). Bayesian MCMC analysis utilised 4 chains, with a chain length of 1,000,000 and a relative burn in of 25% chain length. The scale bar represents the number of substitutions per site. Node labels represent posterior probability.

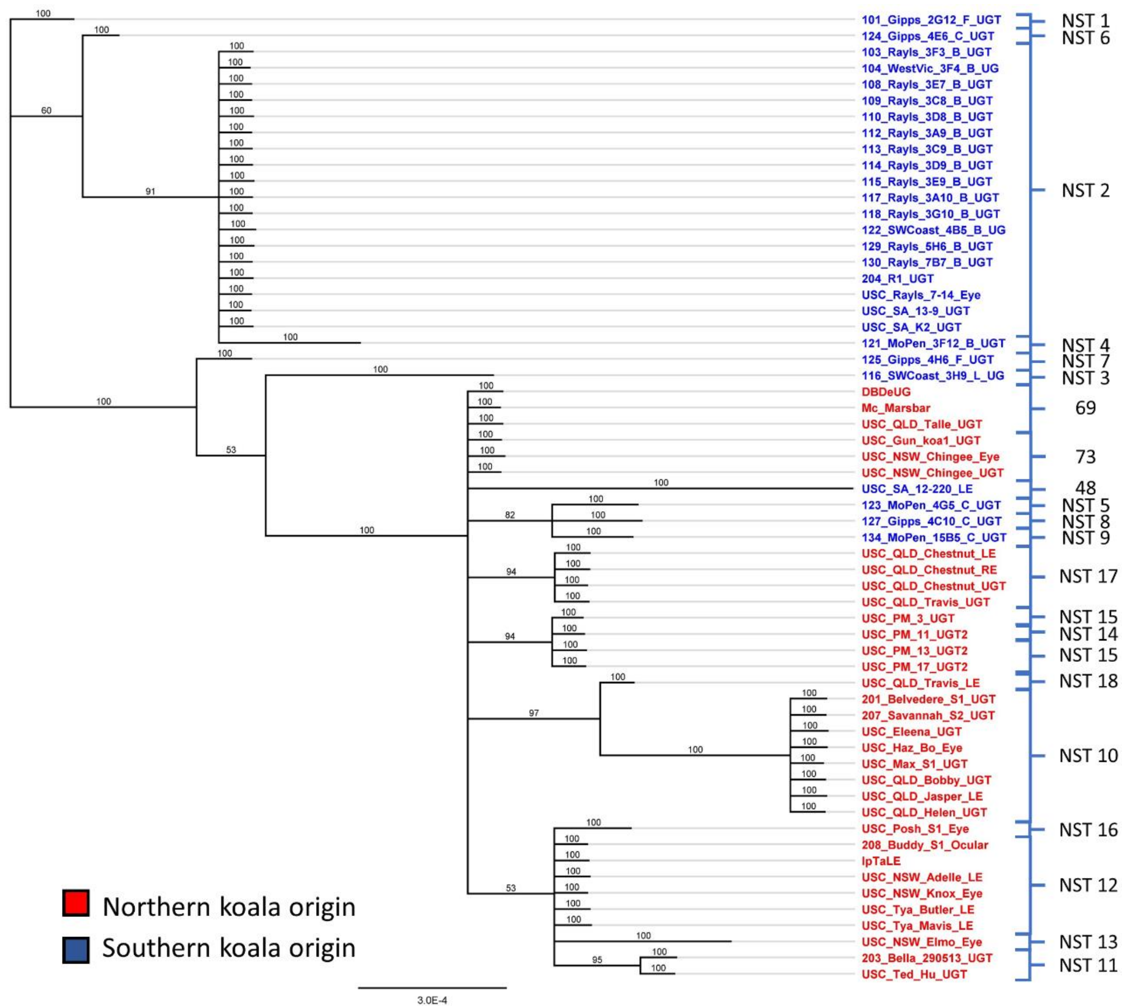



**Figure 7.5.** A phylogenetic tree representing a partial *C. pecorum* concatenated MLST alignment. An MLST alignment was generated with MAFFT (Katoh *et al.* 2002) using 6/7 *Chlamydiales* MLST alleles. *oppA* was not included in the alignment due to it being incomplete in some genomes. The phylogenetic tree was produced using MrBayes (Huelsenbeck & Ronquist 2001), utilising the HKY+I model as determined in jModelTest2 (Darriba *et al.* 2012). Bayesian MCMC analysis utilised 4 chains, with a chain length of 1,000,000 and a relative burn in of 25% chain length. The scale bar represents the number of substitutions per site. Node labels represent posterior probability. ST clusters are named on the right-hand side based on the online pubmlst.org *Chlamydiales* database (Pannekoek *et al.* 2008; Jolley & Maiden 2010), with NST representing a novel sequence type.

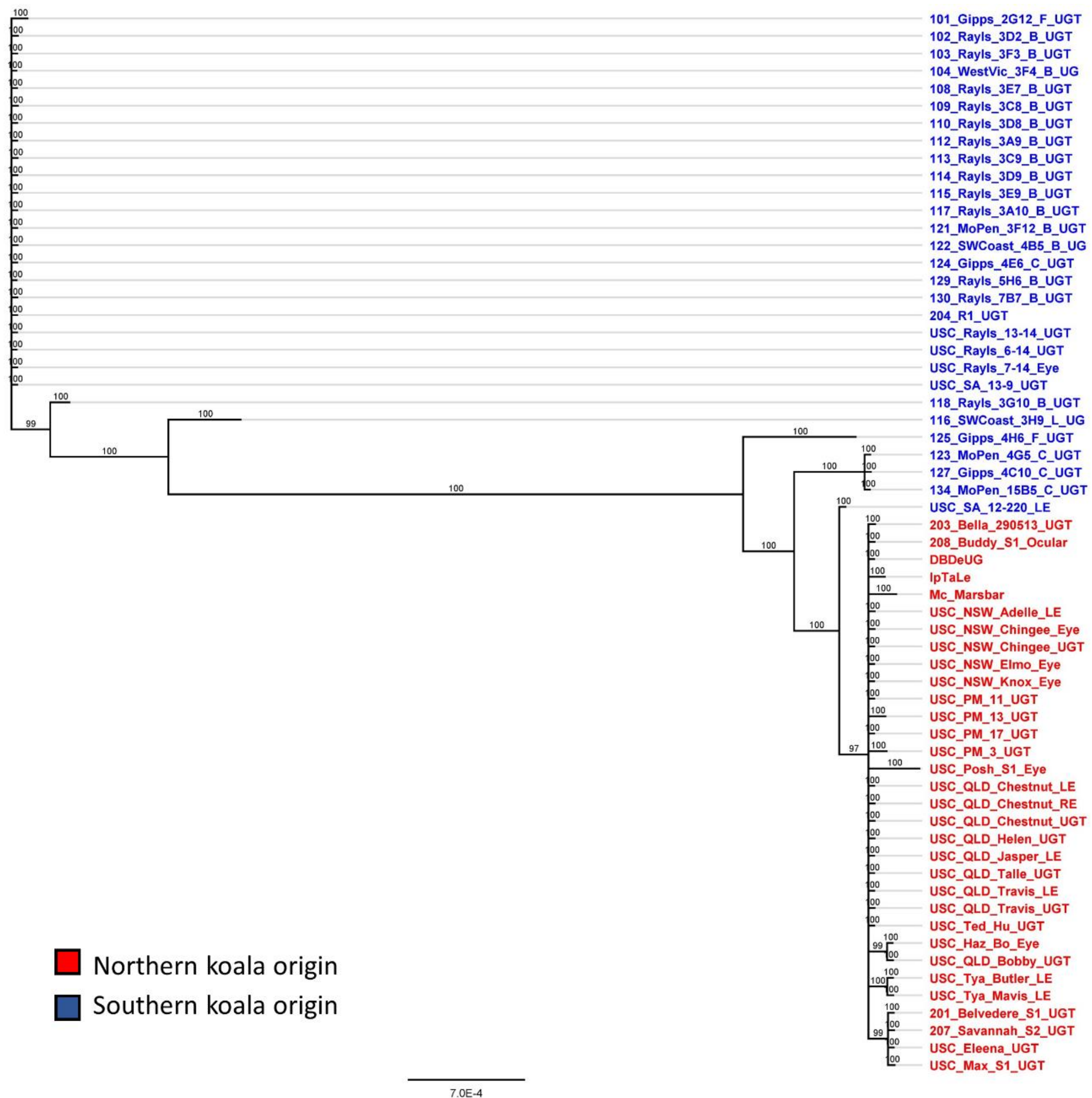



**Figure 7.6.** A phylogenetic tree representing the *C. pecorum* plasmid alignment. An alignment was generated with MAFFT (Katoh *et al.* 2002) and phylogeny was produced using MrBayes (Huelsenbeck & Ronquist 2001), using the HKY+I+G model as determined in jModelTest2 (Darriba *et al.* 2012). Bayesian MCMC analysis utilised 4 chains, with a chain length of 1,000,000 and a relative burn in of 25% chain length. The scale bar represents the number of substitutions per site. Node labels represent posterior probability.

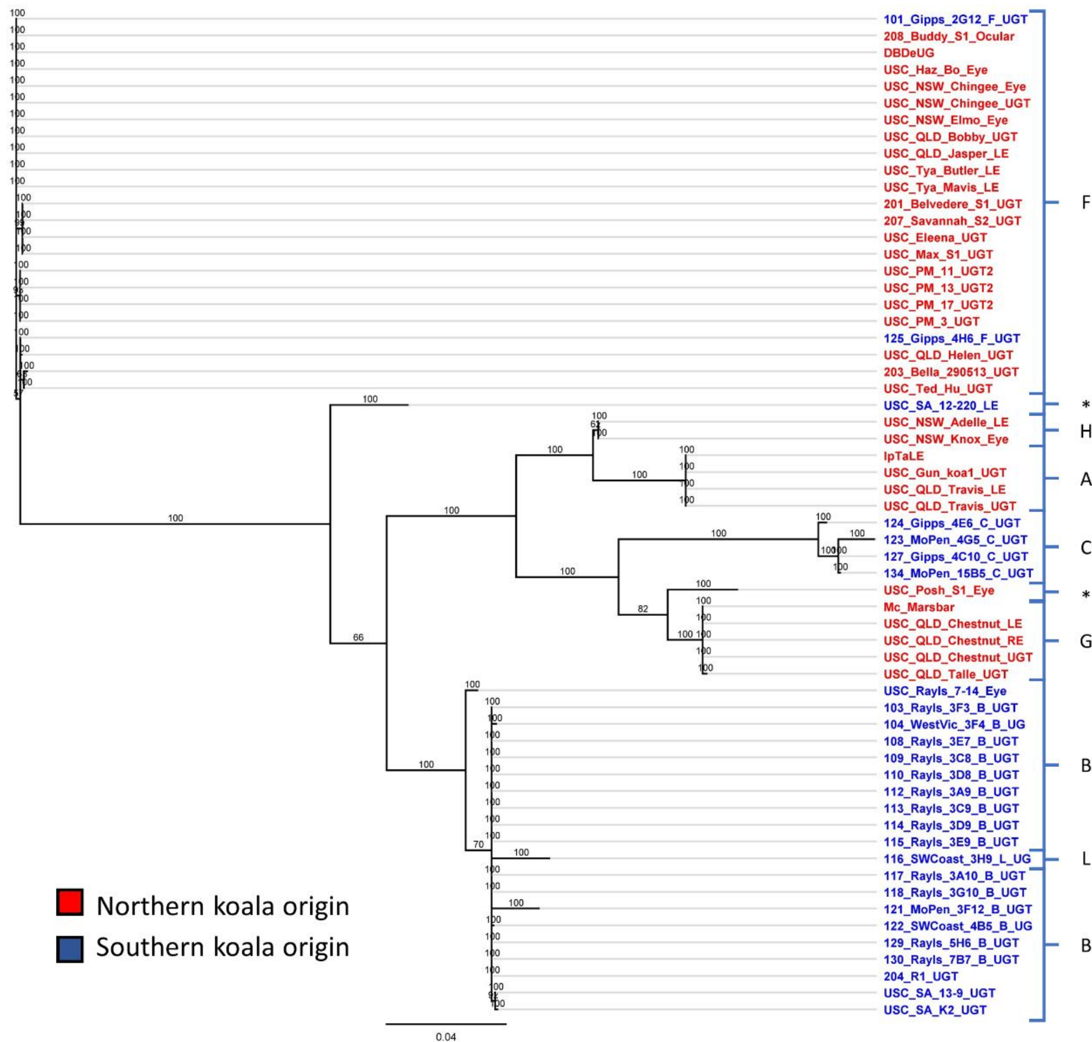



**Figure 7.7.** A phylogenetic tree representing the *C. pecorum ompA* alignment. An alignment was generated with MAFFT (Katoh *et al.* 2002) and phylogeny was produced using MrBayes (Huelsenbeck & Ronquist 2001), using the GTR+I model as determined in jModelTest2 (Darriba *et al.* 2012). Bayesian MCMC analysis utilised 4 chains, with a chain length of 1,000,000 and a relative burn in of 25% chain length. The scale bar represents the number of substitutions per site. Node labels represent posterior probability. *ompA* genotypes are listed on the right-hand side of the tree, with * representing possible novel genotypes with >1% nucleotide difference from previously described genotypes detected in koalas, and nucleotide sequence present in a continuous contig.

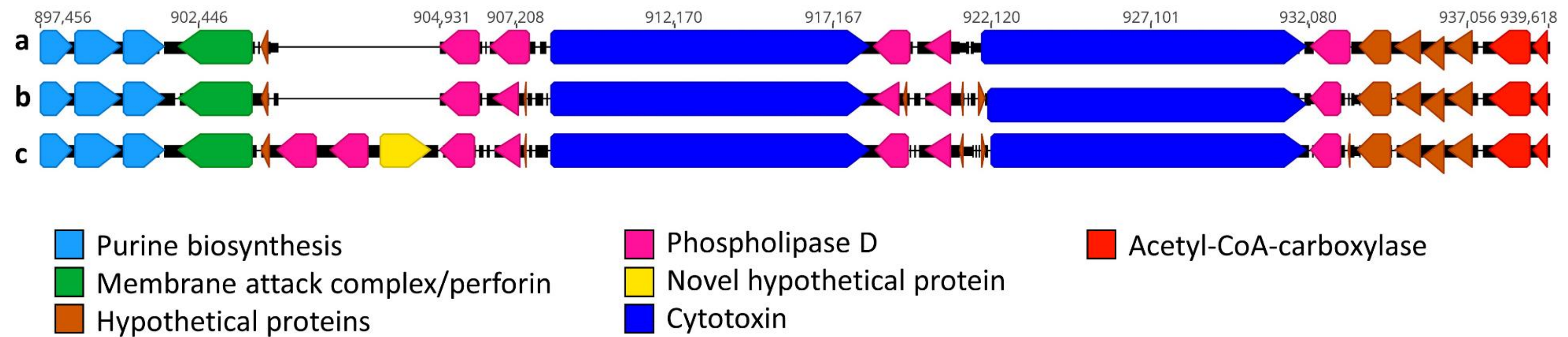



**Figure 7.8.** Alignment of open reading frames and intergenic regions within the plasticity zone of three *C. pecorum* genomes. **a)** E58 reference genome; **b)** 116_SWCoast_3H9_L_UGT; **c)** 129_RayIs_5H6_B_UGT. Truncation of 5' end of coding regions of Phospholipase D and cytotoxin B genes are shown in **b)** and **c)**. Black bars beneath coding regions represent aligned nucleotides. Numbers above alignment represent nucleotide position in relation to E58 genome beginning at the malate dehydrogenase coding region.

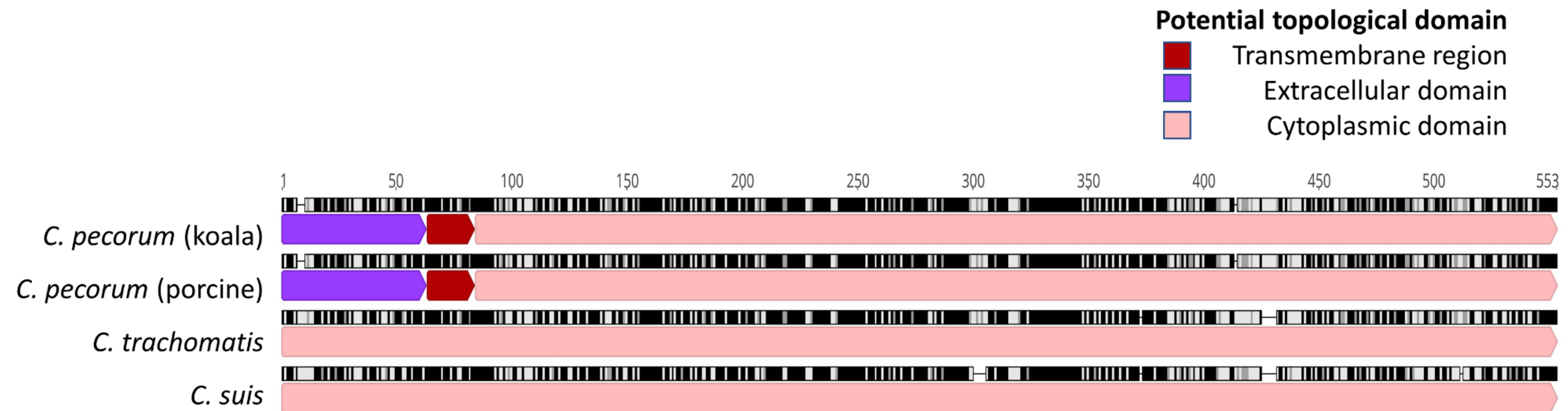



**Figure 7.9.** Amino acid alignment, generated with MAFFT, of a putative membrane protein encoded on the genome of *C. pecorum* from koalas (this study) and pigs (NCBI RefSeq accession number: WP_058787635), *C. trachomatis* (accession number: WP_012727824) and *C. suis* (accession number: WP_080141409). Amino acid similarity is highlighted in each sequence, based on Blosum62 (Henikoff & Henikoff 1992) with a threshold score of 1. 100% similar amino acids in black, 60% or less similar amino acids are white, with shades of grey between these values. Transmembrane Hidden Markov models within the Geneious plugin Transmembrane Prediction Tool version 0.9 were used to predict topological domains.

# 8. General Discussion

## 8.1 Overview of research aims

This body of research aimed to address gaps in our knowledge relating to the prevalence and genotypes of *C. pecorum* and KoRV in Victorian koala populations. To fill these knowledge gaps this study used large numbers of samples collected from koala populations that had been studied previously and also expanded investigations to include new koala populations. In addition to this, genotyping of *C. pecorum* and KoRV samples were undertaken to enable the comparison of genotypes detected in different populations in Victoria, and to allow more accurate comparisons to previously published research from northern koala populations. An investigation into the bacterial diversity of the urogenital tract of the female koala was then undertaken in order to identify other potential causes of the syndrome described as 'wet bottom'. The results from these investigations can be used to strengthen decision making in respect to risk assessment and management of future koala translocations. In addition the investigations enhance our understanding of KoRV and *C. pecorum* in koalas at a national level.

## 8.2 *C. pecorum*, KoRV and wet bottom

The prevalence of *C. pecorum* in the sampled Victorian koala populations ranged from 1 – 45%. The *ompA* genotyping results showed that the genotypes were relatively homogeneous within populations and most were genotype B. Genome sequencing results indicated that genomes within this genotype group were relatively clonal. This probably reflects that koala populations in Victoria are largely derived by translocations from only a few sources. Historical records suggest that the first translocated population to be impacted by *C. pecorum* was on Phillip Island, which was established through translocations of koalas from French Island and Gippsland (Martin & Handasyde 1999). The Gippsland koala population, particularly koalas from the Strzelecki Ranges in west Gippsland, are the only historical remnant populations in Victoria, as demonstrated by several studies investigating genetic profiles of koalas across Australia (Houlden *et al.* 1996; Houlden *et al.* 1999; Lau *et al.* 2014). The French Island koala population was initially founded using Gippsland koalas, and thus all koala populations in Victoria investigated in this study can trace their lineage back to that region. The largest diversity of *ompA* genotypes was identified in Gippsland, which is consistent with

recent *ompA* genotyping studies performed in this region (Wedrowicz *et al.* 2016). It may be that *ompA* genotype B, which is the dominant strain in two regions founded through translocations (Raymond Island and the far western Victoria region), has a reduced pathogenicity compared to other genotypes. This would allow infected individuals to continue to reproduce and thus survive, explaining the predominance of this genotype. It is possible that highly pathogenic strains of *C. pecorum* were translocated into new populations but infected individuals died from disease prior to reproducing, causing these strains to be lost. Genotype F, which has been detected in all northern populations to date (Kollipara 2013; Desclozeaux *et al.* 2017), was only detected in Gippsland in the genotyping study, and as a minor strain in one Raymond Island koala in the genomics study. In the genotyping study (chapter 4) higher normalised *C. pecorum* genomic copy numbers in the small number of cases of genotype F suggests this genotype may be more pathogenic. The detection of a divergent *C. pecorum* genome in the Cape Otway region on the south-west coast of Victoria (*ompA* genotype L), as well as the novel *ompA* genotype N in French Island koalas, may also offer some additional insights into the history of *C. pecorum* spread in koalas. This may include tentative confirmation of undocumented historical translocations of koalas into these regions. The MLST pattern of the novel strain detected on French Island shared a common ancestor with livestock strains of *C. pecorum*. It is unclear whether there is ongoing transmission from livestock to koalas.

Examination of complete, or near complete, *C. pecorum* genome sequences revealed evolutionary differences between samples from Victorian koalas and those collected from koalas in other states that were not always apparent when just the *ompA* sequences were examined. For example, the *ompA* genotype F samples in Victorian koalas clustered with *ompA* genotype F strains from Queensland and New South Wales using *ompA* sequence, but formed their own clades using genome alignments. This may indicate that the genotype F strain in Victorian populations are historical remnants, that are divergent from their northern counterparts. Complete genome sequencing also revealed that *ompA* genotype C samples from Victorian koalas were evolutionary divergent from *C. pecorum* genomes detected both in other Victorian koalas and from northern koalas. The origin of this genotype in koalas is unknown, but it has previously only been detected in Victorian koalas (Jackson *et al.* 1997; Higgins *et al.* 2012). Full genome sequencing also showed clear evidence of recombination in *C. pecorum*

genomes, highlighting the difficulty of producing accurate evolutionary analysis using single gene techniques.

A recent investigation of koala mitochondrial DNA showed that koalas from Cape Otway had a haplotype similar to koalas from New South Wales (Neaves *et al.* 2016), that supports suggestions that undocumented translocations of New South Wales koalas to this region have occurred. Previously it was thought that these populations existed solely through translocations of French Island koalas (Martin 1989). Undocumented translocations may help to explain the presence of a novel *C. pecorum* genome (*ompA* genotype L) in this population. Although *ompA* genotype L has not been detected in New South Wales koalas, only one New South Wales region containing koalas with this same haplotype has been included in previous *C. pecorum* studies and thus further sampling is required to explore this hypothesis. The same study found that haplotypes were shared between koalas from multiple New South Wales regions and the Mount Lofty Ranges koala population in South Australia (Neaves *et al.* 2016). Whilst official records exist of koala translocations from Victorian populations to South Australia (Martin 1989), there are also historical details of koalas of Queensland or New South Wales origin being released into South Australia from captive colonies in that state (Robinson & Bergin 1978). This provides explanation as to why *C. pecorum* genomes from South Australian koalas cluster both with Victorian types and those from northern koala populations. It could also explain the more common occurrence, and greater severity, of ocular infection in South Australian koalas (Funnell *et al.* 2013; Speight *et al.* 2016) compared to Victorian koalas. Speight *et al.* (2016) identified ocular *C. pecorum* infection in 50/65 koalas from the Mount Lofty Ranges and Eyre Peninsula in South Australia. More than half of the individuals tested (34/65) were found to have both ocular and urogenital infection. Only one ocular *C. pecorum* infection was identified in Victorian koalas in this present study, which also contrasts with northern koalas, where ocular infection is found at a similar prevalence to urogenital tract infection (Jackson *et al.* 1999; Polkinghorne *et al.* 2013)

Studies in chapter 5 found that the prevalence of KoRV in the sampled koala populations ranged from 17 – 40%. This prevalence is similar to, or lower than, the prevalence reported in Victorian populations previously, and lower than the prevalence detected in populations in Queensland and New South Wales (Tarlinton *et al.* 2006; Simmons *et al.* 2012). This study was unable to detect KoRV-B in the sampled

population, raising the possibility that this neoplasia-associated variant of KoRV is not present in Victorian koalas. In Victorian koalas KoRV infection was not associated with *C. pecorum* infection, nor with urogenital tract abnormalities. KoRV-B infection has since been found to be associated with the manifestation of chlamydial disease in Queensland koalas. Waugh *et al.* (2017) determined that the likelihood of an individual infected with *C. pecorum* progressing to a disease state was higher in koalas with KoRV-B present. This is consistent with the findings from this present study, as Victorian koalas were infected only with KoRV-A, not KoRV-B, and were not more likely to have *C. pecorum* induced disease if they were infected with KoRV. Recently, Chappell *et al.* (2017) identified several novel genotypes of KoRV that had nucleotide changes to the envelope coding region, as well as variants within each genotype, some of which had SNPs within the *env* gene primer binding regions used in the KoRV genotype studies described in chapter 5. Possibly providing an explanation for instances where *env* could not be amplified successfully. This highlights that other genotypes, not just KoRV-A and B, could be present in Victorian koala populations. In future studies, it would be helpful to perform amplification and deep sequencing of the whole envelope coding region in order to better examine the diversity of KoRV genotypes in Victorian koalas.

The findings in chapter 5 revealed that KoRV was significantly associated with wet bottom in koalas. However, the impact of KoRV on the urogenital microbiome of koalas could not be elucidated as the samples utilised in the microbiome study were from koalas with an unknown KoRV infection status, as blood samples were not available from these animals. Further studies of the microbiome of koalas should consider the impact of the presence or absence of KoRV on the microbiome of koalas with clinical disease. Immunosuppression elicited by KoRV could increase the likelihood of koalas contracting urogenital infections. KoRV-B has been shown to impact cytokine expression in captive koalas (Maher & Higgins 2016), and it is possible that a similar phenomenon exists in individuals infected with KoRV-A.

*Chlamydia* species have long been considered a likely cause of urogenital tract infection in koalas (Dickens 1976), and research in the field has focused heavily on this pathogen. However, in other species urogenital tract infections are not limited to one causative agent (Reviewed in Reid & Sobel 1987; Lockhart *et al.* 1996), and so investigations into other potential agents are indicated. Foundational research in this area is described in

chapter 6, which investigated potential bacterial causes of wet bottom using microbial diversity analysis. Through the comparison of koalas with no apparent clinical signs of disease and those with wet bottom, potential causative agents from two bacterial families were identified. These were the *Aerococcaceae* and *Tissierellaceae* families, both of which may have been overlooked in early culture based studies. This is due to difficulties in differentiating *Aerococcaceae* from *Streptococcaceae* family members (Martín *et al.* 2007; de Jong *et al.* 2010), who share the order *Lactobacillales*, and due to the anaerobic growth requirements of the *Tissierellaceae* (Ezaki *et al.* 2001). Early studies identified *Streptococcus*-like organisms in the reproductive tract of koalas with pyometritis (Obendorf 1981), in addition to *E. coli* in some instances (McKenzie 1981; Obendorf 1981). An *E. coli*-like sequence was identified as the most prevalent sequence in one of the koalas in this study displaying wet bottom. These findings, although only incorporating a small number of individuals, suggest that there may be multiple causative bacterial agents of wet bottom in koalas. The sample archive used in chapters 3 and 4 promise to be a valuable resource for screening for potential pathogens in a large number of samples using targeted PCRs for specific bacterial genera.

## 8.3 Health status of Victorian koalas

When considered in concert, the findings of this thesis show that whilst *C. pecorum* and KoRV are both present in Victorian koalas, the impact of these infections varies across regional populations. In particular, the over-abundance of koalas in regions such as Mt Eccles in the far west of Victoria (McLean & Handasyde 2007), Cape Otway in the south-west coastal region of Victoria (Whisson *et al.* 2016), and French Island (McLean & Handasyde 2007), can be partly explained by very low prevalence of *C. pecorum* infection (1 – 21%), which would in turn have an impact on replacement rate within the population. Furthermore, a low prevalence of KoRV infection, relatively low KoRV loads in infected animals (also identified in Simmons *et al.* 2012; Wedrowicz *et al.* 2016) and the absence of the KoRV-B genotype may also contribute to koala health. The presence of KoRV-B in Queensland populations has been shown to increase the prevalence of chlamydial disease (Waugh *et al.* 2017), however its impact on detectable shedding of *C. pecorum* has not been established, and it is therefore unclear if KoRV contributes to *C. pecorum* transmission. The increased capability for *C. pecorum* to cause reproductive changes in Queensland populations may explain why fertility is lower in these populations than in Victorian populations.

### 8.3.1 Implications for koala translocations

This research will aid decision making regarding the use of koalas from Victoria as a source population for re-establishing koala populations in other Australian states, and in regions of Victoria if required. If translocations are to be used as a means of re-establishing these populations, then the results from this study shows that utilising koalas free of *Chlamydia* and KoRV could be achieved.

To ensure that all translocated animals were free of these pathogens the animals would mostly likely need to be held for long periods before translocation, although this could be reduced by the development of more rapid and accurate diagnostics. Standard protocols dictate that koalas should not be held in crates for more than 24 hours during translocation procedures (Menkhorst 2004). The development of more rapid and accurate diagnostic tests that can be used in the field would assist with identifying koalas free from KoRV and *Chlamydia*, thus allowing them to be selected for translocation more rapidly. The development of such diagnostic tests should be prioritised. Currently, KoRV diagnosis relies on qPCR (Tarlinton *et al.* 2005). *C. pecorum* diagnosis is more varied in the literature, but includes qPCR (Markey *et al.* 2007; Robertson *et al.* 2009) and ELISA (Carey *et al.* 2010). Although qPCR is rapid, this technique requires specialised laboratory reagents and equipment for the correct extraction, amplification, and detection of the target genes, precluding their use in a field setting. ELISA methods previously utilised for koala *C. pecorum* are limited by their requirement for anti-koala antibodies. These antibodies are not sold commercially, and requires the collection of koala serum and subsequent inoculation into another species, such as rabbits (Wilkinson *et al.* 1991) or sheep (Carey *et al.* 2010), to produce the antisera. The solid-phase ELISA 'Clearview' *Chlamydia* test, which can be used in the field, is not recommended for modern field diagnostics due to the poor sensitivity of the assay (43.2%) (Hanger *et al.* 2013).

Disease risk analysis techniques can be utilised to minimise the risk associated with translocation of animals to a new location (Hartley & Sainsbury 2017). The results described in this thesis increases our knowledge of Victorian koalas, making any potential risk analysis more robust. Disease risk analysis requires knowledge on pathogen prevalence in both source populations and in the regions individuals are to be

translocated to, as well as the likelihood of these diseases impacting other species. Knowledge of the pathogen prevalence in the target region is vital to the success of a translocation. Research conducted in the 1990s highlighted the problematic nature of introducing naïve koalas into a region where *C. pecorum* prevalence was not assessed (Santamaria & Schlagloth 2016). Of 30 koalas that were translocated from French Island to mainland Victoria, 16/17 of koalas recaptured 19 months later were positive for *Chlamydia* antibodies, and 8 were PCR positive for *C. pecorum* (Santamaria & Schlagloth 2016). All recaptured koalas were outwardly healthy based on body condition scores, however only 1/12 females had live progeny. Translocation of koalas to areas where there are no current *C. pecorum* infections would be preferential, regardless of their own infection status, even if this requires prior assessment of forests to ensure no koala populations exist. This is due to the potential for mixed infections of *C. pecorum* in koalas, which was highlighted in chapter 7, and has also been demonstrated previously (Bachmann *et al.* 2015). Therefore, it should not be assumed that individuals with *C. pecorum* infections are protected from future infection with different, potentially more pathogenic strains. Chapter 3 highlighted that *C. pecorum* infections occur in an isolated koala population previously not found to carry the infection, possibly through cross-species transmission from livestock. Livestock shedding *C. pecorum* onto pastural lands through faeces (Osman *et al.* 2011; Yang *et al.* 2014; Li *et al.* 2016), may impact newly established koala populations should individuals be translocated to regions adjacent to these areas. The exact method of transmission from livestock is unknown, but it is likely that koalas encounter contaminated soil when traversing agricultural land (White 1999; Davies *et al.* 2013).

### 8.3.2 Implications for KoRV and *C. pecorum* vaccine development

The findings of this research have broader implications for the development of tools to combat KoRV and *C. pecorum* in koalas, particularly relating to the recent advances in vaccine development.

Vaccination of koalas against *C. pecorum* using a polyvalent recombinant protein vaccine containing MOMP (genotypes A, F & G) has shown mixed success (Khan *et al.* 2014). Waugh *et al.* (2016b) showed that, compared to a control group, vaccination did not prevent new infections, did not substantially reduce *C. pecorum* load over the course of a twelve-month trial, and did not reduce the likelihood of infection progressing to clinical disease. A study in the same region of Queensland trialling the same polyvalent

MOMP vaccine, as well as a *C. pecorum* vaccine using PmpG as an antigen, also had mixed results (Desclozeaux *et al.* 2017). Both vaccines produced systemic humoral and cell mediated immune responses in koalas, as measured by comparing IgG and cytokine (IFN-γ and IL-17) levels at inoculation and six-months later. At the end of the six-month trial period, the recombinant MOMP vaccine reduced *C. pecorum* burden in six koalas that were infected at the beginning of the trial. The PmpG vaccine group contained no infected koalas at the beginning of the trial, but three koalas contracted *C. pecorum* by the end of the six-month period. Koalas in both vaccinated groups (5/42 koalas in total) suffered new infections within one to three months after the trial, and novel *ompA* genotypes (named genotype O) were detected in two individuals. Results in chapters 3, 4, and 7 showed that *C. pecorum* in Victorian koalas is predominately genotype B, and that mixed infections are less common, suggesting that using MOMP as a target could be successful in Victoria, providing the vaccine afforded long-term protection. Development of targeted vaccines whereby the *ompA* genotypes used to produce antigens are tailored to the known *ompA* genotypes present in a particular region might be achievable for protecting animals translocated to regions where a particular *ompA* genotype is known to be dominant. This may however, increase selection pressure for other genotypes to become established, and further research is required as to the likelihood of this outcome.

KoRV vaccines are currently in the early stages of development, and there is some conjecture over the likelihood of their effectiveness. The basis for the development of these vaccines comes from the successful development of recombinant vaccines against FeLV using envelope proteins (Clark *et al.* 1991; Langhammer *et al.* 2006). The immunogenic proteins transcribed by the envelope region of the KoRV provirus, specifically p15E and gp70, have been trialled as vaccine antigens with mixed success. Initially, neutralising antibodies were shown to be produced against these proteins when inoculated into rats and goats (Fiebig *et al.* 2015a). However, the inoculation of KoRV positive captive koalas failed to produce a similar response, suggesting individuals with endogenous virus recognised the antigens as self (Fiebig *et al.* 2015b), as occurs with other endogenous gammaretroviruses such as PERV (Keller et al. 2014). Subsequent studies conversely found that seven koalas, from Queensland (4), New South Wales (1), and South Australia (2), had p15E and/or gp70 specific antibodies from natural infection (Waugh *et al.* 2016a). Inoculation of three of these individuals with the

vaccine antigens, in conjunction with adjuvants (poly I:C, host defence peptide, and polyphosphazine), was not able to induce a quantifiable increase in antibody presence. One individual without apparent gp70 antibodies pre-vaccination did produce anti-gp70 post-immunisation, based on western blot results. These studies show the requirement for further investigation and development of these vaccines. The effectiveness of vaccination as a means of combatting KoRV would be difficult to determine in northern koalas, considering the virus is endogenous. Research in chapter 5 highlighted that Victorian koalas have a relatively low prevalence of KoRV-A and no detectable KoRV-B. It is not known if KoRV PCR-negative koalas are likely to have been exposed to KoRV previously, and the current scenario in Victoria presents a unique opportunity to serologically determine whether animals without detectable KoRV provirus in their genome have previously mounted an antibody response to the virus. This would significantly advance our understanding both of the virus itself and the ability of the koala to combat infection.

### 8.3.3 Implications for future prevalence and genotyping studies

The findings of this research provide some confidence in the widely used *C. pecorum* typing methodology, such as *ompA* genotyping and MLST analysis. The phylogenetic trees produced from target genes used in these typing methods were largely congruent with trees produced from full genomes of the same strains in regard to strain clustering by population of origin. The main outlier was the *ompA* genotype F, as discussed in section 8.2. However, whilst population clustering was largely consistent between trees, branching patterns were different within the broader clades, highlighting divergence between single genes and whole genome sequences. The use of single gene genotyping in the future is likely to be driven largely by the intended outcome of the typing system. If the method intends to reflect full genome evolution, targeting a gene with a high nucleotide diversity but neutral selection is be preferential. The results in chapter 7 suggest a gene such as *tyrS* (encoding tyrosine-tRNA ligase), which had a greater number of segregating sites than *ompA*, a lower omega value, and far fewer haplotypes, would allow more accurate evolutionary analysis of a large sample size without the costs associated with MLST and full genome sequencing.

## 8.4 Future research directions

### 8.4.1 KoRV infection in southern koalas

As previously outlined, southern koalas offer a unique opportunity to test the ability of koalas to mount an immune response to KoRV infection. Future work in this area would be enhanced by the development of an effective test for the detection of serum antibody against KoRV. The development of useful serological tests for wildlife has been hampered by the requirement of producing species-specific antibodies. Investigations into the binding ability of marsupial serum immunoglobulins to three immunogenic proteins (staphylococcal protein A, streptococcal protein G, and peptostreptococcal protein L) have shown promising results as alternatives to anti-koala antibodies (Vaz *et al.* 2015). Koala immunoglobulins bound efficiently to low concentrations of proteins G and L in both ELISA and immunoblot tests. These proteins, conjugated with horse radish peroxidase, can provide the basis for serological tests to help learn about the exposure history of southern koalas to KoRV. Other avenues of research which would be valuable to pursue in southern koalas include investigating ancient DNA in southern koala museum specimens, and modelling transmission dynamics of the virus based on the results. Ancient DNA methods were previously successful at dating KoRV incursion into the koala genome as far back as the 1870s (Ávila-Arcos *et al.* 2013). This study did not incorporate samples from Victorian koalas, and assessing the presence of KoRV in historical specimens might allow us to determine if KoRV presence pre-dates the translocation of koalas to French Island. This would be helpful for determining if the koala population on French Island was founded with KoRV present. If the French Island population was founded with KoRV present this would potentially enable modelling of the spread of the infection in that population.

### 8.4.2 Diagnostic testing to aid translocation

Serological tests are valuable for the assessment of past exposure to pathogens but antigen-based tests are also required in order to best inform decisions on koala translocations. A possible avenue for rapid field-based diagnostics is loop-mediated isothermal amplification (LAMP) assays. LAMP assays are a form of nucleic acid amplification that can be undertaken at a stable temperature (Notomi *et al.* 2000), rather than temperature cycling as traditional PCR requires (Saiki *et al.* 1988). The capability to amplify DNA at a stable temperature means the process only requires rudimentary

equipment and thus can be conducted on samples in the field. With the addition of fluorophores for the detection of real time results, this process can be completed in under an hour. The use of LAMP to detect *Chlamydia* spp. in clinical cases has already been achieved (Kawai *et al.* 2009; Gandelman *et al.* 2010), and the near complete genomes of *C. pecorum* produced through this project provide researchers with the resources to produce robust primers for detection of conserved target genes. Similarly, reverse transcriptase LAMP has been successfully used on plasma samples from humans to detect HIV-1 (Curtis *et al.* 2008), and could be adapted to test for KoRV in the field. The development of such diagnostic tools would allow real time detection of infections, and allow population managers to confidently select appropriate individuals for translocation based on infection status rather than solely on clinical signs of disease.

### 8.4.3 Cross species transmissions of *C. pecorum*

The *C. pecorum ompA* genotype N from two French Island koalas (chapter 3) had a novel MLST allele pattern that clustered most closely with livestock strains. This highlights the potential for cross species transmission as an explanation for *C. pecorum* detection in a koala population from which the pathogen had not been previously detected using modern molecular techniques. Additionally, the French Island population has had no recorded introductions of new koalas since its initial establishment over 100 years ago, but does have currently active livestock producers. The detection of a novel genotype, specific to the Cape Otway region (genotype L), which was thought to have been established entirely with French Island koala, also raises the possibility of continuing transmissions from an unknown source. Livestock are known to be affected by *C. pecorum* infection, detailed in chapter 1, but they can also be a reservoir host. Research in China suggested that *C. pecorum* is endemic in the gastrointestinal tract of dairy cattle in this country (Li *et al.* 2016), but little published molecular epidemiological research exists on *C. pecorum* in livestock in Australia. One study of *C. pecorum* in lambs in Victoria, South Australia, Western Australia, and New South Wales, used a qPCR targeting *ompA* to screen faecal samples (Yang *et al.* 2014). The lowest prevalence was detected in Victoria (ranging from 4.9 – 18.3%) whilst the highest was in New South Wales (42.5 – 80.8%), in part mirroring the prevalence of *C. pecorum* in koalas, as described in chapter 4 and in previous work (Kollipara *et al.* 2013b; Polkinghorne *et al.* 2013). Unfortunately, no *ompA* genotyping nor MLST analysis was undertaken in this study of *C. pecorum* in Australian lambs. The only

MLST analysis previously undertaken on a substantial number of Australian livestock was on sheep in New South Wales (n = 62) (Jelocnik *et al.* 2013; Jelocnik *et al.* 2014b). One out of twelve identified sequence types, ST69, which was associated with conjunctivitis in sheep, has also been detected in koalas (Jelocnik *et al.* 2013; Jelocnik *et al.* 2014b). Only three *C. pecorum* infections from Australian cattle have been typed using MLST (Jelocnik *et al.* 2013; Jelocnik *et al.* 2014a), all from cases of encephalomyelitis. This lack of data highlights that a genetic survey of *C. pecorum* in livestock species in Victoria, and across Australia, is required. Doing so will help to investigate whether *C. pecorum* in different koala populations has evolved from a common ancestor, or from multiple cross-species transmissions from livestock. If a greater understanding is achieved on the relationship between *C. pecorum* in livestock and koalas, this would help to inform future management decisions.

### 8.4.4 Other potential pathogens in koalas

There have been sporadic reports of other pathogens, or potential pathogens, detected in koalas as detailed in chapter 1. Recently case studies have revealed the severe impact of *S. scabiei* on koalas in southern populations, identifying the mite in 58 koalas in South Australia and Victoria between 2008 and 2015 (Speight *et al.* 2017). These koalas presented with gross pathology including skin thickening, crusting and deep fissures into the dermis. Further work to assess the prevalence of *S. scabiei* infestation in other koala populations is indicated, as well as further work to investigate any potential associations between *S. scabiei* infection and concurrent infection with KoRV or *C. pecorum*. Interestingly, the cytokine that has been shown to be a marker for chlamydial disease severity, IL-17A (Mathew *et al.* 2014) has also been implicated in the development of 'crusted' scabies pathogenesis in porcine models of disease (Mounsey *et al.* 2015). The possible link between these pathogens in koalas warrants further investigation.

Two viral families, *Herpesviridae* and *Papillomaviridae,* are also of interest, due to their possible associations with neoplasia, which is an issue in northern and captive populations of koalas. Multiple members of both these viral families have been identified in koalas previously (Antonsson & McMillan 2006; Vaz *et al.* 2011; Vaz *et al.* 2012). Both koala herpesviruses detected to date have been gammaherpesviruses. Gammaherpesviruses in humans include Epstein-Barr virus (human herpesvirus 4) and Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8), both of which cause

neoplasia in immunocompromised patients (Epstein *et al.* 1964; Chang *et al.* 1994). Specific strains of human papilloma virus have also been shown to be responsible for a majority of cervical cancers (Bosch *et al.* 1995; Muñoz *et al.* 2003). Associations have been detected between progression to cervical cancer caused by human papillomavirus and immunosuppression due to HIV (Moscicki *et al.* 2000). Determining whether a similar relationship exists between KoRV and koala papillomavirus species will aid our understanding of neoplasia in koalas. A broad survey of herpesviruses in Australian marsupials, which included 99 koalas, 98 of which were from Victoria, found a significant association between the detection of herpesvirus and *C. pecorum* (Stalder et al. 2015). Koalas with herpesvirus were 60 times more likely to be detected with *C. pecorum.* This association has not been investigated further to date. Wider surveys of both koala herpesviruses and papillomaviruses would be worthwhile in an effort to establish whether they are involved in gross changes of the reproductive tract such as cysts, or any of the various clinical diseases collectively denoted as “koala AIDS” (Hanger & Loader 2014).

#### 8.4.4.1 The use of metagenomics to further understand disease in koalas

Research on the urogenital microbiome of koalas (chapter 6) proved useful in identifying potential causative agents of wet bottom in koalas, as well as establishing foundational knowledge on the normal flora of the urogenital tract of female koalas. Two primary constraints of the methods utilised are: 1) the curated 16S rRNA databases rely on the prior sequencing of organisms to characterise bacterial species (DeSantis *et al.* 2006), thus correct taxonomic matches from a novel environment are likely to be limited; and 2) the methodology cannot be used to identify other pathogens such as viruses, fungi, protozoa, and complex parasites. The first constraint resulted in difficulties classifying the majority of the OTUs found in the urogenital tract beyond the genus level. Culture of organisms that appeared to be either at a high abundance in clinically normal koalas, or those displaying wet bottom, would be valuable. Doing so would allow genomic sequencing of novel organisms, as well as biochemical testing to further characterise their potential role in the microbiome. The exploration of non-bacterial pathogens is more difficult, both in respect to increased cost and the detection of previously uncharacterised DNA sequences. In a similar fashion to the use of the 16S rRNA gene of bacteria, the ITS gene can be used to investigate the diversity of fungal species in a sample (Schoch *et al.* 2012). Shotgun sequencing of all genomic DNA

present in a sample will help to resolve non-bacterial aetiological agents, not just of wet bottom, but other clinical signs of disease in koalas such as those reported by Hanger and Loader (2014), including polycystic kidney disease and carcinomas of the renal tract. Such approaches have been applied to other species to identify eukaryotic virus families associated with disease (Hewson *et al.* 2014; Moreno *et al.* 2017). Utilising similar metagenomic techniques will allow the identification of previously undetected causative agents of clinical disease in koalas, which can then be investigated through targeted approaches.

## 8.5 Concluding remarks

The knowledge obtained by this body of research provides completely new information about the state of *C. pecorum* and KoRV infection in Victorian koalas, as well as clarifying questions raised by previous research. The prevalence surveys, as well as the more detailed genotyping studies, provides future researchers and koala population managers a more thorough understanding of the infection status of koalas in different Victorian populations. The microbiome component of this research lays the ground work for identifying other causative agents of clinical disease in koalas, whilst the genomic comparison research will assist with elucidating the origins of *C. pecorum* in koalas once more livestock origin *C. pecorum* genomes are sequenced and available for comparison. Furthermore, the increased understanding of the variation between *C. pecorum* strains affecting koalas in broader geographic regions than previously described will assist with determining new genomic regions of interest for both vaccine and diagnostic development.

# 9. References

**Aagaard, K., K. Riehle, J. Ma, N. Segata, T.-A. Mistretta, C. Coarfa, S. Raza, S. Rosenbaum, I. Van den Veyver, A. Milosavljevic, D. Gevers, C. Huttenhower, J. Petrosino & J. Versalovic (2012).** "A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy." *PLoS ONE* **7**, e36466 DOI: 10.1371/journal.pone.0036466.

**Adams, M. J., E. J. Lefkowitz, A. M. Q. King, B. Harrach, R. L. Harrison, N. J. Knowles, A. M. Kropinski, M. Krupovic, J. H. Kuhn, A. R. Mushegian, M. Nibert, S. Sabanadzovic, H. Sanfaçon, S. G. Siddell, P. Simmonds, A. Varsani, F. M. Zerbini, A. E. Gorbalenya & A. J. Davison (2017).** "Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses." *Arch Virol* **162**(8), 2505-2538.

**Akaike, H. (1987).** "Factor analysis and AIC." *Psychometrika* **52**(3), 317-332.

**Alfano, N., A. Courtiol, H. Vielgrader, P. Timms, A. L. Roca & A. D. Greenwood (2015).** "Variation in koala microbiomes within and between individuals: effect of body region and captivity status." *Sci Rep* **5**, 10189.

**Alfano, N., J. Michaux, S. Morand, K. Aplin, K. Tsangaras, U. Löber, P.-H. Fabre, Y. Fitriana, G. Semiadi, Y. Ishida, K. M. Helgen, A. L. Roca, M. V. Eiden & A. D. Greenwood (2016).** "An endogenous gibbon ape leukemia virus (GALV) identified in a rodent (*Melomys burtoni* subsp.) from Wallacea (Indonesia)." *J Virol* **90**(18), 8169-8180.

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

**Anai, Y., H. Ochi, S. Watanabe, S. Nakagawa, M. Kawamura, T. Gojobori & K. Nishigaki (2012).** "Infectious endogenous retroviruses in cats and emergence of recombinant viruses." *J Virol* **86**(16), 8634-8644.

**Antonsson, A. & N. A. J. McMillan (2006).** "Papillomavirus in healthy skin of Australian animals." *J Gen Virol* **87**(11), 3195-3200.

**Atlas of Living Australia. (2017).** "Koala distribution." Retrieved 05/07/2017, from https://www.ala.org.au/.

**Ávila-Arcos, M. C., S. Y. W. Ho, Y. Ishida, N. Nikolaidis, K. Tsangaras, K. Hönig, R. Medina, M. Rasmussen, S. L. Fordyce, S. Calvignac-Spencer, E. Willerslev, M. T. P. Gilbert, K. M. Helgen, A. L. Roca & A. D. Greenwood (2013).** "One hundred twenty years of koala retrovirus evolution determined from museum skins." *Mol Biol Evol* **30**(2), 299-304.

**Bachmann, N. L., T. A. Fraser, C. Bertelli, M. Jelocnik, A. K. Gillett, O. Funnell, C. Flanagan, G. S. Myers, P. Timms & A. Polkinghorne (2014).** "Comparative genomics of koala, cattle and sheep strains of *Chlamydia pecorum*." *BMC Genomics* **15**, 667.

**Bachmann, N. L., M. J. Sullivan, M. Jelocnik, G. S. A. Myers, P. Timms & A. Polkinghorne (2015).** "Culture-independent genome sequencing of clinical samples reveals an unexpected heterogeneity of infections by *Chlamydia pecorum*." *J Clin Microbiol* **53**(5), 1573-1581.

**Backhouse, T. C. & A. Bolliger (1961).** "Morbidity and mortality in the koala (*Phascolarctos cinereus*)." *Aust J Zool* **9**(1), 24-37.

**Baltimore, D. (1970).** "Viral RNA-dependent DNA polymerase: RNA-dependent DNA polymerase in virions of RNA tumour viruses." *Nature* **226**(5252), 1209-1211.

**Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev & P. A. Pevzner (2012).** "SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing." *J Comput Biol* **19**(5), 455-477.

**Barbosa, A., J. Austen, A. Gillett, K. Warren, A. Paparini, P. Irwin & U. Ryan (2016).** "First report of *Trypanosoma vegrandis* in koalas (*Phascolarctos cinereus*)." *Parasitol Int* **65**(4), 316-318.

**Barker, C. J., A. Gillett, A. Polkinghorne & P. Timms (2013).** "Investigation of the koala (*Phascolarctos cinereus*) hindgut microbiome via 16S pyrosequencing." *Vet Microbiol* **167**(3–4), 554-564.

**Bazala, E. & J. Renda (2005).** "Latent chlamydial infections: the probable cause of a wide spectrum of human diseases." *Med Hypotheses* **65**(3), 578-584.

**Bedson, S. P. & J. O. W. Bland (1932).** "A morphological study of psittacosis virus, with the description of a developmental cycle." *Br J Exp Pathol* **13**(5), 461-466.

**Beem, M. O. & E. M. Saxon (1977).** "Respiratory-tract colonization and a distinctive pneumonia syndrome in infants infected with *Chlamydia trachomatis*." *New Engl J Med* **296**(6), 306-310.

**Belland, R. J., M. A. Scidmore, D. D. Crane, D. M. Hogan, W. Whitmire, G. McClarty & H. D. Caldwell (2001).** "*Chlamydia trachomatis* cytotoxicity associated with complete and partial cytotoxin genes." *Proc Natl Acad Sci USA* **98**(24), 13984-13989.

**Benjamini, Y. & Y. Hochberg (1995).** "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *J Roy Stat Soc Ser B (Stat Method)* **57**(1), 289-300.

**Blanshard, W. H. & K. Bodley (2008).** Koalas. *Medicine of Australian Mammals*. Vogelnest, L. & R. Woods. Collingwood, Victoria, Australia, CSIRO Publishing**:** 227-327.

**Bodetti, T. J., K. Viggers, K. Warren, R. Swan, S. Conaghty, C. Sims & P. Timms (2003).** "Wide range of *Chlamydiales* types detected in native Australian mammals." *Vet Microbiol* **96**(2), 177-187.

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

**Bosch, F. X., M. M. Manos, N. Muñoz, M. Sherman, A. M. Jansen, J. Peto, M. H. Schiffman, V. Moreno, R. Kurman & K. V. Shan (1995).** "Prevalence of human papillomavirus in cervical cancer: a worldwide perspective." *J Natl Cancer Inst* **87**(11), 796-802.

**Boskey, E. R., K. M. Telsch, K. J. Whaley, T. R. Moench & R. A. Cone (1999).** "Acid production by vaginal flora in vitro is consistent with the rate and extent of vaginal acidification." *Infect Immun* **67**(10), 5170-5175.

**Branley, J., N. L. Bachmann, M. Jelocnik, G. S. A. Myers & A. Polkinghorne (2016).** "Australian human and parrot *Chlamydia psittaci* strains cluster within the highly virulent 6BC clade of this important zoonotic pathogen." *Sci Rep* **6**, 30019 DOI: 10.1038/srep30019.

**Bray, J. R. & J. T. Curtis (1957).** "An ordination of the upland forest communities of southern Wisconsin." *Ecol Monogr* **27**(4), 325-349.

**Brown, A. S., A. A. Girjes, M. F. Lavin, P. Timms & J. B. Woolcock (1987).** "Chlamydial disease in koalas." *Aust Vet J* **64**(11), 346-350.

**Brown, A. S. & R. G. Grice (1984).** "Isolation of *Chlamydia psittaci* from koalas (*Phascolarctos cinereus*)." *Aust Vet J* **61**(12), 413.

**Brunelle, B. W. & G. F. Sensabaugh (2006).** "The *ompA* gene in *Chlamydia trachomatis* differs in phylogeny and rate of evolution from other regions of the genome." *Infect Immun* **74**(1), 578-585.

**Brynildsrud, O., J. Bohlin, L. Scheffer & V. Eldholm (2016).** "Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary." *Genome Biol* **17**, 238.

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

**Burnard, D., H. Weaver, A. Gillett, J. Loader, C. Flanagan & A. Polkinghorne (2017).** "Novel *Chlamydiales* genotypes identified in ticks from Australian wildlife." *Parasit Vectors* **10**(1), 46.

**Byrne, G. I. (2010).** "*Chlamydia trachomatis* strains and virulence: rethinking links to infection prevalence and disease severity." *J Infect Dis* **201**(S2), S126-S133.

**Byrne, G. I. & J. W. Moulder (1978).** "Parasite-specified phagocytosis of *Chlamydia psittaci* and *Chlamydia trachomatis* by L and HeLa cells." *Infect Immun* **19**(2), 598-606.

**Caldwell, H. D., J. Kromhout & J. Schachter (1981).** "Purification and partial characterization of the major outer membrane protein of *Chlamydia trachomatis*." *Infect Immun* **31**(3), 1161-1176.

**Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer & T. L. Madden (2009).** "BLAST+: architecture and applications." *BMC Bioinformatics* **10**, 421.

**Canfield, P. J. (1987).** "A mortality survey of free range koalas from the north coast of New South Wales." *Aust Vet J* **64**(11), 325-328.

**Canfield, P. J., A. S. Brown, W. R. Kelly & R. H. Sutton (1987).** "Spontaneous lymphoid neoplasia in the koala (*Phascolarctos cinereus*)." *J Comp Pathol* **97**(2), 171-178.

**Canfield, P. J., W. J. Hartley & J. P. Dubey (1990).** "Lesions of toxoplasmosis in Australian marsupials." *J Comp Pathol* **103**(2), 159-167.

**Canfield, P. J., D. N. Love, G. Mearns & E. Farram (1991).** "Chlamydial infection in a colony of captive koalas." *Aust Vet J* **68**(5), 167-169.

**Canfield, P. J., C. J. Oxenford, D. N. Love & R. K. Dickens (1983).** "Pyometra and pyovagina in koalas." *Aust Vet J* **60**(11), 337-338.

**Canfield, P. J., J. M. Sabine & D. N. Love (1988).** "Virus particles associated with leukaemia in a koala." *Aust Vet J* **65**(10), 327-328.

**Caporaso, J. G., K. Bittinger, F. D. Bushman, T. Z. DeSantis, G. L. Andersen & R. Knight (2010a).** "PyNAST: a flexible tool for aligning sequences to a template alignment." *Bioinformatics* **26**(2), 266-267.

**Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Pena, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld & R. Knight (2010b).** "QIIME allows analysis of high-throughput community sequencing data." *Nat Methods* **7**(5), 335-336.

**Carey, A. J., P. Timms, G. Rawlinson, J. Brumm, K. Nilsson, J. M. Harris & K. W. Beagley (2010).** "A multi-subunit chlamydial vaccine induces antibody and cell-mediated immunity in immunized koalas (*Phascolarctos cinereus*): comparison of three different adjuvants." *Am J Reprod Immunol* **63**(2), 161-172.

**Carlson, J. H., S. Hughes, D. Hogan, G. Cieplak, D. E. Sturdevant, G. McClarty, H. D. Caldwell & R. J. Belland (2004).** "Polymorphisms in the *Chlamydia trachomatis c*ytotoxin locus associated with ocular and genital Isolates." *Infect Immun* **72**(12), 7063-7072.

**Carlson, J. H., S. F. Porcella, G. McClarty & H. D. Caldwell (2005).** "Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and genitotropic strains." *Infect Immun* **73**(10), 6407-6418.

**Carlson, J. H., W. M. Whitmire, D. D. Crane, L. Wicke, K. Virtaneva, D. E. Sturdevant, J. J. Kupko, S. F. Porcella, N. Martinez-Orengo, R. A. Heinzen, L. Kari & H. D. Caldwell (2008).** "The *Chlamydia trachomatis* plasmid is a transcriptional regulator of chromosomal genes and a virulence factor." *Infect Immun* **76**(6), 2273-2283.

**Castresana, J. (2000).** "Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis." *Mol Biol Evol* **17**(4), 540-552.

**Cattori, V., R. Tandon, B. Riond, A. C. Pepin, H. Lutz & R. Hofmann-Lehmann (2009).** "The kinetics of feline leukaemia virus shedding in experimentally infected cats are associated with infection outcome." *Vet Microbiol* **133**(3), 292-296.

**Chang, Y., E. Cesarman, M. S. Pessin, F. Lee, J. Culpepper, D. M. Knowles & P. S. Moore (1994).** "Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma." *Science* **266**(5192), 1865-1869.

**Chao, A. (1984).** "Nonparametric estimation of the number of classes in a population." *Scand J Stat* **11**(4), 265-270.

**Chappell, K. J., J. C. Brealey, A. A. Amarilla, D. Watterson, L. Hulse, C. Palmieri, S. D. Johnston, E. C. Holmes, J. Meers & P. R. Young (2017).** "Phylogenetic diversity of koala retrovirus within a wild koala population." *J Virol* **91**(3), e01820-01816.

**Chhour, K.-L., L. A. Hinds, E. M. Deane & N. A. Jacques (2008).** "The microbiome of the cloacal openings of the urogenital and anal tracts of the tammar wallaby, *Macropus eugenii*." *Microbiology* **154**(5), 1535-1543.

**Christiansen, M. T., A. C. Brown, S. Kundu, H. J. Tutill, R. Williams, J. R. Brown, J. Holdstock, M. J. Holland, S. Stevenson, J. Dave, C. Y. W. Tong, K. Einer-Jensen, D. P. Depledge & J. Breuer (2014).** "Whole-genome enrichment and sequencing of *Chlamydia trachomatis* directly from clinical samples." *BMC Infect Dis* **14**(1), 591.

**Clark, K., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell & E. W. Sayers (2016).** "GenBank." *Nucleic Acids Res* **44**(Database issue), D67-D72.

**Clark, N., N. N. Kushner, C. B. Barrett, C. R. Kensil, D. Salsbury & S. Cotter (1991).** "Efficacy and safety field trials of a recombinant DNA vaccine against feline leukemia virus infection." *J Am Vet Med Assoc* **199**(10), 1433-1443.

**Cockram, F. A. & A. R. B. Jackson (1974).** "Isolation of a *Chlamydia* from cases of keratoconjunctivitis in koalas." *Aust Vet J* **50**(2), 82-83.

**Cockram, F. A. & A. R. B. Jackson (1981).** "Keratoconjunctivitis of the koala, *Phascolarctos cinereus*, caused by *Chlamydia psittaci*." *J Wildl Dis* **17**(4), 497-504.

**Coffin, J. M., S. H. Hughes & H. E. Varmus (1997).** *Retroviruses*. New York, USA, Cold Spring Harber Laboratory Press.

**Collier, L. H. & J. Sowa (1958).** "Isolation of trachoma virus in embryonate eggs." *The Lancet* **271**(7028), 993-996.

**Connolly, J. H., P. J. Canfield, S. Hemsley & A. J. Spencer (1998).** "Lymphoid neoplasia in the koala." *Aust Vet J* **76**(12), 819-825.

**Connolly, J. H., M. B. Krockenberger, R. Malik, P. J. Canfield, D. I. Wigney & D. B. Muir (1999).** "Asymptomatic carriage of *Cryptococcus neoformans* in the nasal cavity of the koala (*Phascolarctos cinereus*)." *Med Mycol* **37**(5), 331-338.

**Croucher, N. J., A. J. Page, T. R. Connor, A. J. Delaney, J. A. Keane, S. D. Bentley, J. Parkhill & S. R. Harris (2015).** "Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins." *Nucleic Acids Res* **43**(3), e13.

**Cui, J., G. Tachedjian, M. Tachedjian, E. C. Holmes, S. Zhang & L.-F. Wang (2012).** "Identification of diverse groups of endogenous gammaretroviruses in mega and microbats." *J Gen Virol* **93**, 2037-2045.

**Cunningham, F. G., J. C. Hauth, L. C. Gilstrap, W. N. P. Herbert & S. S. Kappus (1978).** "The bacterial pathogenesis of acute pelvic inflammatory disease." *Obstet Gynecol* **52**(2), 161-164.

**Curtis, K. A., D. L. Rudolph & S. M. Owen (2008).** "Rapid detection of HIV-1 by reverse-transcription, loop-mediated isothermal amplification (RT-LAMP)." *J Virol Methods* **151**(2), 264-270.

**Darriba, D., G. L. Taboada, R. Doallo & D. Posada (2012).** "jModelTest 2: more models, new heuristics and parallel computing." *Nat Methods* **9**(8), 772-772.

**Davies, N., G. Gramotnev, L. Seabrook, A. Bradley, G. Baxter, J. Rhodes, D. Lunney & C. McAlpine (2013).** "Movement patterns of an arboreal marsupial at the edge of its range: a case study of the koala." *Movement Ecology* **1** DOI: 10.1186/2051-3933-1-8.

**Davis, D., C. Trisdell & M. Hardy (2001).** The role of economics in managing wildlife tourism. *Status Assessment of Wildlife Tourism in Australia*. Higginbottom, K., K. Rann & D. Davis, CRC Tourism. **3**.

**Day, T., N. Mideo & S. Alizon (2008).** "Why is HIV not vector-borne?" *Evolutionary Applications* **1**, 17-27.

**de Jong, M. F. C., R. Soetekouw, R. W. ten Kate & D. Veenendaal (2010).** "*Aerococcus urinae*: severe and fatal bloodstream infections and endocarditis." *J Clin Microbiol* **48**(9), 3445-3447.

**de Oliveira, S. M., P. J. Murray, D. L. de Villiers & G. S. Baxter (2014).** "Ecology and movement of urban koalas adjacent to linear infrastructure in coastal south-east Queensland." *Aust Mammal* **36**(1), 45-54.

**Dean, D. & R. S. Stephens (1994).** "Identification of individual genotypes of *Chlamydia trachomatis* from experimentally mixed serovars and mixed infections among trachoma patients." *J Clin Microbiol* **32**(6), 1506-1510.

**Delassus, S., P. Sonigo & S. Wain-Hobson (1989).** "Genetic organization of gibbon ape leukemia virus." *Virology* **173**(1), 205-213.

**Delchambre, M., D. Gheysen, D. Thines, C. Thiriart, E. Jacobs, E. Verdin, M. Horth, A. Burny & F. Bex (1989).** "The GAG precursor of simian immunodeficiency virus assembles into virus-like particles." *The EMBO Journal* **8**(9), 2653-2660.

**Department of Environment and Resource Management (2009).** Decline of the koala coast koala population: population status in 2008. Brisbane, Australia, Department of Environment and Resource Management.

**DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu & G. L. Andersen (2006).** "Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB." *Appl Environ Microbiol* **72**(7), 5069-5072.

**Desclozeaux, M., A. Robbins, M. Jelocnik, S. A. Khan, J. Hanger, V. Gerdts, A. Potter, A. Polkinghorne & P. Timms (2017).** "Immunization of a wild koala population with a recombinant *Chlamydia pecorum* major outer membrane protein (MOMP) or polymorphic membrane protein (PMP) based vaccine: new insights into immune response, protection and clearance." *PLoS ONE* **12**(6), e0178786.

**Devereaux, L. N., A. Polkinghorne, A. Meijer & P. Timms (2003).** "Molecular evidence for novel chlamydial infections in the koala (*Phascolarctos cinereus*)." *Syst Appl Microbiol* **26**(2), 245-253.

**Dickens, R. K. (1976).** The koala in health and disease. Refresher Course for Veterinarians Proceedings No 29, Taronga Zoo, The University of Sydney, New South Wales, Australia.

**Dique, D. S., J. Thompson, H. J. Preece, G. C. Penfold, D. L. de Villiers & R. S. Leslie (2003).** "Koala mortality on roads in south-east Queensland: the koala speed-zone trial." *Wildl Res* **30**(4), 419-426.

**Edgar, R. C. (2010).** "Search and clustering orders of magnitude faster than BLAST." *Bioinformatics* **26**(19), 2460-2461.

**Edgar, R. C. (2013).** "UPARSE: highly accurate OTU sequences from microbial amplicon reads." *Nat Methods* **10**(10), 996-998.

**Edgar, R. C. & H. Flyvbjerg (2015).** "Error filtering, pair assembly and error correction for next-generation sequencing reads." *Bioinformatics* **31**(21), 3476-3482.

**Epstein, M. A., B. G. Achong & Y. M. Barr (1964).** "Virus particles in cultured lymphoblasts from Burkitt's lymphoma." *The Lancet* **283**(7335), 702-703.

**Everett, K. D. E., R. M. Bush & A. A. Andersen (1999).** "Emended description of the order *Chlamydiales*, proposal of *Parachlamydiaceae* fam. nov. and *Simkaniaceae* fam. nov., each containing one monotypic genus, revised taxonomy of the family *Chlamydiaceae*, including a new genus and five new species, and standards for the identification of organisms." *Int J Syst Bacteriol* **49**(2), 415-440.

**Ezaki, T., Y. Kawamura, N. Li, Z. Y. Li, L. Zhao & S. Shu (2001).** "Proposal of the genera *Anaerococcus* gen. nov., *Peptoniphilus* gen. nov. and *Gallicola* gen. nov. for members of the genus *Peptostreptococcus*." *Int J Syst Evol Microbiol* **51**, 1521-1528.

**Faith, D. P. (1992).** "Conservation evaluation and phylogenetic diversity." *Biol Conserv* **61**(1), 1-10.

**Fan, H. (2014).** "Leukemogenesis by murine leukemia viruses: lessons for koala retrovirus (KoRV)." *Tech Rep Aust Mus (online)* **24**, 83-88.

**Fiebig, U., B. Dieckhoff, C. Wurzbacher, A. Moller, R. Kurth & J. Denner (2015a).** "Induction of neutralizing antibodies specific for the envelope proteins of the koala retrovirus by immunization with recombinant proteins or with DNA." *Virol J* **12**, 68.

**Fiebig, U., M. G. Hartmann, N. Bannert, R. Kurth & J. Denner (2006).** "Transspecies transmission of the endogenous koala retrovirus." *J Virol* **80**(11), 5651-5654.

**Fiebig, U., M. Keller & J. Denner (2016).** "Detection of koala retrovirus subgroup B (KoRV-B) in animals housed at European zoos." *Arch Virol* **161**(12), 3549-3553.

**Fiebig, U., M. Keller, A. Möller, P. Timms & J. Denner (2015b).** "Lack of antiviral antibody response in koalas infected with koala retroviruses (KoRV)." *Virus Res* **198**, 30-34.

**Finn, R. D., J. Clements & S. R. Eddy (2011).** "HMMER web server: interactive sequence similarity searching." *Nucleic Acids Res* **39**(Web Server issue), W29-W37.

**Flanagan, C. (2009).** *Koala rehabilitation manual*. Port Macquarie, New South Wales, Australia, Koala Preservation Society.

**Frutos, M. C., M. S. Monetti, L. G. Vaulet, M. E. Cadario, M. R. Fermepin, V. E. Ré & C. G. Cuffini (2015).** "Genetic diversity of *Chlamydia* among captive birds from central Argentina." *Avian Pathol* **44**(1), 50-56.

**Fukushi, H. & K. Hirai (1992).** "Proposal of *Chlamydia Pecorum* sp. nov. for *Chlamydia* strains derived from ruminants." *Int J Syst Bacteriol* **42**(2), 306-308.

**Funnell, O., L. Johnson, L. Woolford, W. Boardman, A. Polkinghorne & D. McLelland (2013).** "Conjunctivitis associated with *Chlamydia pecorum* in three koalas (*Phascolarctos cinereus*) in the Mount Lofty Ranges, South Australia." *J Wildl Dis* **49**(4), 1066-1069.

**Gandelman, O. A., V. L. Church, C. A. Moore, G. Kiddle, C. A. Carne, S. Parmar, H. Jalal, L. C. Tisi & J. A. H. Murray (2010).** "Novel bioluminescent quantitative detection of nucleic acid amplification in real-time." *PLoS ONE* **5**, e14155 DOI: 10.1371/journal.pone.0014155.

**Gillett, A. K. (2014).** "An examination of disease in captive Australian koalas (*Phascolarctos cinereus*) and potential links to koala retrovirus (KoRV)." *Tech Rep Aust Mus (online)* **24**, 39-45.

**Girjes, A. A., A. Hugall, D. M. Graham, T. F. McCaul & M. F. Lavin (1993).** "Comparison of type I and type II *Chlamydia psittaci* strains infecting koalas (*Phascolarctos cinereus*)." *Vet Microbiol* **37**(1–2), 65-83.

**Girjes, A. A., A. F. Hugall, P. Timms & M. F. Lavin (1988).** "Two distinct forms of *Chlamydia psittaci* associated with disease and infertility in *Phascolarctos cinereus* (koala)." *Infect Immun* **56**(8), 1897-1900.

**Glassick, T., P. Giffard & P. Timms (1996).** "Outer membrane protein 2 gene sequences indicate that *Chlamydia pecorum* and *Chlamydia pneumoniae* cause infections in koalas." *Syst Appl Microbiol* **19**(3), 457-464.

**Goldman, N. & Z. Yang (1994).** "A codon-based model of nucleotide substitution for protein-coding DNA sequences." *Mol Biol Evol* **11**(5), 725-736.

**Gomes, J. P., W. J. Bruno, A. Nunes, N. Santos, C. Florindo, M. J. Borrego & D. Dean (2007).** "Evolution of *Chlamydia trachomatis* diversity occurs by widespread interstrain recombination involving hotspots." *Genome Res* **17**(1), 50-60.

**Gordon, G., P. Menkhorst, T. Robinson, D. Lunney, R. Martin & M. Ellis. (2008).** "*Phascolarctos cinereus*." *IUCN 2013 IUCN Red List of Threatened Species Version 20132* Retrieved 26 March 2014., 2014, from http://www.iucnredlist.org/details/full/16892/0.

**Grayston, J. T., C.-C. Kuo, L. A. Campbell & S.-P. Wang (1989).** "*Chlamydia pneumoniae* sp. nov. for *Chlamydia* sp. strain TWAR." *Int J Syst Bacteriol* **39**(1), 88-90.

**Greco, G., M. Corrente, D. Buonavoglia, G. Campanile, R. Di Palo, V. Martella, A. L. Bellacicco, M. D'Abramo & C. Buonavoglia (2008).** "Epizootic abortion related to infections by *Chlamydophila abortus* and *Chlamydophila pecorum* in water buffalo (*Bubalus bubalis*)." *Theriogenology* **69**(9), 1061-1069.

**Griffith, J. E. (2010).** *Studies into the diagnosis, treatment and management of chlamydiosis in koalas*. PhD PhD Thesis, The University of Sydney.

**Griffith, J. E., N. K. Dhand, M. B. Krockenberger & D. P. Higgins (2013).** "A retrospective study of admission trends of koalas to a rehabilitation facility over 30 years." *J Wildl Dis* **49**(1), 18-28.

**Griffith, J. E. & Damien P. Higgins (2012).** "Diagnosis, treatment and outcomes for koala chlamydiosis at a rehabilitation facility (1995–2005)." *Aust Vet J* **90**(11), 457-463.

**Guindon, S. & O. Gascuel (2003).** "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood." *Syst Biol* **52**(5), 696-704.

**Gurevich, A., V. Saveliev, N. Vyahhi & G. Tesler (2013).** "QUAST: quality assessment tool for genome assemblies." *Bioinformatics* **29**(8), 1072-1075.

**Hadfield, J., N. J. Croucher, R. J. Goater, K. Abudahab, D. M. Aanensen & S. R. Harris (2017).** "Phandango: an interactive viewer for bacterial population genomics." *bioRxiv* DOI: 10.1101/119545.

**Hanger, J., J. Loader, C. Wan, K. W. Beagley, P. Timms & A. Polkinghorne (2013).** "Comparison of antigen detection and quantitative PCR in the detection of chlamydial infection in koalas (*Phascolarctos cinereus*)." *Vet J* **195**(3), 391-393.

**Hanger, J. J., L. D. Bromham, J. J. McKee, T. M. O'Brien & W. F. Robinson (2000).** "The nucleotide sequence of koala (*Phascolarctos cinereus*) retrovirus: a novel type C endogenous virus related to gibbon ape leukemia virus." *J Virol* **74**(9), 4264-4272.

**Hanger, J. J. & J. Loader (2014).** "Disease in wild koalas (*Phascolarctos cinereus*) with possible koala retrovirus involvement." *Tech Rep Aust Mus (online)* **24**, 19-29 DOI: 10.3853/j.1835-4211.24.2014.1609.

**Harris, S. R., I. N. Clarke, H. M. B. Seth-Smith, A. W. Solomon, L. T. Cutcliffe, P. Marsh, R. J. Skilton, M. J. Holland, D. Mabey, R. W. Peeling, D. A. Lewis, B. G. Spratt, M. Unemo, K. Persson, C. Bjartling, R. Brunham, H. J. C. de Vries, S. A. Morre, A. Speksnijder, C. M. Bebear, M. Clerc, B. de Barbeyrac, J. Parkhill & N. R. Thomson (2012).** "Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing." *Nat Genet* **44**(4), 413-419.

**Hartley, M. & A. Sainsbury (2017).** "Methods of disease risk analysis in wildlife translocations for conservation purposes." *EcoHealth* **14**(1), 16-29.

**Hasegawa, M., H. Kishino & T.-a. Yano (1985).** "Dating of the human-ape splitting by a molecular clock of mitochondrial DNA." *J Mol Evol* **22**(2), 160-174.

**Hatch, T. P. (1996).** "Disulfide cross-linked envelope proteins: the functional equivalent of peptidoglycan in *Chlamydiae*?" *J Bacteriol* **178**(1), 1-5.

**Hatch, T. P., E. Al-Hossainy & J. A. Silverman (1982).** "Adenine nucleotide and lysine transport in *Chlamydia psittaci*." *J Bacteriol* **150**(2), 662-670.

**Hatch, T. P., I. Allan & J. H. Pearce (1984).** "Structural and polypeptide differences between envelopes of infective and reproductive life cycle forms of *Chlamydia* spp." *J Bacteriol* **157**(1), 13-20.

**Hatch, T. P., D. W. Vance, Jr. & E. Al-Hossainy (1981).** "Identification of a major envelope protein in *Chlamydia* spp." *J Bacteriol* **146**(1), 426-429.

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

**Hayward, W. S., B. G. Neel & S. M. Astrin (1981).** "Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis." *Nature* **290**(5806), 475-480.

**Hemsley, S. & P. J. Canfield (1997).** "Histopathological and immunohistochemical investigation of naturally occurring chlamydial conjunctivitis and urogenital inflammation in koalas (*Phascolarctos cinereus*)." *J Comp Pathol* **116**(3), 273-290.

**Henikoff, S. & J. G. Henikoff (1992).** "Amino acid substitution matrices from protein blocks." *Proc Natl Acad Sci USA* **89**(22), 10915-10919.

**Heuschele, W. P. & J. R. Hayes (1961).** "Acute leukemia in a New South Wales koala (*Phascolarctos c. cinereus*)." *Cancer Res* **21**(10), 1394-1395.

**Hewson, I., J. B. Button, B. M. Gudenkauf, B. Miner, A. L. Newton, J. K. Gaydos, J. Wynne, C. L. Groves, G. Hendler, M. Murray, S. Fradkin, M. Breitbart, E. Fahsbender, K. D. Lafferty, A. M. Kilpatrick, C. M. Miner, P. Raimondi, L. Lahner, C. S. Friedman, S. Daniels, M. Haulena, J. Marliave, C. A. Burge, M. E. Eisenlord & C. D. Harvell (2014).** "Densovirus associated with sea-star wasting disease and mass mortality." *Proc Natl Acad Sci USA* **111**(48), 17278-17283.

**Higgins, D. P., T. Beninati, M. Meek, J. Irish & J. E. Griffith (2012).** "Within-population diversity of koala *Chlamydophila pecorum* at *ompA* VD1-VD3 and the ORF663 hypothetical gene." *Vet Microbiol* **156**(3–4), 353-358.

**Higgins, D. P., S. Hemsley & P. J. Canfield (2005).** "Immuno-histochemical demonstration of the role of *Chlamydiaceae* in renal, uterine and salpingeal disease of the koala, and demonstration of *Chlamydiaceae* in novel sites." *J Comp Pathol* **133**(2–3), 164-174.

**Hobbs, M., A. Pavasovic, A. G. King, P. J. Prentis, M. D. B. Eldridge, Z. Chen, D. J. Colgan, A. Polkinghorne, M. R. Wilkins, C. Flanagan, A. Gillett, J. Hanger, R. N. Johnson & P. Timms (2014).** "A transcriptome resource for the koala (*Phascolarctos cinereus*): insights into koala retrovirus transcription and sequence diversity." *BMC Genomics* **15**, 786.

**Hogg, C. J., A. V. Lee, C. Srb & C. Hibbard (2016).** "Metapopulation management of an Endangered species with limited genetic diversity in the presence of disease: the Tasmanian devil *Sarcophilus harrisii*." *Int Zoo Yearb* **51**, 137-153.

**Holland, S. M., A. P. Hudson, L. Bobo, J. A. Whittum-Hudson, R. P. Viscidi, T. C. Quinn & H. R. Taylor (1992).** "Demonstration of chlamydial RNA and DNA during a culture-negative state." *Infect Immun* **60**(5), 2040-2047.

**Holzwarth, N., A. Pospischil, N. Marreros, M.-P. Ryser-Degiorgis, F. Mavrot, J. Frey, R. Thoma & N. Borel (2011).** "Alpine ibex (*Capra i. ibex*) is not a reservoir for chlamydial infections of domestic ruminants and humans." *Eur J Wildl Res* **57**(2), 233-240.

**Houlden, B. A., B. H. Costello, D. Sharkey, E. V. Fowler, A. Melzer, W. Ellis, F. Carrick, P. R. Baverstock & M. S. Elphinstone (1999).** "Phylogeographic differentiation in the mitochondrial control region in the koala, *Phascolarctos cinereus* (Goldfuss 1817)." *Mol Ecol* **8**(6), 999-1011.

**Houlden, B. A., P. R. England, A. C. Taylor, W. D. Greville & W. B. Sherwin (1996).** "Low genetic variability of the koala *Phascolarctos cinereus* in south-eastern Australia following a severe population bottleneck." *Mol Ecol* **5**(2), 269-281.

**Hrdina, F. & G. Gordon (2004).** "The koala and possum trade in Queensland, 1906-1936." *Aust Zool* **32**(4), 543-585.
**Huelsenbeck, J. P. & F. Ronquist (2001).** "MRBAYES: Bayesian inference of phylogenetic trees." *Bioinformatics* **17**(8), 754-755.
**Huerta-Cepas, J., F. Serra & P. Bork (2016).** "ETE 3: reconstruction, analysis, and visualization of phylogenomic data." *Mol Biol Evol* **33**(6), 1635-1638.
**Human Microbiome Project Consortium (2012).** "Structure, function and diversity of the healthy human microbiome." *Nature* **486**(7402), 207-214.
**Hundloe, T., C. Hamilton & L. Wilks (1997).** *Koalas and tourism: an economic evaluation*, The Australia Institute.
**Hyatt, D., G.-L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer & L. J. Hauser (2010).** "Prodigal: prokaryotic gene recognition and translation initiation site identification." *BMC Bioinformatics* **11**, 119.
**Hybiske, K. & R. S. Stephens (2007).** "Mechanisms of host cell exit by the intracellular bacterium *Chlamydia*." *Proc Natl Acad Sci USA* **104**(27), 11430-11435.
**Ishida, Y., C. McCallister, N. Nikolaidis, K. Tsangaras, K. M. Helgen, A. D. Greenwood & A. L. Roca (2015a).** "Sequence variation of koala retrovirus transmembrane protein p15E among koalas from different geographic regions." *Virology* **475**, 28-36.
**Ishida, Y., K. Zhao, A. D. Greenwood & A. L. Roca (2015b).** "Proliferation of endogenous retroviruses in the early stages of a host germ line invasion." *Mol Biol Evol* **32**(1), 109-120.
**Jackson, M., P. Giffard & P. Timms (1997).** "Outer membrane protein A gene sequencing demonstrates the polyphyletic nature of koala *Chlamydia pecorum* isolates." *Syst Appl Microbiol* **20**(2), 187-200.
**Jackson, M., N. White, P. Giffard & P. Timms (1999).** "Epizootiology of *Chlamydia* infections in two free-range koala populations." *Vet Microbiol* **65**(4), 255-264.
**Jarrett, W. F. H., E. M. Crawford, W. B. Martin & F. Davie (1964).** "Leukaemia in the cat: a virus-like particle associated with leukaemia (lymphosarcoma)." *Nature* **202**(4932), 567-568.
**Jeffares, D. C., B. Tomiczek, V. Sojo & M. dos Reis (2015).** "A beginners guide to estimating the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome." *Methods Mol Biol* **1201**, 65-90.
**Jelocnik, M., N. Bachmann, B. Kaltenboeck, C. Waugh, L. Woolford, K. N. Speight, A. Gillett, D. P. Higgins, C. Flanagan, G. S. A. Myers, P. Timms & A. Polkinghorne (2015).** "Genetic diversity in the plasticity zone and the presence of the chlamydial plasmid differentiates *Chlamydia pecorum* strains from pigs, sheep, cattle, and koalas." *BMC Genomics* **16**, 893.
**Jelocnik, M., N. L. Bachmann, H. Seth-Smith, N. R. Thomson, P. Timms & A. M. Polkinghorne (2016).** "Molecular characterisation of the *Chlamydia pecorum* plasmid from porcine, ovine, bovine, and koala strains indicates plasmid-strain co-evolution." *PeerJ* **4** DOI: 10.7717/peerj.1661.
**Jelocnik, M., D. Forshaw, J. Cotter, D. Roberts, P. Timms & A. Polkinghorne (2014a).** "Molecular and pathological insights into *Chlamydia pecorum*-associated sporadic bovine encephalomyelitis (SBE) in Western Australia." *BMC Vet Res* **10**, 121.
**Jelocnik, M., F. D. Frentiu, P. Timms & A. Polkinghorne (2013).** "Multilocus sequence analysis provides insights into molecular epidemiology of *Chlamydia pecorum* infections in Australian sheep, cattle, and koalas." *J Clin Microbiol* **51**(8), 2625-2632.
**Jelocnik, M., E. Walker, Y. Pannekoek, J. Ellem, P. Timms & A. Polkinghorne (2014b).** "Evaluation of the relationship between *Chlamydia pecorum* sequence types and disease using a species-specific multi-locus sequence typing scheme (MLST)." *Vet Microbiol* **174**(1–2), 214-222.

**Johnson, W. E. & J. M. Coffin (1999).** "Constructing primate phylogenies from ancient retrovirus sequences." *Proc Natl Acad Sci USA* **96**(18), 10254-10260.

**Jolley, K. & M. Maiden (2010).** "BIGSdb: Scalable analysis of bacterial genome variation at the population level." *BMC Bioinformatics* **11**, 595.

**Kapos, V., A. Balmford, R. Aveling, P. Bubb, P. Carey, A. Entwistle, J. Hopkins, T. Mulliken, R. Safford, A. Stattersfield, M. Walpole & A. Manica (2008).** "Calibrating conservation: new tools for measuring success." *Conservation Letters* **1**(4), 155-164.

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

**Kawai, Y., N. Miyashita, F. Kishi, M. Tabuchi, K. Oda, T. Yamaguchi, K. Kawasaki, T. Yamazaki & K. Ouchi (2009).** "Development and evaluation of a loop-mediated isothermal amplification method for the rapid detection of *Chlamydophila pneumoniae*." *Eur J Clin Microbiol Infect Dis* **28**(7), 801-805.

**Kawakami, T. G., S. D. Huff, P. M. Buckley, D. L. Dungworth, S. P. Synder & R. V. Gilden (1972).** "C-type virus associated with gibbon lymphosarcoma." *Nat New Biol* **235**(58), 170-171.

**Kawakami, T. G., L. Sun & T. S. McDowell (1977).** "Infectious primate type-C virus shed by healthy gibbons." *Nature* **268**(5619), 448-450.

**Keat, A. C., B. J. Thomas, D. Taylor-Robinson, G. D. Pegrum, R. N. Maini & J. T. Scott (1980).** "Evidence of *Chlamydia trachomatis* infection in sexually acquired reactive arthritis." *Ann Rheum Dis* **39**(5), 431-437.

**Keller, M., B. Petersen, H. Niemann & J. Denner (2014).** "Lack of antibody response in pigs immunized with the transmembrane envelope protein of porcine endogenous retroviruses." *J Gen Virol* **95**(8), 1827-1831.

**Khan, S. A., C. Waugh, G. Rawlinson, J. Brumm, K. Nilsson, V. Gerdts, A. Potter, A. Polkinghorne, K. Beagley & P. Timms (2014).** "Vaccination of koalas (*Phascolarctos cinereus*) with a recombinant chlamydial major outer membrane protein adjuvanted with poly I:C, a host defense peptide and polyphosphazine, elicits strong and long lasting cellular and humoral immune responses." *Vaccine* **32**(44), 5781-5786.

**Klebanoff, S. J. & R. W. Coombs (1991).** "Viricidal effect of *Lactobacillus acidophilus* on Human Immunodeficiency Virus Type 1: possible role in heterosexual transmission." *J Exp Med* **174**(1), 289-292.

**Kolbe, D. L. & S. R. Eddy (2011).** "Fast filtering for RNA homology search." *Bioinformatics* **27**(22), 3102-3109.

**Kollipara, A. (2013).** *Investigation of genetic diversity and development of vaccine for Chlamydia pecorum infections in koala (Phascolarctos cinereus)*. PhD, Queensland University of Technology.

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

**Kollipara, A., A. Polkinghorne, K. W. Beagley & P. Timms (2013a).** "Vaccination of koalas with a recombinant *Chlamydia pecorum* major outer membrane protein induces antibodies of different specificity compared to those following a natural live infection." *PLoS ONE* **8**(9), e74808.

**Kollipara, A., A. Polkinghorne, C. Wan, P. Kanyoka, J. Hanger, J. Loader, J. Callaghan, A. Bell, W. Ellis, S. Fitzgibbon, A. Melzer, K. Beagley & P. Timms (2013b).** "Genetic diversity of *Chlamydia pecorum* strains in wild koala locations across Australia and the implications for a recombinant *C. pecorum* major outer membrane protein based vaccine." *Vet Microbiol* **167**(3–4), 513-522.

**Kollipara, A., C. Wan, G. Rawlinson, J. Brumm, K. Nilsson, A. Polkinghorne, K. Beagley & P. Timms (2013c).** "Antigenic specificity of a monovalent versus polyvalent MOMP based *Chlamydia pecorum* vaccine in koalas (*Phascolarctos cinereus*)." *Vaccine* **31**(8), 1217-1223.

**Kontchou, C. Y. & R. Blondeau (1990).** "Isolation and characterization of hydrogen peroxide producing *Aerococcus* sp. from soil samples." *FEMS Microbiol Lett* **68**(3), 323-327.

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

**Krockenberger, M. B., P. J. Canfield & R. Malik (2002).** "*Cryptococcus neoformans* in the koala (*Phascolarctos cinereus*): colonization by *C. n.* var. *gattii* and investigation of environmental sources." *Med Mycol* **40**(3), 263-272.

**Langhammer, S., J. Hübner, R. Kurth & J. Denner (2006).** "Antibodies neutralizing feline leukaemia virus (FeLV) in cats immunized with the transmembrane envelope protein p15E." *Immunology* **117**(2), 229-237.

**Laslett, D. & B. Canback (2004).** "ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences." *Nucleic Acids Res* **32**(1), 11-16.

**Lau, Q., W. Jaratlerdsiri, J. E. Griffith, J. Gongora & D. P. Higgins (2014).** "MHC class II diversity of koala (*Phascolarctos cinereus*) populations across their range." *Heredity* **113**(4), 287-296.

**Lee, K. E., J. M. Seddon, S. W. Corley, W. A. H. Ellis, S. D. Johnston, D. L. de Villiers, H. J. Preece & F. N. Carrick (2010).** "Genetic variation and structuring in the threatened koala populations of Southeast Queensland." *Conserv Genet* **11**(6), 2091-2103.

**Legione, A. R., J. Amery-Gale, M. Lynch, L. Haynes, J. R. Gilkerson, F. M. Sansom & J. M. Devlin (2016a).** "*Chlamydia pecorum* infection in free-ranging koalas (*Phascolarctos cinereus*) on French Island, Victoria, Australia." *J Wildl Dis* **52**(2), 426-429.

**Legione, A. R., J. L. S. Patterson, P. Whiteley, S. M. Firestone, M. Curnick, K. Bodley, M. Lynch, J. R. Gilkerson, F. M. Sansom & J. M. Devlin (2017).** "Koala retrovirus genotyping analyses reveal a low prevalence of KoRV-A in Victorian koalas and an association with clinical disease." *J Med Microbiol* **66**(2), 236-244.

**Legione, A. R., J. L. S. Patterson, P. L. Whiteley, J. Amery-Gale, M. Lynch, L. Haynes, J. R. Gilkerson, A. Polkinghorne, J. M. Devlin & F. M. Sansom (2016b).** "Identification of unusual *Chlamydia pecorum* genotypes in Victorian koalas (*Phascolarctos cinereus*) and clinical variables associated with infection." *J Med Microbiol* **65**(5), 420-428.

**Lenz, J., D. Celander, R. L. Crowther, R. Patarca, D. W. Perkins & W. A. Haseltine (1984).** "Determination of the leukaemogenicity of a murine retrovirus by sequences within the long terminal repeat." *Nature* **308**(5958), 467-470.

**Lewis, F. (1954).** "The rehabilitation of the koala in Victoria." *Vic Nat* **70**, 197-200.

**Li, H. & R. Durbin (2010).** "Fast and accurate long-read alignment with Burrows-Wheeler transform." *Bioinformatics* **26**(5), 589-595.

**Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis & R. Durbin (2009).** "The sequence alignment/map format and SAMtools." *Bioinformatics* **25**(16), 2078-2079.

**Li, J., W. Guo, B. Kaltenboeck, K. Sachse, Y. Yang, G. Lu, J. Zhang, L. Luan, J. You, K. Huang, H. Qiu, Y. Wang, M. Li, Z. Yang & C. Wang (2016).** "*Chlamydia pecorum* is the endemic intestinal species in cattle while *C. gallinacea*, *C. psittaci* and *C. pneumoniae* associate with sporadic systemic infection." *Vet Microbiol* **193**, 93-99.

**Li, Y., E. Golemis, J. W. Hartley & N. Hopkins (1987).** "Disease specificity of nondefective Friend and Moloney murine leukemia viruses is controlled by a small number of nucleotides." *J Virol* **61**(3), 693-700.

**Lieber, M. M., C. J. Sherr, G. J. Todaro, R. E. Benveniste, R. Callahan & H. G. Coon (1975).** "Isolation from the asian mouse *Mus caroli* of an endogenous type C virus related to infectious primate type C viruses." *Proc Natl Acad Sci USA* **72**(6), 2315-2319.

**Liu, B., G. Eydallin, R. P. Maharjan, L. Feng, L. Wang & T. Ferenci (2017).** "Natural *Escherichia coli* isolates rapidly acquire genetic changes upon laboratory domestication." *Microbiology* **163**(1), 22-30.

**Lockhart, A. B., P. H. Thrall & J. Antonovics (1996).** "Sexually transmitted diseases in animals: ecological and evolutionary implications." *Biological Reviews* **71**(3), 415-471.

**Love, M. I., W. Huber & S. Anders (2014).** "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biol* **15**, 550.

**Lozupone, C. & R. Knight (2005).** "UniFrac: a new phylogenetic method for comparing microbial communities." *Appl Environ Microbiol* **71**(12), 8228-8235.

**Lunney, D., S. Gresser, L. E. O'Neill, A. Matthews & J. Rhodes (2007).** "The impact of fire and dogs on koalas at Port Stephens, New South Wales, using population viability analysis." *Pac Conserv Biol* **13**(3), 189-201.

**Mackie, J. T., A. K. Gillett, C. Palmieri, T. Feng & D. P. Higgins (2016).** "Pneumonia due to *Chlamydia pecorum* in a koala (*Phascolarctos cinereus*)." *J Comp Pathol* **155**(4), 356-360.

**Madsen, M., G. H. Sørensen & B. Aalbaek (1990).** "Summer mastitis in heifers: a bacteriological examination of secretions from clinical cases of summer mastitis in Denmark." *Vet Microbiol* **22**(4), 319-328.

**Maher, I. E. & D. P. Higgins (2016).** "Altered immune cytokine expression associated with KoRV B infection and season in captive koalas." *PLoS ONE* **11**(10), e0163780.

**Maiden, M. C. J., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman & B. G. Spratt (1998).** "Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms." *Proc Natl Acad Sci USA* **95**(6), 3140-3145.

**Mårdh, P.-A., T. Ripa, L. Svensson & L. Weström (1977).** "*Chlamydia trachomatis* infection in patients with acute salpingitis." *New Engl J Med* **296**(24), 1377-1379.

**Markey, B., C. Wan, J. Hanger, C. Phillips & P. Timms (2007).** "Use of quantitative real-time PCR to monitor the shedding and treatment of *Chlamydiae* in the koala (*Phascolarctos cinereus*)." *Vet Microbiol* **120**(3–4), 334-342.

**Marschner, C., C. Flanagan, D. P. Higgins & M. B. Krockenberger (2014).** "Validation of ultrasonography in detecting structural disease of the urogenital tract of the koala, *Phascolarctos cinereus*." *Aust Vet J* **92**(5), 177-178.

**Marsh, J., A. Kollipara, P. Timms & A. Polkinghorne (2011).** "Novel molecular markers of *Chlamydia pecorum* genetic diversity in the koala (*Phascolarctos cinereus*)." *BMC Microbiol* **11**, 77.

**Martin, H. L., B. A. Richardson, P. M. Nyange, L. Lavreys, S. L. Hillier, B. Chohan, K. Mandaliya, J. O. Ndinya-Achola, J. Bwayo & J. Kreiss (1999a).** "Vaginal *Lactobacilli*, microbial flora, and risk of Human Immunodeficiency Virus Type 1 and sexually transmitted disease acquisition." *J Infect Dis* **180**(6), 1863-1868.

**Martin, J., E. Herniou, J. Cook, R. W. O'Neill & M. Tristem (1999b).** "Interclass transmission and phyletic host tracking in murine leukemia virus-related retroviruses." *J Virol* **73**(3), 2442-2449.

**Martin, R. W. (1981).** "Age-specific fertility in three populations of the koala, *Phascolarctos cinereus* Goldfuss, in Victoria." *Wildl Res* **8**(2), 275-283.

**Martin, R. W. (1985a).** "Overbrowsing, and decline of a population of the koala, *Phascolarctos cinereus*, in Victoria. 1. Food preference and food tree defoliation." *Wildl Res* **12**(3), 355-365.

**Martin, R. W. (1985b).** "Overbrowsing, and decline of a population of the koala, *Phascolarctos cinereus*, in Victoria. 2. Population condition." *Wildl Res* **12**(3), 367-375.

**Martin, R. W. (1989).** Draft managment plan for the conservation of the koala (*Phascolarctos cinereus*) in Victoria. *Technical Report Series*. Melbourne, Vic, Australia, Arthur Rylah Institute for Environmental Research.

**Martin, R. W. & K. A. Handasyde (1990).** *Translocations and the re-establishment of koala populations in Victoria (1944–1988): the implications for New South Wales*. Koala Summit: managing koalas in New South Wales, New South Wales National Parks and Wildlife Service, Sydney.

**Martin, R. W. & K. A. Handasyde (1999).** *The koala: natural history, conservation and management*. Sydney, NSW, Australia, UNSW Press Ltd.

**Martín, V., A. I. Vela, M. Gilbert, J. Cebolla, J. Goyache, L. Domínguez & J. F. Fernández-Garayzábal (2007).** "Characterization of *Aerococcus viridans* isolates from swine clinical specimens." *J Clin Microbiol* **45**(9), 3053-3057.

**Masters, P., T. Duka, S. Berris & G. Moss (2004).** "Koalas on Kangaroo Island: from introduction to pest status in less than a century." *Wildl Res* **31**(3), 267-272.

**Mathew, M., C. Waugh, K. W. Beagley, P. Timms & A. Polkinghorne (2014).** "Interleukin 17A is an immune marker for chlamydial disease severity and pathogenesis in the koala (*Phascolarctos cinereus*)." *Dev Comp Immunol* **46**(2), 423-429.

**Mattingley, A. (1901).** "Natural history notes." *Vic Nat* **18**, 55.

**McColl, K. A., R. W. Martin, L. J. Gleeson, K. A. Handasyde & A. K. Lee (1984).** "*Chlamydia* infection and infertility in the female koala (*Phascolarctos cinereus*)." *Vet Rec* **115**(25-26), 655.

**McFadden, M., D. Hunter, M. Evans, B. Scheele, R. Pietsch & P. Harlow (2016).** Re-introduction of the northern corroboree frog in the Northern Brindabella Mountains, New South Wales, Australia. *Global Re-introduction Perspectives: 2016 Case-studies from around the globe*. Soorae, P. S. Gland, Switzerland & Abu Dhabi, UAE, IUCN/SSC Re-introduction Specialist Group & Environment Agency-Abu Dhabi**:** 35 - 39.

**McInnes, L. M., A. Gillett, J. Hanger, S. A. Reid & U. M. Ryan (2011).** "The potential impact of native Australian trypanosome infections on the health of koalas (*Phascolarctos cinereus*)." *Parasitology* **138**(7), 873-883.

**McKean, S. C., J. K. Davies & R. J. Moore (2007).** "Expression of phospholipase D, the major virulence factor of *Corynebacterium pseudotuberculosis*, is regulated by multiple environmental factors and plays a role in macrophage death." *Microbiology* **153**(7), 2203-2211.

**McKenzie, R. A. (1981).** "Observations on diseases of free-living and captive koalas (*Phascolarctos cinereus*)." *Aust Vet J* **57**(5), 243-246.

**McLean, N. (2003).** *Ecology and management of overabundant koala (Phascolarctos cinereus) populations*. PhD, University of Melbourne.

**McLean, N. & K. A. Handasyde (2007).** "Sexual maturity, factors affecting the breeding season and breeding in consecutive seasons in populations of overabundant Victorian koalas (*Phascolarctos cinereus*)." *Aust J Zool* **54**(6), 385-392.

**McMurdie, P. J. & S. Holmes (2014).** "Waste not, want not: why rarefying microbiome data is inadmissible." *PLoS Comp Biol* **10**(4), e1003531.

**McNutt, S. H. & E. F. Waller (1940).** "Sporadic bovine encephalomyelitis (Buss disease)." *Cornell Vet* **30**, 437-448.

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

**Mendoza, R., M. M. Anderson & J. Overbaugh (2006).** "A putative thiamine transport protein is a receptor for feline leukemia virus subgroup A." *J Virol* **80**(7), 3378-3385.

**Menkhorst, P. (2004).** *Victoria's koala management strategy*. Victoria, Australia, Department of Sustainability and Environment.

**Meyer, K. F. & B. Eddie (1933).** "Latent psittacosis infections in shell parrakeets." *Proc Soc Exp Biol Med* **30**(4), 484-488.

**Miller, A. D., L. Bonham, J. Alfano, H. P. Kiem, T. Reynolds & G. Wolgamot (1996).** "A novel murine retrovirus identified during testing for helper virus in human gene transfer trials." *J Virol* **70**(3), 1804-1809.

**Minoche, A. E., J. C. Dohm & H. Himmelbauer (2011).** "Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems." *Genome Biol* **12** DOI: 10.1186/gb-2011-12-11-r112.

**Mitchell, P. J., S. McOrist & R. Bilney (1987).** "Epidemiology of *Mycobacterium ulcerans* infection in koalas (*Phascolarctos cinereus*) on Raymond Island, southeastern Australia." *J Wildl Dis* **23**(3), 386-390.

**Miyazawa, T., T. Shojima, R. Yoshikawa & T. Ohata (2011).** "Isolation of koala retroviruses from koalas in Japan." *J Vet Med Sci* **73**(1), 65 - 70.

**Mojica, S., H. Huot Creasy, S. Daugherty, T. D. Read, T. Kim, B. Kaltenboeck, P. Bavoil & G. S. A. Myers (2011).** "Genome sequence of the obligate intracellular animal pathogen *Chlamydia pecorum* E58." *J Bacteriol* **193**(14), 3690.

**Moloney, J. B. (1960).** "Biological studies on a lymphoid-leukemia virus extracted from sarcoma 37. 1. Origin and introductory investigations." *J Natl Cancer Inst* **24**(4), 933-951.

**Moreno, P. S., J. Wagner, C. S. Mansfield, M. Stevens, J. R. Gilkerson & C. D. Kirkwood (2017).** "Characterisation of the canine faecal virome in healthy dogs and dogs with acute diarrhoea using shotgun metagenomics." *PLoS ONE* **12**(6), e0178433.

**Morgan, X. C., T. L. Tickle, H. Sokol, D. Gevers, K. L. Devaney, D. V. Ward, J. A. Reyes, S. A. Shah, N. LeLeiko, S. B. Snapper, A. Bousvaros, J. Korzenik, B. E. Sands, R. J. Xavier & C. Huttenhower (2012).** "Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment." *Genome Biol* **13**, R79 DOI: 10.1186/gb-2012-13-9-r79.

**Morgulis, A., G. Coulouris, Y. Raytselis, T. L. Madden, R. Agarwala & A. A. Schaffer (2008).** "Database indexing for production MegaBLAST searches." *Bioinformatics* **24**(16), 1757-1764.

**Morris, K. M., D. O'Meally, T. Zaw, X. Song, A. Gillett, M. P. Molloy, A. Polkinghorne & K. Belov (2016).** "Characterisation of the immune compounds in koala milk using a combined transcriptomic and proteomic approach." *Sci Rep* **6**, 35011 DOI: 10.1038/srep35011.

**Moscicki, A.-B., J. H. Ellenberg, S. H. Vermund, C. A. Holland, T. Darragh, P. A. Crowley-Nowick, L. Levin & C. M. Wilson (2000).** "Prevalence of and risks for cervical human papillomavirus infection and squamous intraepithelial lesions in adolescent girls: impact of infection with human immunodeficiency virus." *Arch Pediatr Adolesc Med* **154**(2), 127-134.

**Moulder, J. W. (1991).** "Interaction of *Chlamydiae* and host cells in vitro." *Microbiol Rev* **55**(1), 143-190.

**Moulder, J. W., N. J. Levy & L. P. Schulman (1980).** "Persistent infection of mouse fibroblasts (L cells) with *Chlamydia psittaci*: evidence for a cryptic chlamydial form." *Infect Immun* **30**(3), 874-883.

**Mounsey, K. E., H. C. Murray, H. Bielefeldt-Ohmann, C. Pasay, D. C. Holt, B. J. Currie, S. F. Walton & J. S. McCarthy (2015).** "Prospective study in a porcine model of *Sarcoptes scabiei* indicates the association of Th2 and Th17 pathways with the clinical severity of scabies." *PLoS Negl Trop Dis* **9**(3), e0003498.

**Muñoz, N., F. X. Bosch, S. de Sanjosé, R. Herrero, X. Castellsagué, K. V. Shah, P. J. F. Snijders & C. J. L. M. Meijer (2003).** "Epidemiologic classification of human papillomavirus types associated with cervical cancer." *New Engl J Med* **2003**(348), 518-527.

**Murthy, A. K., B. P. Arulanandam & G. Zhong (2012).** *Chlamydia* vaccine: progress and challenges. *Intracellular Pathogens I: Chlamydiales*. Tan, M. & P. Bauvoil. Washington, DC, USA, ASM Press**:** 311-333.

**Neaves, L. E., G. J. Frankham, S. Dennison, S. FitzGibbon, C. Flannagan, A. Gillett, E. Hynes, K. Handasyde, K. M. Helgen, K. Tsangaras, A. D. Greenwood, M. D. B. Eldridge & R. N. Johnson (2016).** "Phylogeography of the koala, (*Phascolarctos cinereus*), and harmonising data to inform conservation." *PLoS ONE* **11**(9).

**Nielsen, R. & Z. Yang (1998).** "Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene." *Genetics* **148**(3), 929-936.

**Notomi, T., H. Okayama, H. Masubuchi, T. Yonekawa, K. Watanabe, N. Amino & T. Hase (2000).** "Loop-mediated isothermal amplification of DNA." *Nucleic Acids Res* **28**(12), e63-e63.

**O'Brien, S. J. & J. F. Evermann (1988).** "Interactive influence of infectious disease and genetic diversity in natural populations." *Trends Ecol Evol* **3**(10), 254-259.

**O'Brien, S. J., M. E. Roelke, L. Marker, A. Newman, C. A. Winkler, D. Meltzer, L. Colly, J. F. Evermann, M. Bush & D. E. Wildt (1985).** "Genetic basis for species vulnerability in the cheetah." *Science* **227**(4693), 1428-1434.

**O'Brien, T. M., J. J. Hanger, J. J. McKee & W. F. Robinson (1997).** *The isolation, characterisation and partial gene sequence of a retrovirus from koalas*. A Conference on the Status of the Koala, Soldiers Point, Port Stephens, NSW, Australian Koala Foundation.

**O'Hara, B., S. V. Johann, H. P. Klinger, D. G. Blair, H. Rubinson, K. J. Dunn, P. Sass, S. M. Vitek & T. Robins (1990).** "Characterization of a human gene conferring sensitivity to infection by gibbon ape leukemia virus." *Cell Growth Differ* **1**(3), 119-127.

**Obendorf, D. L. (1981).** "Pathology of the female reproductive tract in the koala, *Phascolarctos cinereus* (Goldfuss), from Victoria, Australia." *J Wildl Dis* **17**(4), 587-592.

**Obendorf, D. L. (1983).** "Causes of mortality and morbidity of wild koalas, *Phascolarctos cinereus* (Goldfuss), in Victoria, Australia." *J Wildl Dis* **19**(2), 123-131.

**Oliveira, N. M., K. B. Farrell & M. V. Eiden (2006).** "*In vitro* characterization of a koala retrovirus." *J Virol* **80**(6), 3104-3107.

**Ooshiro, M., S. Konnai, Y. Katagiri, M. Afuso, N. Arakaki, O. Tsuha, S. Murata & K. Ohashi (2013).** "Horizontal transmission of bovine leukemia virus from lymphocytotic cattle, and beneficial effects of insect vector control." *Vet Rec* **173**(21), 527-528.

**Orange-bellied Parrot Recovery Team. (2016a).** "Autumn 2016 orange-bellied parrot recovery program summary." Retrieved 17/04/2017, from birdlife.org.au.

**Orange-bellied Parrot Recovery Team (2016b).** *Communique from the national orange-bellied parrot recovery team*. Orange-bellied parrot recovery team workshop, Hobart, Tasmania, Australia.

**Osawa, R., P. S. Bird, D. J. Harbrow, K. Ogimoto & G. J. Seymour (1993a).** "Microbiological studies of the intestinal microflora of the koala, *Phascolarctos cinereus*. 1. Colonization of the cecal wall by tannin-protein-complex-degrading *Enterobacteria*." *Aust J Zool* **41**(6), 599-609.

**Osawa, R., W. H. Blanshard & P. G. O'Callaghan (1992).** "Microflora of the pouch of the koala (*Phascolarctos cinereus*)." *J Wildl Dis* **28**(2), 276-280.

**Osawa, R., W. H. Blanshard & P. G. O'Callaghan (1993b).** "Microbiological studies of the intestinal microflora of the koala, *Phascolarctos cinereus* 2. Pap, a special maternal feces consumed by juvenile koalas." *Aust J Zool* **41**(6), 611-620.

**Osman, K. M., H. A. Ali, J. A. ElJakee & H. M. Galal (2011).** "*Chlamydophila psittaci* and *Chlamydophila pecorum* infections in goats and sheep in Egypt." *Rev Sci Tech* **30**(3), 939-948.

**Page, A. J., C. A. Cummins, M. Hunt, V. K. Wong, S. Reuter, M. T. Holden, M. Fookes, D. Falush, J. A. Keane & J. Parkhill (2015).** "Roary: rapid large-scale prokaryote pan genome analysis." *Bioinformatics* **31**(22), 3691-3693.

**Page, L. A. (1968).** "Proposal for the recognition of two species in the genus *Chlamydia* Jones, Rake, and Stearns, 1945." *Int J Syst Evol Microbiol* **18**(1), 51-66.

**Pannekoek, Y., G. Morelli, B. Kusecek, S. Morre, J. Ossewaarde, A. Langerak & A. van der Ende (2008).** "Multi locus sequence typing of *Chlamydiales*: clonal groupings within the obligate intracellular bacteria *Chlamydia trachomatis*." *BMC Microbiol* **8**(1), 42.

**Paradis, E. (2010).** "pegas: an R package for population genetics with an integrated–modular approach." *Bioinformatics* **26**(3), 419-420.

**Paradis, E., J. Claude & K. Strimmer (2004).** "APE: Analyses of Phylogenetics and Evolution in R language." *Bioinformatics* **20**(2), 289-290.

**Parris, H. S. (1948).** "Koalas on the lower Goulburn." *Vic Nat* **64**(10), 192-193.

**Patterson, J. L. S. (2012).** *Chlamydophila and koala retrovirus (KoRV) infection in southern populations of koalas (Phascolarctos cinereus).* . Master of Veterinary Science (Clinical), The University of Melbourne.

**Patterson, J. L. S., M. Lynch, G. A. Anderson, A. H. Noormohammadi, A. R. Legione, J. R. Gilkerson & J. M. Devlin (2015).** "The prevalence and clinical significance of *Chlamydia* infection in island and mainland populations of Victorian koalas (*Phascolarctos cinereus*)." *J Wildl Dis* **51**(2), 309-317.

**Pearce, M. M., E. E. Hilt, A. B. Rosenfeld, M. J. Zilliox, K. Thomas-White, C. Fok, S. Kliethermes, P. C. Schreckenberger, L. Brubaker, X. Gai & A. J. Wolfe (2014).** "The female urinary microbiome: a comparison of women with and without urgency urinary incontinence." *mBio* **5**(4).

**Peters, A., E. I. Patterson, B. G. B. Baker, M. Holdsworth, S. Sarker, S. A. Ghorashi & S. R. Raidal (2014).** "Evidence of psittacine beak and feather disease virus spillover into wild critically endangered orange-bellied parrots (*Neophema chrysogaster*)." *J Wildl Dis* **50**(2), 288-296.

**Petersen, T. N., S. Brunak, G. von Heijne & H. Nielsen (2011).** "SignalP 4.0: discriminating signal peptides from transmembrane regions." *Nat Methods* **8**(10), 785-786.

**Phillips, S. S. (2000).** "Population trends and the koala conservation debate." *Conserv Biol* **14**(3), 650-659.

**Plassan, J.-R. (1816).** Bulletin des sciences. Paris, Société Philomathique de Paris. **1816**.

**Polkinghorne, A., J. Hanger & P. Timms (2013).** "Recent advances in understanding the biology, epidemiology and control of chlamydial infections in koalas." *Vet Microbiol* **165**(3–4), 214-223.

**Price, M. N., P. S. Dehal & A. P. Arkin (2009).** "FastTree: computing large minimum evolution trees with profiles instead of a distance matrix." *Mol Biol Evol* **26**(7), 1641-1650.

**Pringle, C. R. (1998).** "Virus taxonomy–San Diego 1998." *Arch Virol* **143**(7), 1449-1459.

**Pritchard, R. (2012).** "Update on the orange-bellied parrot recovery program."   Retrieved 17/04/2017.

**Quinlan, A. R. & I. M. Hall (2010).** "BEDTools: a flexible suite of utilities for comparing genomic features." *Bioinformatics* **26**(6), 841-842.

**R Core Team (2016).** R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing.

**Rasmussen, S. J., L. Eckmann, A. J. Quayle, L. Shen, Y. X. Zhang, D. J. Anderson, J. Fierer, R. S. Stephens & M. F. Kagnoff (1997).** "Secretion of proinflammatory cytokines by

epithelial cells in response to *Chlamydia* infection suggests a central role for epithelial cells in chlamydial pathogenesis." *J Clin Invest* **99**(1), 77-87.

**Ravel, J., P. Gajer, Z. Abdo, G. M. Schneider, S. S. K. Koenig, S. L. McCulle, S. Karlebach, R. Gorle, J. Russell, C. O. Tacket, R. M. Brotman, C. C. Davis, K. Ault, L. Peralta & L. J. Forney (2011).** "Vaginal microbiome of reproductive-age women." *Proc Natl Acad Sci USA* **108**(Supplement 1), 4680-4687.

**Regenscheit, N., N. Holzwarth, G. Greub, S. Aeby, A. Pospischil & N. Borel (2012).** "Deer as a potential wildlife reservoir for *Parachlamydia* species." *Vet J* **193**(2), 589-592.

**Reid, G. & J. D. Sobel (1987).** "Bacterial adherence in the pathogenesis of urinary tract infection: a review." *Rev Infect Dis* **9**(3), 470-487.

**Rice, P., I. Longden & A. Bleasby (2000).** "EMBOSS: The European Molecular Biology Open Software Suite." *Trends Genet* **16**(6), 276-277.

**Ridderhof, J. C. & R. C. Barnes (1989).** "Fusion of inclusions following superinfection of HeLa cells by two serovars of *Chlamydia trachomatis*." *Infect Immun* **57**(10), 3189-3193.

**Robertson, T., S. Bibby, D. O'Rourke, T. Belfiore, H. Lambie & A. H. Noormohammadi (2009).** "Characterization of *Chlamydiaceae* species using PCR and high resolution melt curve analysis of the 16S rRNA gene." *J Appl Microbiol* **107**(6), 2017-2028.

**Robinson, A. C. & T. J. Bergin (1978).** The koala in South Australia. The Koala: Proceedings of the Taronga Symposium on Koala Biology, Management and Medicine, Taronga Zoo, New South Wales, Australia, Zoological Parks Board of New South Wales.

**Robinson, D. F. & L. R. Foulds (1981).** "Comparison of phylogenetic trees." *Math Biosci* **53**(1), 131-147.

**Robinson, W. F., T. M. O'Brien, J. Hanger & J. J. McKee (1996).** *Leukaemia in koalas (Phascolarctos cinereus): the presence of viral particles in cultured lymphocytes*. A Conference on the Status of the Koala, Greenmount, Coolangatta, Queensland, Australia, Australian Koala Foundation.

**Roelke, M. E., J. S. Martenson & S. J. O'Brien (1993).** "The consequences of demographic reduction and genetic depletion in the endangered Florida panther." *Curr Biol* **3**(6), 340-350.

**Roulis, E., N. Bachmann, M. Humphrys, G. Myers, W. Huston, A. Polkinghorne & P. Timms (2015).** "Phylogenetic analysis of human *Chlamydia pneumoniae* strains reveals a distinct Australian indigenous clade that predates European exploration of the continent." *BMC Genomics* **16**, 1094.

**Ruiz-Rodriguez, C. T., Y. Ishida, A. D. Greenwood & A. L. Roca (2014).** "Development of 14 microsatellite markers in the Queensland koala (*Phascolarctos cinereus adustus*) using next generation sequencing technology." *Conservation Genet Resour* **6**(2), 429-431.

**Ruiz-Rodriguez, C. T., Y. Ishida, N. D. Murray, S. J. O'Brien, J. A. M. Graves, A. D. Greenwood & A. L. Roca (2016).** "Koalas (*Phascolarctos cinereus*) from Queensland are genetically distinct from two populations in Victoria." *J Hered* **107**(7), 573-580.

**Sabet, S. F., J. Simmons & H. D. Caldwell (1984).** "Enhancement of *Chlamydia trachomatis* infectious progeny by cultivation of HeLa 229 cells treated with DEAE-dextran and cycloheximide." *J Clin Microbiol* **20**(2), 217-222.

**Sachse, K., K. Laroucau, K. Riege, S. Wehner, M. Dilcher, H. H. Creasy, M. Weidmann, G. Myers, F. Vorimore, N. Vicari, S. Magnino, E. Liebler-Tenorio, A. Ruettger, P. M. Bavoil, F. T. Hufert, R. Rosselló-Móra & M. Marz (2014).** "Evidence for the existence of two new members of the family *Chlamydiaceae* and proposal of *Chlamydia avium* sp. nov. and *Chlamydia gallinacea* sp. nov." *Syst Appl Microbiol* **37**(2), 79-88.

**Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis & H. A. Erlich (1988).** "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase." *Science* **239**(4839), 487-491.

**Sait, M., M. Livingstone, E. Clark, N. Wheelhouse, L. Spalding, B. Markey, S. Magnino, F. Lainson, G. Myers & D. Longbottom (2014).** "Genome sequencing and comparative analysis of three *Chlamydia pecorum* strains associated with different pathogenic outcomes." *BMC Genomics* **15**, 23.

**Santamaria, F. & R. Schlagloth (2016).** "The effect of *Chlamydia* on translocated *Chlamydia*-naïve koalas: a case study." *Aust Zool* **38**(2), 192-202.

**Schoch, C. L., K. A. Seifert, S. Huhndorf, V. Robert, J. L. Spouge, C. A. Levesque, W. Chen & Fungal Barcoding Consortium (2012).** "Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*." *Proc Natl Acad Sci USA* **109**(16), 6241-6246.

**Schwartzberg, P., J. Colicelli & S. P. Goff (1984).** "Construction and analysis of deletion mutations in the *pol* gene of Moloney murine leukemia virus: a new viral function required for productive infection." *Cell* **37**(3), 1043-1052.

**Seddon, P. J. (1999).** "Persistence without intervention: assessing success in wildlife reintroductions." *Trends Ecol Evol* **14**(12), 503.

**Seemann, T. (2014).** "Prokka: rapid prokaryotic genome annotation." *Bioinformatics* **30**(14), 2068-2069.

**Seth-Smith, H. M. B., S. R. Harris, K. Persson, P. Marsh, A. Barron, A. Bignell, C. Bjartling, L. Clark, L. T. Cutcliffe, P. R. Lambden, N. Lennard, S. J. Lockey, M. A. Quail, O. Salim, R. J. Skilton, Y. Wang, M. J. Holland, J. Parkhill, N. R. Thomson & I. N. Clarke (2009).** "Co-evolution of genomes and plasmids within *Chlamydia trachomatis* and the emergence in Sweden of a new variant strain." *BMC Genomics* **10**, 239.

**Seth-Smith, H. M. B., S. R. Harris, R. J. Skilton, F. M. Radebe, D. Golparian, E. Shipitsyna, P. T. Duy, P. Scott, L. T. Cutcliffe, C. O'Neill, S. Parmar, R. Pitt, S. Baker, C. A. Ison, P. Marsh, H. Jalal, D. A. Lewis, M. Unemo, I. N. Clarke, J. Parkhill & N. R. Thomson (2013).** "Whole-genome sequences of *Chlamydia trachomatis* directly from clinical samples without culture." *Genome Res* **23**(5), 855-866.

**Shannon, C. E. (1948).** "A mathematical theory of communication." *Bell Syst Tech J* **27**(3), 379-423.

**Shimode, S., S. Nakagawa, R. Yoshikawa, T. Shojima & T. Miyazawa (2014).** "Heterogeneity of koala retrovirus isolates." *FEBS Lett* **588**(1), 41-46.

**Shojima, T., R. Yoshikawa, S. Hoshino, S. Shimode, S. Nakagawa, T. Ohata, R. Nakaoka & T. Miyazawa (2013).** "Identification of a novel subgroup of koala retrovirus from koalas in Japanese zoos." *J Virol* **87**(17), 9943-9948.

**Short, M. K., S. A. Okenquist & J. Lenz (1987).** "Correlation of leukemogenic potential of murine retroviruses with transcriptional tissue preference of the viral long terminal repeats." *J Virol* **61**(4), 1067-1072.

**Simmons, G., D. Clarke, J. McKee, P. Young & J. Meers (2014).** "Discovery of a novel retrovirus sequence in an Australian native rodent (*Melomys burtoni*): a putative link between gibbon ape leukemia virus and koala retrovirus." *PLoS ONE* **9**(9), e106954.

**Simmons, G. S., P. R. Young, J. J. Hanger, K. Jones, D. Clarke, J. J. McKee & J. Meers (2012).** "Prevalence of koala retrovirus in geographically diverse populations in Australia." *Aust Vet J* **90**(10), 404-409.

**Smith, P. & J. Smith (1990).** "Decline of the urban koala (*Phascolarctos cinereus*) population in Warringah Shire, Sydney." *Aust Zool* **26**(3), 109-129.

**Speight, K. N., A. Polkinghorne, R. Penn, W. Boardman, P. Timms, T. Fraser, K. Johnson, R. Faull, S. Bate & L. Woolford (2016).** "Prevalence and pathologic features of *Chlamydia pecorum* infections in South Australian koalas (*Phascolarctos cinereus*)." *J Wildl Dis* **52**(2), 301-306.

**Speight, K. N., P. L. Whiteley, L. Woolford, P. J. Duignan, B. Bacci, S. Lathe, W. Boardman, T. F. Scheelings, O. Funnell, G. Underwood & M. A. Stevenson (2017).** "Outbreaks of

sarcoptic mange in free-ranging koala populations in Victoria and South Australia: a case series." *Aust Vet J* **95**(7), 244-249.

**Spencer, A., J. & P. J. Canfield (1996).** "Lymphoid neoplasia in the koala (*Phascolarctos cinereus*): A review and classification of 31 cases." *J Zoo Wildl Med* **27**(3), 303-314.

**Spielman, D., B. W. Brook, D. A. Briscoe & R. Frankham (2004).** "Does inbreeding and loss of genetic diversity decrease disease resistance?" *Conserv Genet* **5**(4), 439-448.

**Stalder, K., P. K. Vaz, J. R. Gilkerson, R. Baker, P. Whiteley, N. Ficorilli, L. Tatarczuch, T. Portas, K. Skogvold, G. A. Anderson & J. M. Devlin (2015).** "Prevalence and clinical significance of herpesvirus infection in populations of Australian marsupials." *PLoS ONE* **10**(7), e0133807.

**Stamatakis, A. (2014).** "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies." *Bioinformatics* **30**(9), 1312-1313.

**Starks, J., P. Brown, R. Loyn & P. Menkhorst (1992).** "Twelve years of winter counts of the orange-bellied parrot *Neophema chrysogaster*." *Aust Bird Watch* **14**(8), 305-312.

**Stephens, R. S., G. Myers, M. Eppinger & P. M. Bavoil (2009).** "Divergence without difference: phylogenetics and taxonomy of *Chlamydia* resolved." *FEMS Immunol Med Microbiol* **55**(2), 115-119.

**Strahan, R. (1978).** What is a koala? The Koala: Proceedings of the Taronga Symposium of Koala Biology, Management and Medicine, Taronga Zoo, New South Wales, Australia, Zoological Parks Board of NSW.

**Strahan, R. & R. Martin (1981).** *The koala: little fact, much emotion*. Symposium on the Biology of Rare and Endangered Species in Australia, Canberra, Australia, Australian Academy of Science.

**Strandström, H., P. Veijalainen, V. Moennig, G. Hunsmann, H. Schwarz & W. Schäfer (1974).** "C-type particles produced by a permanent cell line from a leukemic pig." *Virology* **57**(1), 175-178.

**Streitenberger, S. A., J. A. Lopez-Mas, A. Sanchez-Ferrer & F. Garcia-Carmona (2001).** "Highly efficient *Aerococcus viridans* L-alpha-glycerophosphate oxidase production in the presence of $H_2O_2$-decomposing agent: purification and kinetic characterization." *Appl Microbiol Biotechnol* **57**(3), 329-333.

**Subtil, A., C. Parsot & A. Dautry-Varsat (2001).** "Secretion of predicted Inc proteins of *Chlamydia pneumoniae* by a heterologous type III machinery." *Mol Microbiol* **39**(3), 792-800.

**Tabart, D. (2002).** "Koala myth becomes scientific fact." Retrieved 2015.

**Tajima, F. (1989).** "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism." *Genetics* **123**(3), 585-595.

**Takeuchi, Y., R. G. Vile, G. Simpson, B. O'Hara, M. K. Collins & R. A. Weiss (1992).** "Feline leukemia virus subgroup B uses the same cell surface receptor as gibbon ape leukemia virus." *J Virol* **66**(2), 1219-1222.

**Tamura, A. & N. Higashi (1963).** "Purification and chemical composition of meningopneumonitis virus." *Virology* **20**(4), 596-604.

**Tamura, A. & G. P. Manire (1967).** "Preparation and chemical composition of the cell membranes of developmental reticulate forms of meningopneumonitis organisms." *J Bacteriol* **94**(4), 1184-1188.

**Tamura, A., A. Matsumoto, G. P. Manire & N. Higashi (1971).** "Electron microscopic observations on the structure of the envelopes of mature elementary bodies and developmental reticulate forms of *Chlamydia psittaci*." *J Bacteriol* **105**(1), 355-360.

**Tarlinton, R., J. Meers, J. Hanger & P. Young (2005).** "Real-time reverse transcriptase PCR for the endogenous koala retrovirus reveals an association between plasma viral load and neoplastic disease in koalas." *J Gen Virol* **86**(3), 783-787.

**Tarlinton, R. E., J. Meers & P. R. Young (2006).** "Retroviral invasion of the koala genome." *Nature* **442**(7098), 79-81.

**Temin, H. M. & S. Mizutani (1970).** "Viral RNA-dependent DNA polymerase: RNA-dependent DNA polymerase in virions of Rous sarcoma virus." *Nature* **226**(5252), 1211-1213.

**The UniProt Consortium (2017).** "UniProt: the universal protein knowledgebase." *Nucleic Acids Res* **45**(D1), D158-D169.

**Thompson, J. D., D. G. Higgins & T. J. Gibson (1994).** "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." *Nucleic Acids Res* **22**(22), 4673-4680.

**Thomson, N. R., M. T. G. Holden, C. Carder, N. Lennard, S. J. Lockey, P. Marsh, P. Skipp, C. D. O'Connor, I. Goodhead, H. Norbertzcak, B. Harris, D. Ormond, R. Rance, M. A. Quail, J. Parkhill, R. S. Stephens & I. N. Clarke (2008).** "*Chlamydia trachomatis*: genome sequence analysis of lymphogranuloma venereum isolates." *Genome Res* **18**(1), 161-171.

**Thomson, N. R., C. Yeats, K. Bell, M. T. G. Holden, S. D. Bentley, M. Livingstone, A. M. Cerdeño-Tárraga, B. Harris, J. Doggett, D. Ormond, K. Mungall, K. Clarke, T. Feltwell, Z. Hance, M. Sanders, M. A. Quail, C. Price, B. G. Barrell, J. Parkhill & D. Longbottom (2005).** "The *Chlamydophila abortus* genome sequence reveals an array of variable proteins that contribute to interspecies variation." *Genome Res* **15**(5), 629-640.

**Timms, P. (2014).** "KoRV and *Chlamydia*: are they co-culprits?" *Tech Rep Aust Mus (online)* **24**, 89-90.

**Tovar, C., R. J. Pye, A. Kreiss, Y. Cheng, G. K. Brown, J. Darby, R. C. Malley, H. V. T. Siddle, K. Skjødt, J. Kaufman, A. Silva, A. Baz Morelli, A. T. Papenfuss, L. M. Corcoran, J. M. Murphy, M. J. Pearse, K. Belov, A. B. Lyons & G. M. Woods (2017).** "Regression of devil facial tumour disease following immunotherapy in immunised Tasmanian devils." *Sci Rep* **7**, 43827 DOI: 10.1038/srep43827.

**Trask, A. E., E. M. Bignal, D. I. McCracken, P. Monaghan, S. B. Piertney & J. M. Reid (2016).** "Evidence of the phenotypic expression of a lethal recessive allele under inbreeding in a wild population of conservation concern." *J Anim Ecol* **85**(4), 879-891.

**Troughton, E. (1941).** *Furred animals of Australia*. Sydney, NSW, Australia, Angus and Roberson Ltd.

**Tsangaras, K., M. C. Siracusa, N. Nikolaidis, Y. Ishida, P. Cui, H. Vielgrader, K. M. Helgen, A. L. Roca & A. D. Greenwood (2014).** "Hybridization capture reveals evolution and conservation across the entire koala retrovirus genome." *PLoS ONE* **9**(4), e95633.

**Turnbaugh, P. J., V. K. Ridaura, J. J. Faith, F. E. Rey, R. Knight & J. I. Gordon (2009).** "The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice." *Sci Transl Med* **1**(6), 6ra14.

**van der Woude, M. W. & A. J. Bäumler (2004).** "Phase and antigenic variation in bacteria." *Clin Microbiol Rev* **17**(3), 581-611.

**Vaz, P., P. L. Whiteley, C. R. Wilks, G. F. Browning, J. R. Gilkerson, N. Ficorilli & J. M. Devlin (2012).** "Detection of a second novel gammaherpesvirus in a free-ranging koala (*Phascolarctos cinereus*)." *J Wildl Dis* **48**(1), 226-229.

**Vaz, P., P. L. Whiteley, C. R. Wilks, P. J. Duignan, N. Ficorilli, J. R. Gilkerson, G. F. Browning & J. M. Devlin (2011).** "Detection of a novel gammaherpesvirus in koalas (*Phascolarctos cinereus*)." *J Wildl Dis* **47**(3), 787-791.

**Vaz, P. K., C. A. Hartley, G. F. Browning & J. M. Devlin (2015).** "Marsupial and monotreme serum immunoglobulin binding by proteins A, G and L and anti-kangaroo antibody." *J Immunol Methods* **427**, 94-99.

**Vázquez-Baeza, Y., M. Pirrung, A. Gonzalez & R. Knight (2013).** "EMPeror: a tool for visualizing high-throughput microbial community data." *GigaScience* **2**, 16.

**Verkooyen, R. P., D. Willemse, S. C. A. M. Hiep-van Casteren, S. A. Mousavi Joulandan, R. J. Snijder, J. M. M. van den Bosch, H. P. T. van Helden, M. F. Peeters & H. A. Verbrugh (1998).** "Evaluation of PCR, culture, and serology for diagnosis of *Chlamydia pneumoniae* respiratory infections." *J Clin Microbiol* **36**(8), 2301-2307.

**Viratyosin, W., L. A. Campbell, C.-C. Kuo & D. D. Rockey (2002).** "Intrastrain and interstrain genetic variation within a paralogous gene family in *Chlamydia pneumoniae*." *BMC Microbiol* **2**, 38.

**Wan, C., J. Loader, J. Hanger, K. W. Beagley, P. Timms & A. Polkinghorne (2011).** "Using quantitative polymerase chain reaction to correlate *Chlamydia pecorum* infectious load with ocular, urinary and reproductive tract disease in the koala (*Phascolarctos cinereus*)." *Aust Vet J* **89**(10), 409-412.

**Warneke, R. M. (1978).** *The status of the koala in Victoria*. The Koala: Proceedings of the Taronga Symposium of Koala Biology, Management and Medicine, Taronga Zoo, New South Wales, Australia, Zoological Parks Board of NSW.

**Waugh, C., A. Gillett, A. Polkinghorne & P. Timms (2016a).** "Serum antibody response to koala retrovirus antigens varies in free-ranging koalas (*Phascolarctos cinereus*) in Australia: implications for vaccine design." *J Wildl Dis* **52**(2), null.

**Waugh, C., S. A. Khan, S. Carver, J. Hanger, J. Loader, A. Polkinghorne, K. Beagley & P. Timms (2016b).** "A prototype recombinant-protein based *Chlamydia pecorum* vaccine results in reduced chlamydial burden and less clinical disease in free-ranging koalas (*Phascolarctos cinereus*)." *PLoS ONE* **11**(1), e0146934.

**Waugh, C. A., J. Hanger, J. Loader, A. King, M. Hobbs, R. Johnson & P. Timms (2017).** "Infection with koala retrovirus subgroup B (KoRV-B), but not KoRV-A, is associated with chlamydial disease in free-ranging koalas (Phascolarctos cinereus)." *Sci Rep* **7**(1), 134.

**Wedrowicz, F., T. Saxton, J. Mosse, W. Wright & F. E. Hogan (2016).** "A non-invasive tool for assessing pathogen prevalence in koala (*Phascolarctos cinereus*) populations: detection of *Chlamydia pecorum* and koala retrovirus (KoRV) DNA in genetic material sourced from scats." *Conservation Genet Resour* **8**(4), 511-521.

**Weiss, E. (1965).** "Adenosine triphosphate and other requirements for the utilization of glucose by agents of the psittacosis-trachoma group." *J Bacteriol* **90**(1), 243-253.

**Weiss, R. A. (1993).** Cellular receptors and viral glycoproteins involved in retrovirus entry. *The Retroviridae*. Levy, J. A. Boston, MA, Springer US**:** 1-108.

**Whisson, D. A., V. Dixon, M. L. Taylor & A. Melzer (2016).** "Failure to respond to food resource decline has catastrophic consequences for koalas in a high-density population in southern Australia." *PLoS ONE* **11**(1), e0144348.

**White, N. A. (1999).** "Ecology of the koala (*Phascolarctos cinereus*) in rural south-east Queensland, Australia." *Wildl Res* **26**(6), 731-744.

**White, N. A. & P. Timms (1994).** "*Chlamydia psittaci* in a koala *(Phascolarctos cinereus)* population in south-east Queensland." *Wildl Res* **21**(1), 41-47.

**Wiesenfeld, H. C., S. L. Hillier, M. A. Krohn, D. V. Landers & R. L. Sweet (2003).** "Bacterial vaginosis is a strong predictor of *Neisseria gonorrhoeae* and *Chlamydia trachomatis* infection." *Clin Infect Dis* **36**(5), 663-668.

**Wilkinson, R., M. Allanson, V. Kolega, D. Lawrence & S. Neville (1991).** "Purification and initial characterisation of koala immunoglobulins." *Vet Immunol Immunopathol* **29**(1), 189-195.

**Wilson, C. A., S. Laeeq, A. Ritzhaupt, W. Colon-Moran & F. K. Yoshimura (2003).** "Sequence analysis of porcine endogenous retrovirus long terminal repeats and identification of transcriptional regulatory regions." *J Virol* **77**(1), 142-149.

**Worley, M., B. Rideout, A. Shima & D. Janssen (1993).** Opportunistic infections, cancer and hematologic disorders associated with retrovirus infection in the koala. Proceedings of American Association of Zoo Veterinarians, St. Louis, Media, PA: AAZV.

**Xu, W., K. Gorman, J. Santiago, K. Kluska & M. Eiden (2015).** "Genetic diversity of koala retroviral envelopes." *Viruses* **7**(3), 1258.

**Xu, W., C. K. Stadler, K. Gorman, N. Jensen, D. Kim, H. Zheng, S. Tang, W. M. Switzer, G. W. Pye & M. V. Eiden (2013).** "An exogenous retrovirus isolated from koalas with malignant neoplasias in a US zoo." *Proc Natl Acad Sci USA* **110**(28), 11547-11552.

**Xu, Y., C. Moser, W. A. Al-Soud, S. Sørensen, N. Høiby, P. H. Nielsen & T. R. Thomsen (2012).** "Culture-dependent and-independent investigations of microbial diversity on urinary catheters." *J Clin Microbiol* **50**(12), 3901-3908.

**Yang, R., C. Jacobson, G. Gardner, I. Carmichael, A. J. D. Campbell & U. Ryan (2014).** "Longitudinal prevalence and faecal shedding of *Chlamydia pecorum* in sheep." *Vet J* **201**(3), 322-326.

**Yang, Z. (2007).** "PAML 4: Phylogenetic Analysis by Maximum Likelihood." *Mol Biol Evol* **24**(8), 1586-1591.

**Yang, Z. & R. Nielsen (1998).** "Synonymous and nonsynonymous rate variation in nuclear genes of mammals." *J Mol Evol* **46**(4), 409-418.

**Yang, Z., R. Nielsen, N. Goldman & A.-M. K. Pedersen (2000).** "Codon-substitution models for heterogeneous selection pressure at amino acid sites." *Genetics* **155**(1), 431-449.

**Yang, Z., W. S. W. Wong & R. Nielsen (2005).** "Bayes empirical bayes inference of amino acid sites under positive selection." *Mol Biol Evol* **22**(4), 1107-1118.

**Yeruva, L., N. Spencer, A. K. Bowlin, Y. Wang & R. G. Rank (2013).** "Chlamydial infection of the gastrointestinal tract: a reservoir for persistent infection." *Pathogens and Disease* **68**(3), 88-95.

**Young, P. R. (2014).** "Koala retrovirus (KoRV) and its variants." *Tech Rep Aust Mus (online)* **24**, 59-60.

**Yousef Mohamad, K., B. Kaltenboeck, K. S. Rahman, S. Magnino, K. Sachse & A. Rodolakis (2014).** "Host adaptation of *Chlamydia pecorum* towards low virulence evident in co-evolution of the *ompA*, *incA*, and ORF663 loci." *PLoS ONE* **9**(8), e103615.

**Yousef Mohamad, K., S. M. Roche, G. Myers, P. M. Bavoil, K. Laroucau, S. Magnino, S. Laurent, D. Rasschaert & A. Rodolakis (2008).** "Preliminary phylogenetic identification of virulent *Chlamydophila pecorum* strains." *Infect, Genet Evol* **8**(6), 764-771.

**Yu, Y., C. Lee, J. Kim & S. Hwang (2005).** "Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction." *Biotechnol Bioeng* **89**(6), 670-679.

**Zhang, Q., C. Kwoh, S. Attorri & J. E. Clarridge (2000).** "*Aerococcus urinae* in urinary tract infections." *J Clin Microbiol* **38**(4), 1703-1705.

**Zhou, X., C. J. Brown, Z. Abdo, C. C. Davis, M. A. Hansmann, P. Joyce, J. A. Foster & L. J. Forney (2007).** "Differences in the composition of vaginal microbial communities found in healthy caucasian and black women." *ISME J* **1**(2), 121-133.

# Appendices

**Appendix 1.** Members of the genus *Chlamydia* and their typical reservoir host, or commonly infected host

| ***Chlamydia*** **species** | **Typical/common/reservoir hosts** | **Year classified** | **Reference** |
|---|---|---|---|
| *Chlamydia trachomatis* | Humans (*Homo sapiens*) | 1968 | Page (1968) |
| *Chlamydia psittaci* | Psittacine and columbine birds | 1968 | Page (1968) |
| *Chlamydia pneumoniae* | Humans (*Homo sapiens*) | 1989 | Grayston *et al.* (1989) |
| *Chlamydia pecorum* | Koalas (*Phascolarctos cinereus*)<br>Cattle (*Bos Taurus*)<br>Sheep (*Ovis aries*) | 1992 | Fukushi and Hirai (1992) |
| *Chlamydia felis* | Felines (*Felis catus*) | 1999 | Everett *et al.* (1999) |
| *Chlamydia caviae* | Guinea pigs (*Cavia porcellus*) | 1999 | Everett *et al.* (1999) |
| *Chlamydia muridarum* | Mice (*Mus musculus*) | 1999 | Everett *et al.* (1999) |
| *Chlamydia suis* | Pigs (*Sus scrofa*) | 1999 | Everett *et al.* (1999) |
| *Chlamydia abortus* | Cattle (*Bos Taurus*)<br>Sheep (*Ovis aries*) | 1999 | Everett *et al.* (1999) |
| *Chlamydia avium* | Psittacine and columbine birds | 2014 | Sachse *et al.* (2014) |
| *Chlamydia gallinacea* | Chickens (*Gallus gallus*) | 2014 | Sachse *et al.* (2014) |

**Appendix 2.** Wet bottom scoring criteria replicated verbatim from Griffith (2010), which in turn was based on a criteria developed by Flanagan (2009)

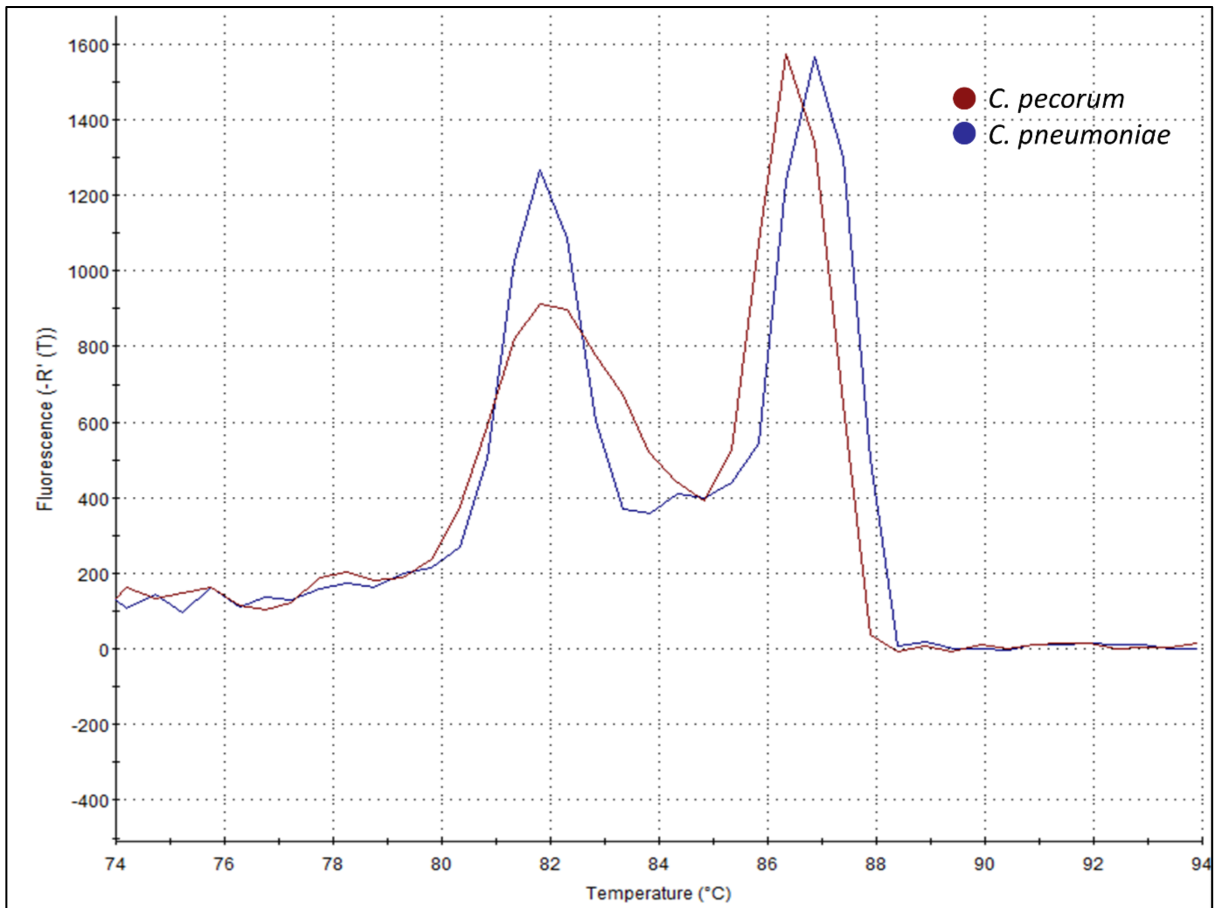
| Grade | Description |
|---|---|
| 0 | Normal fur, normal cloaca |
| 0.5 | Discolouration of fur around cloaca |
| 1 | Slight discolouration of fur around cloaca |
| | Evidence of mild fresh urine leakage |
| | Slight “wet bottom” smell |
| 2 | Slight discolouration of fur around cloaca/tail area |
| | Occasional urine dribbles |
| | Mild yet discernible “wet bottom” smell |
| 3 | Discolouration of tail area fur more evident |
| | Stronger “wet bottom” smell |
| | Urine discharge, greasy texture evident around cloaca/tail area |
| 4 | Fur stained, greasy, darkened |
| | Strong pungent “wet bottom” smell |
| | Inflammation of the cloacal margins, clitoris, vestibule |
| | Discharge containing urinary calculus/debris |
| 5 | Stained greasy fur covering a large area |
| | Very strong pungent acidic smell |
| | Blood in urine, crying and straining when urinating |
| | Clots, blood in both male and female urine and urogenital tracks |
| | Coat brown dry and lustreless |
| | Cloaca and tail area swollen (oedema) |
| | Grinding teeth |
| 6 | Stained, greasy, wet matted fur around rump/cloaca area |
| | Blood in urine, clots, and constant purulent discharge |
| | Crying, straining, grinding teeth, distressed, flat, ear flicking |
| | Ulcerated, oedematous cloaca/tail area |
| 7 - 10 | Progressive decline |
| | Often becomes maggot infested |
| | Death ultimately results if intervention/removal of suffering does not occur |

**Appendix 3.** Univariable analysis assessing select epidemiological variables relating to animal health and disease as predictors for the presence of *Chlamydia pecorum* DNA in both sexes of koala.

| Variable | *C. pecorum* positive | Prevalence | Odds ratio* | 95% CI | Co-efficient *P* value | Likelihood ratio *P* value^ |
|---|---|---|---|---|---|---|
| **Body condition score** | | | | | | 0.105 |
| ≤ 2 | 10/113 | 8.8% | 0.58 | 0.29 – 1.16 | 0.126 | |
| ≥ 3 | 84/589 | 14.3% | 1.00 | - | - | |
| Not recorded | 31/118 | | | | | |
| **Wet bottom** | | | | | | **0.008** |
| Absent | 83/608 | 13.7% | 1.00 | - | - | |
| Present | 41/187 | 21.9% | 1.78 | 1.17 – 2.69 | 0.007 | |
| Not recorded | 1/25 | | | | | |
| **Urinary tract pathology** | | | | | | 0.736 |
| Absent | 40/169 | 23.7% | 1.00 | - | - | |
| Present | 12/46 | 26.1% | 1.14 | 0.54 – 2.40 | 0.734 | |
| Not recorded | 73/605 | | | | | |

[*] Reference levels are indicated by odds ratio of 1.0.

[^] Results highlighted in bold (log likelihood $P \leq 0.25$) represent variables included in the initial multivariable model. In the final model (n = 702), after stepwise backwards elimination, only wet bottom was identified as a significant factor ($P = 0.002$).




**Appendix 4.** Melt curve of the 16S rRNA region SYBR green qPCR (16SG, Robertson *et al.* (2009)) Melt curve generated at a resolution of 0.3°C. *C. pneumoniae* and *C. pecorum* are differentiated by the shape of the first peak and melting temperature of the second peak.

**Appendix 5.** Multivariable mixed effects logistic regression model assessing clinical variables for associations with the presence of urogenital tract disease in 95 female koalas from Victoria, Australia, sampled between 2010 to 2016.

**a**) Including koala retrovirus detection as a main effect:

| Variable | | OR * | SE(OR) * | 95% CI | *P* value ^ |
|---|---|---|---|---|---|
| **Age** | | | | | |
| | Young | 0.13 | 0.13 | 0.02, 0.93 | 0.042 |
| | Mature adult | 1.00 | — | — | — |
| | Old adult | 1.90 | 1.76 | 0.31, 11.6 | 0.49 |
| **Wet bottom** | | | | | |
| | Present | 10.6 | 7.14 | 2.85, 39.7 | < 0.001 |
| | Absent | 1.00 | — | — | — |
| ***C. pecorum*** | | | | | |
| | Present | 5.71 | 4.96 | 1.04, 31.3 | 0.045 |
| | Absent | 1.00 | — | — | — |
| **KoRV** | | | | | |
| | Present | 1.53 | 0.91 | 0.48, 4.91 | 0.48 |
| | Absent | 1.00 | — | — | — |
| **Constant** | | 0.05 | 0.05 | 0.01, 0.33 | 0.002 |
| | | **Var *** | **SE(Var)** | **95% CI** | **P value ^** |
| **Regional cluster** | | 1.42 | 1.41 | 0.20, 9.90 | <0.001 |

N = 95; Log likelihood = -43.04; $P = 0.005$; df = 7; Akaike information criterion = 100.08

**b**) Including koala retrovirus detection in an interaction term with *C. pecorum*:

| Variable | | OR * | SE(OR) * | 95% CI | *P* value ^ |
|---|---|---|---|---|---|
| **Age** | | | | | |
| | Young | 0.12 | 0.12 | 0.02, 0.88 | 0.037 |
| | Mature adult | 1.00 | — | — | — |
| | Old adult | 2.24 | 2.12 | 0.35, 14.3 | 0.40 |
| **Wet bottom** | | | | | |
| | Present | 11.3 | 7.73 | 2.97, 43.1 | < 0.001 |
| | Absent | 1.00 | — | — | — |
| **Interaction of *C. pecorum* and KoRV** | | | | | |
| | Both detected | 20.5 | 32.0 | 0.97, 433.7 | 0.052 |
| | Only *C. pecorum* | 3.59 | 3.69 | 0.48, 26.85 | 0.21 |
| | Only KoRV | 1.25 | 0.82 | 0.35, 4.50 | 0.73 |
| | Neither detected | 1.00 | — | — | |
| **Constant** | | 0.05 | 0.05 | 0.01, 0.34 | 0.002 |
| | | **Var *** | **SE(Var)** | **95% CI** | **P value ^** |
| **Regional cluster** | | 1.56 | 1.54 | 0.23, 10.8 | <0.001 |

N = 95; Log likelihood = -42.69; $P = 0.010$; df = 8; Akaike information criterion = 101.38

* OR = Odds ratio; SE = Standard error; Var = estimated variance of random effect. Reference levels are indicated by odds ratios of 1.0.

^ *P* values estimated using the likelihood ratio test statistic. Results highlighted in bold ($P < 0.05$).

**Appendix 6.** Absolute abundance of merged reads clustered to assigned operational taxonomic units (OTUs). OTUs were assigned a taxonomy (with 90% identity) from the Greengenes curated database (DeSantis *et al.* 2006). Koalas K1 – K5 were clinically normal (wet bottom absent), whilst koalas K31 – K70 had wet bottom.

| OTU | K1 | K2 | K3 | K4 | K5 | K31 | K49 | K55 | K59 | K70 | Phylum | Class | Order | Family | Genus | Species |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| OTU 1 | 23420 | 18380 | 13250 | 44770 | 1821 | 76350 | 65180 | 131700 | 124500 | 4724 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | -^ |
| OTU 2 | 90660 | 65200 | 60810 | 43370 | 130100 | 3915 | 713 | 955 | 28070 | 19680 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | - |
| OTU 3 | 23 | 188 | 1382 | 6 | 249 | 79 | 100900 | 3 | 9 | 55 | *Proteobacteria* | *Gammaproteobacteria* | *Enterobacteriales* | *Enterobacteriaceae* | - | - |
| OTU 4 | 22860 | 23020 | 30080 | 13500 | 27720 | 122 | 69 | 10 | 1109 | 5671 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | - |
| OTU 5 | 756 | 67 | 3121 | 2115 | 230 | 0 | 32 | 4 | 884 | 62000 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 6 | 1525 | 1115 | 11920 | 34530 | 538 | 23360 | 2491 | 2467 | 1104 | 1078 | Unassigned | - | - | - | - | - |
| OTU 7 | 10010 | 13720 | 9683 | 9116 | 73 | 6158 | 323 | 4228 | 16 | 338 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 8 | 176 | 0 | 284 | 162 | 60 | 3 | 3 | 1 | 97 | 28550 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 9 | 11110 | 11350 | 2591 | 0 | 1 | 1506 | 109 | 2225 | 6 | 410 | *Actinobacteria* | *Coriobacteriia* | *Coriobacteriales* | *Coriobacteriaceae* | - | - |
| OTU 10 | 16210 | 10310 | 8742 | 29050 | 519 | 26610 | 3010 | 15940 | 1950 | 1673 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Facklamia* | - |
| OTU 11 | 316 | 24 | 0 | 1040 | 37 | 1 | 11 | 0 | 639 | 26710 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 12 | 48 | 26 | 959 | 1220 | 46 | 0 | 1 | 1 | 61 | 18010 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 13 | 56 | 17 | 266 | 224 | 13 | 0 | 0 | 4 | 462 | 9511 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 14 | 472 | 4330 | 1134 | 0 | 0 | 3320 | 1063 | 129 | 521 | 54 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 15 | 42 | 886 | 358 | 0 | 5 | 2243 | 89 | 6177 | 2 | 41 | *TM7* | *TM7-3* | *I025* | *Rs-045* | - | - |
| OTU 16 | 3 | 1 | 0 | 1 | 0 | 525 | 41 | 19030 | 185 | 44 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Porphyromonas* | - |
| OTU 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 134 | 8132 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 18 | 65 | 3 | 78 | 220 | 0 | 0 | 1 | 0 | 36 | 13100 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Veillonellaceae* | *Phascolarctobacterium* | - |
| OTU 19 | 46100 | 29040 | 21090 | 15720 | 4699 | 16980 | 2219 | 18590 | 22570 | 3193 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | - |
| OTU 20 | 181 | 34 | 507 | 625 | 20 | 0 | 0 | 1 | 18 | 10660 | *Synergistetes* | *Synergistia* | *Synergistales* | *Synergistaceae* | - | - |
| OTU 21 | 1 | 0 | 0 | 0 | 0 | 1 | 268 | 2074 | 1185 | 8 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Peptoniphilus* | - |
| OTU 22 | 0 | 13 | 15 | 74 | 0 | 0 | 1 | 0 | 0 | 3770 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 23 | 1 | 0 | 0 | 58 | 0 | 0 | 2 | 1 | 68 | 4108 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 24 | 53 | 6 | 0 | 4 | 0 | 0 | 1 | 0 | 4454 | 45 | *Proteobacteria* | *Gammaproteobacteria* | *Enterobacteriales* | *Enterobacteriaceae* | - | - |
| OTU 25 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2761 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 26 | 691 | 511 | 1141 | 215 | 1 | 445 | 4 | 142 | 39 | 15 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Propionibacteriaceae* | - | - |
| OTU 27 | 0 | 0 | 21 | 0 | 0 | 4 | 0 | 1 | 0 | 2103 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 28 | 0 | 7 | 0 | 55 | 13 | 0 | 5 | 0 | 25 | 2898 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 29 | 1 | 0 | 0 | 0 | 0 | 0 | 195 | 90 | 2284 | 29 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Gallicola* | - |
| OTU 30 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2293 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Dysgonomonas* | - |
| OTU 31 | 0 | 0 | 16 | 0 | 0 | 0 | 0 | 0 | 2 | 1696 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 32 | 36 | 4 | 7 | 151 | 1 | 0 | 0 | 0 | 36 | 2767 | *Fusobacteria* | *Fusobacteriia* | *Fusobacteriales* | *Fusobacteriaceae* | *Fusobacterium* | - |
| OTU 33 | 85 | 3 | 276 | 131 | 21 | 0 | 11 | 0 | 62 | 1414 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | *Desulfovibrio* | - |
| OTU 34 | 3 | 1 | 13 | 0 | 30 | 1 | 0 | 0 | 25 | 1764 | *Verrucomicrobia* | *Verrucomicrobiae* | *Verrucomicrobiales* | *Verrucomicrobiaceae* | *Akkermansia* | *muciniphila* |
| OTU 35 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 1577 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 36 | 5 | 0 | 477 | 45 | 0 | 0 | 0 | 1 | 1 | 3139 | *Proteobacteria* | *Gammaproteobacteria* | *Pasteurellales* | *Pasteurellaceae* | *Lonepinella* | *koalarum* |
| OTU 37 | 73 | 25 | 203 | 710 | 121 | 53 | 13 | 0 | 8 | 175 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylobacteriaceae* | *Methylobacterium* | - |
| OTU 38 | 0 | 0 | 0 | 0 | 0 | 46 | 130 | 396 | 913 | 16 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Peptoniphilus* | - |
| OTU 39 | 71 | 3 | 27 | 574 | 3 | 0 | 0 | 0 | 40 | 1191 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | *Ruminococcus* | - |
| OTU 40 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2102 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 41 | 46 | 0 | 60 | 315 | 1 | 0 | 2 | 0 | 1 | 1875 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 42 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2454 | 0 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Peptococcaceae* | *Peptococcus* | - |
| OTU 43 | 0 | 0 | 0 | 0 | 17 | 0 | 0 | 0 | 0 | 1748 | *Planctomycetes* | *vadinHA49* | *PeHg47* | - | - | - |
| OTU 44 | 0 | 1 | 30 | 0 | 0 | 0 | 0 | 0 | 6 | 1134 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 45 | 139 | 27 | 180 | 1414 | 162 | 59 | 29 | 2 | 10 | 344 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylobacteriaceae* | *Methylobacterium* | - |
| OTU 46 | 16 | 2 | 21 | 0 | 7 | 0 | 0 | 0 | 1 | 794 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 47 | 0 | 0 | 0 | 0 | 0 | 0 | 31 | 418 | 130 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *ph2* | - |
| OTU 48 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 30 | 412 | *Synergistetes* | *Synergistia* | *Synergistales* | *Synergistaceae* | *vadinCA02* | - |

| OTU | K1 | K2 | K3 | K4 | K5 | K31 | K49 | K55 | K59 | K70 | Phylum | Class | Order | Family | Genus | Species |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| OTU 49 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 630 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 547 | 0 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Actinomycetaceae* | *Mobiluncus* | - |
| OTU 51 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 169 | 149 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Peptoniphilus* | - |
| OTU 52 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 430 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Veillonellaceae* | *Dialister* | - |
| OTU 53 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 386 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *ph2* | - |
| OTU 54 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 308 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 55 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 261 | 0 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Porphyromonas* | - |
| OTU 56 | 12 | 0 | 0 | 187 | 0 | 0 | 1 | 0 | 0 | 69 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 57 | 15 | 0 | 34 | 92 | 9 | 2 | 0 | 0 | 2 | 22 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylobacteriaceae* | - | - |
| OTU 58 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 218 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 59 | 31 | 0 | 0 | 19 | 0 | 0 | 3 | 0 | 0 | 230 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 60 | 0 | 0 | 0 | 318 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Streptococcaceae* | *Streptococcus* | - |
| OTU 61 | 8 | 0 | 1 | 136 | 10 | 0 | 3 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 62 | 25 | 4 | 30 | 35 | 0 | 86 | 0 | 0 | 0 | 29 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Moraxellaceae* | *Acinetobacter* | *rhizosphaerae* |
| OTU 63 | 37 | 0 | 0 | 255 | 1 | 0 | 0 | 0 | 0 | 63 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 64 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 229 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 65 | 0 | 0 | 0 | 0 | 1 | 60 | 0 | 407 | 0 | 0 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Alcaligenaceae* | *Sutterella* | - |
| OTU 66 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 129 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *ph2* | - |
| OTU 67 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 266 | 0 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Prevotellaceae* | *Prevotella* | - |
| OTU 69 | 41 | 0 | 0 | 89 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 70 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 116 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | *Oscillospira* | - |
| OTU 71 | 25 | 0 | 0 | 127 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 72 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 156 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 73 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 143 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Rikenellaceae* | - | - |
| OTU 74 | 0 | 0 | 0 | 88 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 75 | 31 | 0 | 0 | 166 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 76 | 7 | 0 | 19 | 79 | 1 | 0 | 1 | 1 | 0 | 35 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Pseudomonadaceae* | *Pseudomonas* | - |
| OTU 77 | 7 | 2 | 10 | 24 | 42 | 10 | 6 | 0 | 0 | 63 | *Proteobacteria* | *Alphaproteobacteria* | *Sphingomonadales* | *Sphingomonadaceae* | *Sphingomonas* | *yabuuchiae* |
| OTU 78 | 2 | 1 | 9 | 165 | 6 | 0 | 1 | 2 | 1 | 1 | *Firmicutes* | *Bacilli* | *Bacillales* | *Staphylococcaceae* | *Staphylococcus* | - |
| OTU 79 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 94 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 80 | 0 | 0 | 0 | 79 | 0 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Hyphomicrobiaceae* | *Pedomicrobium* | - |
| OTU 81 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 108 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Oxalobacteraceae* | - | - |
| OTU 82 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 95 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 83 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 98 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | *fragilis* |
| OTU 84 | 0 | 0 | 80 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 86 | 29 | 0 | 73 | 43 | 0 | 0 | 0 | 0 | 0 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 87 | 0 | 0 | 0 | 67 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 88 | 0 | 70 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 89 | 0 | 0 | 0 | 46 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 90 | 10 | 0 | 0 | 0 | 0 | 32 | 0 | 118 | 0 | 0 | *Actinobacteria* | *Coriobacteriia* | *Coriobacteriales* | *Coriobacteriaceae* | - | - |
| OTU 91 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 153 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfarculales* | *Desulfarculaceae* | - | - |
| OTU 92 | 5 | 0 | 11 | 50 | 0 | 0 | 0 | 0 | 2 | 17 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Comamonadaceae* | *Comamonas* | - |
| OTU 93 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 144 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 94 | 0 | 0 | 0 | 131 | 14 | 0 | 0 | 1 | 0 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylocystaceae* | - | - |
| OTU 95 | 0 | 0 | 18 | 39 | 0 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Rhizobiaceae* | *Rhizobium* | *leguminosarum* |
| OTU 96 | 0 | 1 | 0 | 239 | 0 | 0 | 0 | 0 | 1 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 97 | 1 | 2 | 4 | 55 | 3 | 0 | 0 | 0 | 1 | 17 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Comamonadaceae* | - | - |
| OTU 98 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 140 | *Planctomycetes* | *vadinHA49* | *PeHg47* | - | - | - |
| OTU 99 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 71 | 0 | 0 | *Actinobacteria* | *Coriobacteriia* | *Coriobacteriales* | *Coriobacteriaceae* | - | - |
| OTU 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 64 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Mogibacteriaceae]* | *Mogibacterium* | - |
| OTU 101 | 0 | 0 | 2 | 12 | 11 | 6 | 2 | 0 | 0 | 17 | *Bacteroidetes* | *[Saprospirae]* | *[Saprospirales]* | *Chitinophagaceae* | *Sediminibacterium* | - |
| OTU 102 | 0 | 0 | 1 | 52 | 2 | 2 | 1 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | *Blautia* | *producta* |

| OTU | K1 | K2 | K3 | K4 | K5 | K31 | K49 | K55 | K59 | K70 | Phylum | Class | Order | Family | Genus | Species |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| OTU 103 | 14 | 0 | 45 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Unassigned | - | - | - | - | - |
| OTU 105 | 27 | 46 | 18 | 30 | 0 | 122 | 342 | 403 | 99 | 11 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | - |
| OTU 106 | 2 | 0 | 11 | 0 | 41 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Burkholderiaceae* | *Burkholderia* | - |
| OTU 107 | 0 | 0 | 0 | 63 | 0 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Sphingomonadales* | *Sphingomonadaceae* | *Sphingobium* | - |
| OTU 108 | 0 | 0 | 0 | 34 | 0 | 3 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 109 | 0 | 0 | 0 | 0 | 0 | 0 | 36 | 15 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Peptostreptococcaceae* | *Peptostreptococcus* | - |
| OTU 110 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 27 | *Acidobacteria* | *Acidobacteriia* | *Acidobacteriales* | *Acidobacteriaceae* | - | - |
| OTU 111 | 0 | 3 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 37 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 112 | 2 | 0 | 0 | 30 | 0 | 0 | 0 | 0 | 0 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Nocardiaceae* | *Rhodococcus* | *fascians* |
| OTU 113 | 0 | 0 | 5 | 14 | 9 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Rhizobiaceae* | *Agrobacterium* | - |
| OTU 114 | 0 | 0 | 0 | 40 | 10 | 0 | 0 | 0 | 1 | 0 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Oxalobacteraceae* | - | - |
| OTU 115 | 3 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 43 | *Cyanobacteria* | *4C0d-2* | *YS2* | - | - | - |
| OTU 116 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 39 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 117 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 32 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfobacterales* | *Desulfobacteraceae* | - | - |
| OTU 118 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 36 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 119 | 2 | 0 | 0 | 22 | 0 | 0 | 0 | 0 | 0 | 7 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Oxalobacteraceae* | *Ralstonia* | - |
| OTU 120 | 0 | 0 | 46 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *[Paraprevotellaceae]* | *Paraprevotella* | - |
| OTU 121 | 0 | 0 | 0 | 37 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Erysipelotrichi* | *Erysipelotrichales* | *Erysipelotrichaceae* | *Coprobacillus* | - |
| OTU 122 | 0 | 0 | 0 | 59 | 0 | 0 | 0 | 0 | 0 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 123 | 0 | 0 | 0 | 6 | 13 | 1 | 17 | 1 | 5 | 92 | Unassigned | - | - | - | - | - |
| OTU 124 | 33 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Micrococcaceae* | *Rothia* | *mucilaginosa* |
| OTU 125 | 0 | 0 | 54 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfarculales* | *Desulfarculaceae* | - | - |
| OTU 126 | 0 | 0 | 0 | 45 | 1 | 1 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 127 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 25 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 128 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 53 | *Planctomycetes* | *vadinHA49* | *PeHg47* | - | - | - |
| OTU 129 | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Micrococcaceae* | - | - |
| OTU 130 | 0 | 0 | 1 | 26 | 0 | 0 | 1 | 0 | 0 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 131 | 0 | 0 | 15 | 0 | 17 | 2 | 1 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 132 | 4 | 1 | 31 | 0 | 0 | 4 | 1 | 0 | 0 | 16 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Comamonadaceae* | - | - |
| OTU 133 | 0 | 0 | 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Bacteroidetes* | *Cytophagia* | *Cytophagales* | *Cytophagaceae* | *Hymenobacter* | - |
| OTU 134 | 0 | 0 | 0 | 25 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Eubacteriaceae* | *Anaerofustis* | - |
| OTU 135 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 18 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | *uniformis* |
| OTU 136 | 0 | 0 | 0 | 20 | 0 | 3 | 0 | 0 | 0 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 137 | 13 | 0 | 19 | 0 | 7 | 0 | 0 | 0 | 0 | 17 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 138 | 0 | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | *Oscillospira* | - |
| OTU 139 | 0 | 0 | 0 | 18 | 0 | 0 | 0 | 0 | 0 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Micrococcaceae* | *Rothia* | *dentocariosa* |
| OTU 140 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 14 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Moraxellaceae* | *Acinetobacter* | - |
| OTU 141 | 0 | 0 | 0 | 21 | 3 | 2 | 0 | 0 | 0 | 0 | *Actinobacteria* | *Thermoleophilia* | *Solirubrobacterales* | - | - | - |
| OTU 142 | 0 | 0 | 0 | 0 | 0 | 0 | 28 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 143 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | *ovatus* |
| OTU 144 | 5 | 0 | 0 | 0 | 23 | 0 | 0 | 0 | 0 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 145 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 17 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 146 | 8 | 0 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Prevotellaceae* | *Prevotella* | - |
| OTU 147 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 23 | Unassigned | - | - | - | - | - |
| OTU 148 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 31 | 21 | 0 | Unassigned | - | - | - | - | - |
| OTU 149 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | *Desulfovibrio* | - |
| OTU 150 | 0 | 0 | 34 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Bacteroidetes* | *Cytophagia* | *Cytophagales* | *Cytophagaceae* | *Hymenobacter* | - |
| OTU 151 | 0 | 0 | 23 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Kineosporiaceae* | - | - |
| OTU 152 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 31 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 153 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 17 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | *Bilophila* | - |
| OTU 154 | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 155 | 34 | 104 | 14 | 15 | 0 | 103 | 13 | 440 | 0 | 0 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Carnobacteriaceae* | *Trichococcus* | - |

| OTU | K1 | K2 | K3 | K4 | K5 | K31 | K49 | K55 | K59 | K70 | Phylum | Class | Order | Family | Genus | Species |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| OTU 156 | 5 | 1 | 0 | 0 | 11 | 0 | 0 | 0 | 0 | 0 | *Armatimonadetes* | *[Fimbriimonadia]* | *[Fimbriimonadales]* | *[Fimbriimonadaceae]* | *Fimbriimonas* | - |
| OTU 157 | 4 | 0 | 0 | 0 | 0 | 7 | 3 | 0 | 2 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 158 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylocystaceae* | - | - |
| OTU 159 | 0 | 1 | 3 | 0 | 0 | 5 | 1 | 44 | 19 | 0 | Unassigned | - | - | - | - | - |
| OTU 160 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 161 | 0 | 0 | 1 | 0 | 18 | 2 | 0 | 0 | 1 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | *perfringens* |
| OTU 162 | 0 | 1 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 163 | 14 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Gammaproteobacteria* | *Oceanospirillales* | *Halomonadaceae* | - | - |
| OTU 164 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 15 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Moraxellaceae* | *Acinetobacter* | - |
| OTU 165 | 10 | 4 | 25 | 12 | 0 | 58 | 7 | 25 | 7 | 0 | Unassigned | - | - | - | - | - |
| OTU 166 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 22 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 167 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 20 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | *Dorea* | - |
| OTU 168 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | *Oscillospira* | - |
| OTU 169 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Intrasporangiaceae* | *Terracoccus* | - |
| OTU 170 | 0 | 2 | 1 | 0 | 18 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Enterococcaceae* | - | - |
| OTU 171 | 2 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | *Blautia* | - |
| OTU 172 | 4 | 0 | 3 | 7 | 13 | 0 | 2 | 0 | 0 | 0 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Oxalobacteraceae* | *Ralstonia* | - |
| OTU 173 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | *Deferribacteres* | *Deferribacteres* | *Deferribacterales* | *Deferribacteraceae* | *Mucispirillum* | - |
| OTU 174 | 0 | 0 | 0 | 0 | 11 | 2 | 0 | 0 | 0 | 23 | Unassigned | - | - | - | - | - |
| OTU 175 | 5 | 0 | 0 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Oxalobacteraceae* | *Oxalobacter* | *formigenes* |
| OTU 176 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 177 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 27 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | *Bilophila* | - |
| OTU 178 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | Unassigned | - | - | - | - | - |
| OTU 179 | 6 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 180 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhodospirillales* | *Rhodospirillaceae* | - | - |
| OTU 181 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 48 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 182 | 0 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Cyanobacteria* | *4C0d-2* | *MLE1-12* | - | - | - |
| OTU 183 | 1 | 5 | 0 | 1 | 0 | 0 | 0 | 1 | 3 | 63 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Pseudomonadaceae* | *Pseudomonas* | - |
| OTU 184 | 6 | 0 | 0 | 0 | 9 | 0 | 0 | 0 | 0 | 5 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Comamonadaceae* | - | - |
| OTU 185 | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhodospirillales* | *Acetobacteraceae* | - | - |
| OTU 186 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Peptoniphilus* | - |
| OTU 187 | 4 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 8 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Moraxellaceae* | *Enhydrobacter* | - |
| OTU 188 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Betaproteobacteria* | *Neisseriales* | *Neisseriaceae* | *Neisseria* | *subflava* |
| OTU 189 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhodospirillales* | *Acetobacteraceae* | - | - |
| OTU 190 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Nocardioidaceae* | - | - |
| OTU 191 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 39 | *Planctomycetes* | *vadinHA49* | *PeHg47* | - | - | - |
| OTU 192 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | *ovatus* |
| OTU 193 | 0 | 0 | 9 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 194 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 25 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Rikenellaceae* | - | - |
| OTU 195 | 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 196 | 0 | 1 | 1 | 3 | 4 | 5 | 3 | 1 | 1 | 1 | *Proteobacteria* | *Gammaproteobacteria* | *Enterobacteriales* | *Enterobacteriaceae* | *Plesiomonas* | *shigelloides* |
| OTU 197 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | *Acidobacteria* | *Acidobacteriia* | *Acidobacteriales* | *Acidobacteriaceae* | *Terriglobus* | - |
| OTU 199 | 0 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 200 | 2 | 18 | 0 | 0 | 0 | 7 | 0 | 66 | 0 | 0 | Unassigned | - | - | - | - | - |
| OTU 201 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 79 | 0 | 0 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | - |
| OTU 202 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | *Coprococcus* | - |
| OTU 204 | 4 | 0 | 0 | 1 | 4 | 2 | 0 | 2 | 0 | 1 | *Fusobacteria* | *Fusobacteriia* | *Fusobacteriales* | *Leptotrichiaceae* | - | - |
| OTU 205 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 206 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 16 | *Proteobacteria* | *Gammaproteobacteria* | *Pasteurellales* | *Pasteurellaceae* | *Aggregatibacter* | - |
| OTU 207 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | Unassigned | - | - | - | - | - |
| OTU 209 | 0 | 1 | 2 | 1 | 4 | 2 | 1 | 1 | 1 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Peptostreptococcaceae* | - | - |
| OTU 210 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Rikenellaceae* | - | - |

| OTU | K1 | K2 | K3 | K4 | K5 | K31 | K49 | K55 | K59 | K70 | Phylum | Class | Order | Family | Genus | Species |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| OTU 212 | 0 | 0 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Dehalobacteriaceae* | - | - |
| OTU 213 | 0 | 19 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 19 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 214 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Moraxellaceae* | *Acinetobacter* | - |
| OTU 215 | 7 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Veillonellaceae* | *Veillonella* | *dispar* |
| OTU 216 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | Unassigned | - | - | - | - | - |
| OTU 217 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Bradyrhizobiaceae* | - | - |
| OTU 218 | 0 | 0 | 0 | 0 | 6 | 0 | 2 | 0 | 2 | 0 | *Firmicutes* | *Bacilli* | *Bacillales* | - | - | - |
| OTU 219 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhodobacterales* | *Rhodobacteraceae* | *Amaricoccus* | - |
| OTU 220 | 0 | 8 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 45 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 221 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | Unassigned | - | - | - | - | - |
| OTU 222 | 1 | 10 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Unassigned | - | - | - | - | - |
| OTU 223 | 12 | 0 | 0 | 0 | 5 | 0 | 0 | 3 | 0 | 0 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Streptococcaceae* | *Streptococcus* | - |
| OTU 224 | 0 | 1 | 0 | 0 | 0 | 7 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Sphingomonadales* | *Sphingomonadaceae* | - | - |
| OTU 225 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Betaproteobacteria* | *Gallionellales* | *Gallionellaceae* | *Gallionella* | - |
| OTU 226 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | Unassigned | - | - | - | - | - |
| OTU 227 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 492 | 0 | 1 | Unassigned | - | - | - | - | - |
| OTU 228 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | *Ruminococcus* | - |
| OTU 229 | 0 | 5 | 0 | 4 | 0 | 5 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | - | - |
| OTU 230 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 3 | Unassigned | - | - | - | - | - |
| OTU 231 | 0 | 0 | 9 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylocystaceae* | - | - |
| OTU 232 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhodospirillales* | *Acetobacteraceae* | - | - |
| OTU 233 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 2 | Unassigned | - | - | - | - | - |
| OTU 234 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 235 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 236 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Rikenellaceae* | - | - |
| OTU 237 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 238 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 121 | 0 | 0 | Unassigned | - | - | - | - | - |
| OTU 239 | 0 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | *Fusobacteria* | *Fusobacteriia* | *Fusobacteriales* | *Fusobacteriaceae* | *Fusobacterium* | - |
| OTU 240 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 750 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 241 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Anaerococcus* | - |
| OTU 242 | 1 | 0 | 6 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Comamonadaceae* | - | - |
| OTU 243 | 1 | 2 | 2 | 0 | 5 | 1 | 0 | 1 | 1 | 0 | *Fusobacteria* | *Fusobacteriia* | *Fusobacteriales* | *Fusobacteriaceae* | *Cetobacterium* | *somerae* |
| OTU 244 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 17 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 245 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | *Synergistetes* | *Synergistia* | *Synergistales* | *Synergistaceae* | - | - |
| OTU 246 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Actinobacteria* | *Coriobacteriia* | *Coriobacteriales* | *Coriobacteriaceae* | - | - |
| OTU 247 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Rikenellaceae* | - | - |
| OTU 248 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 249 | 0 | 6 | 0 | 0 | 0 | 1 | 1 | 5 | 7 | 0 | Unassigned | - | - | - | - | - |
| OTU 250 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 9 | 47 | *Planctomycetes* | *vadinHA49* | *PeHg47* | - | - | - |
| OTU 251 | 0 | 0 | 0 | 40 | 0 | 0 | 0 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 252 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 19 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Veillonellaceae* | - | - |
| OTU 253 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | Unassigned | - | - | - | - | - |
| OTU 254 | 1 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Actinomycetaceae* | *Actinomyces* | - |
| OTU 255 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 256 | 0 | 0 | 0 | 0 | 5 | 0 | 1 | 0 | 0 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhodospirillales* | *Rhodospirillaceae* | - | - |
| OTU 257 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | Unassigned | - | - | - | - | - |
| OTU 258 | 1 | 0 | 1 | 0 | 0 | 0 | 7 | 0 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | *Roseburia* | - |
| OTU 259 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 260 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Micrococcaceae* | *Kocuria* | *palustris* |
| OTU 261 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 16 | 0 | Unassigned | - | - | - | - | - |

^ Classification could not be made to this level

**Appendix 7.** DESeq2 normalised abundance of merged reads clustered to assigned operational taxonomic units (OTUs). OTUs were assigned a taxonomy (with 90% identity) from the Greengenes curated database (DeSantis *et al.* 2006). Koalas K1 – K5 were clinically normal (wet bottom absent), whilst koalas K31 – K70 had wet bottom.

| OTU | K1 | K2 | K3 | K4 | K5 | K31 | K49 | K55 | K59 | K70 | Phylum | Class | Order | Family | Genus | Species |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| OTU 1 | 14.6 | 14.4 | 13.7 | 15.8 | 11.4 | 17.1 | 16.8 | 18.4 | 18.3 | 12.2 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | -^ |
| OTU 2 | 17.3 | 16.8 | 16.4 | 15.7 | 18.4 | 12.2 | 10.5 | 10.7 | 15.0 | 14.2 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | - |
| OTU 3 | 7.0 | 9.2 | 11.0 | 5.4 | 9.5 | 8.4 | 17.9 | 4.7 | 6.0 | 7.8 | *Proteobacteria* | *Gammaproteobacteria* | *Enterobacteriales* | *Enterobacteriaceae* | - | - |
| OTU 4 | 14.6 | 14.8 | 15.0 | 13.7 | 15.0 | 8.8 | 8.3 | 6.1 | 10.8 | 12.5 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | - |
| OTU 5 | 10.4 | 8.2 | 11.9 | 11.4 | 9.4 | 1.9 | 7.5 | 5.0 | 10.6 | 16.3 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 6 | 11.1 | 10.9 | 13.6 | 15.3 | 10.2 | 14.7 | 11.7 | 11.7 | 10.8 | 10.7 | Unassigned | - | - | - | - | - |
| OTU 7 | 13.3 | 13.9 | 13.3 | 13.2 | 8.3 | 12.8 | 9.7 | 12.3 | 6.6 | 9.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 8 | 9.1 | 2.0 | 9.5 | 9.0 | 8.1 | 4.7 | 4.8 | 3.4 | 8.5 | 14.8 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 9 | 13.5 | 13.6 | 11.7 | 1.7 | 3.4 | 11.2 | 8.7 | 11.5 | 5.5 | 9.8 | *Actinobacteria* | *Coriobacteriia* | *Coriobacteriales* | *Coriobacteriaceae* | - | - |
| OTU 10 | 14.0 | 13.5 | 13.1 | 14.9 | 10.1 | 15.0 | 12.0 | 14.1 | 11.4 | 11.1 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Facklamia* | - |
| OTU 11 | 9.6 | 7.1 | 1.8 | 10.7 | 7.6 | 3.5 | 6.3 | 1.9 | 10.3 | 14.7 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 12 | 7.8 | 7.2 | 10.6 | 10.9 | 7.8 | 1.9 | 3.5 | 3.4 | 8.1 | 14.0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 13 | 7.9 | 6.8 | 9.4 | 9.3 | 6.4 | 1.9 | 2.0 | 5.0 | 10.0 | 13.1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 14 | 10.0 | 12.4 | 10.8 | 1.7 | 1.9 | 12.0 | 10.9 | 8.8 | 10.1 | 7.8 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 15 | 7.6 | 10.7 | 9.7 | 1.7 | 5.3 | 11.6 | 8.5 | 12.7 | 4.2 | 7.5 | *TM7* | *TM7-3* | *I025* | *Rs-045* | - | - |
| OTU 16 | 4.6 | 3.5 | 1.8 | 3.2 | 1.9 | 10.1 | 7.7 | 14.4 | 9.2 | 7.6 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Porphyromonas* | - |
| OTU 17 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 5.0 | 8.9 | 12.9 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 18 | 8.1 | 4.8 | 8.3 | 9.3 | 1.9 | 1.9 | 3.5 | 1.9 | 7.5 | 13.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Veillonellaceae* | *Phascolarctobacterium* | - |
| OTU 19 | 15.9 | 15.2 | 14.4 | 13.9 | 12.4 | 14.2 | 11.6 | 14.3 | 14.6 | 11.8 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | - |
| OTU 20 | 9.1 | 7.5 | 10.0 | 10.2 | 6.9 | 1.9 | 2.0 | 3.4 | 6.8 | 13.3 | *Synergistetes* | *Synergistia* | *Synergistales* | *Synergistaceae* | - | - |
| OTU 21 | 3.3 | 2.0 | 1.8 | 1.7 | 1.9 | 3.5 | 9.6 | 11.5 | 10.9 | 5.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Peptoniphilus* | - |
| OTU 22 | 1.8 | 6.5 | 6.5 | 8.2 | 1.9 | 1.9 | 3.5 | 1.9 | 1.9 | 12.0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 23 | 3.3 | 2.0 | 1.8 | 7.9 | 1.9 | 1.9 | 4.3 | 3.4 | 8.2 | 12.1 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 24 | 7.9 | 5.6 | 1.8 | 4.9 | 1.9 | 1.9 | 3.5 | 1.9 | 12.3 | 7.6 | *Proteobacteria* | *Gammaproteobacteria* | *Enterobacteriales* | *Enterobacteriaceae* | - | - |
| OTU 25 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 3.5 | 2.0 | 11.8 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 26 | 10.3 | 10.1 | 10.8 | 9.2 | 3.4 | 10.0 | 5.1 | 8.9 | 7.6 | 6.4 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Propionibacteriaceae* | - | - |
| OTU 27 | 1.8 | 2.0 | 6.9 | 1.7 | 1.9 | 5.1 | 2.0 | 3.4 | 1.9 | 11.3 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 28 | 1.8 | 5.8 | 1.8 | 7.9 | 6.4 | 1.9 | 5.4 | 1.9 | 7.1 | 11.7 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 29 | 3.3 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 9.3 | 8.5 | 11.6 | 7.1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Gallicola* | - |
| OTU 30 | 3.3 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 3.4 | 1.9 | 11.4 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Dysgonomonas* | - |
| OTU 31 | 1.8 | 2.0 | 6.6 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 4.2 | 11.1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 32 | 7.5 | 5.1 | 5.6 | 8.9 | 3.4 | 1.9 | 2.0 | 1.9 | 7.5 | 11.6 | *Fusobacteria* | *Fusobacteriia* | *Fusobacteriales* | *Fusobacteriaceae* | *Fusobacterium* | - |
| OTU 33 | 8.3 | 4.8 | 9.5 | 8.8 | 7.0 | 1.9 | 6.3 | 1.9 | 8.1 | 10.9 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | *Desulfovibrio* | - |
| OTU 34 | 4.6 | 3.5 | 6.3 | 1.7 | 7.3 | 3.5 | 2.0 | 1.9 | 7.1 | 11.1 | *Verrucomicrobia* | *Verrucomicrobiae* | *Verrucomicrobiales* | *Verrucomicrobiaceae* | *Akkermansia* | *muciniphila* |
| OTU 35 | 1.8 | 2.0 | 1.8 | 6.0 | 1.9 | 1.9 | 2.0 | 11.2 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 36 | 5.2 | 2.0 | 10.0 | 7.7 | 1.9 | 1.9 | 2.0 | 3.4 | 3.4 | 11.8 | *Proteobacteria* | *Gammaproteobacteria* | *Pasteurellales* | *Pasteurellaceae* | *Lonepinella* | *koalarum* |
| OTU 37 | 8.2 | 7.2 | 9.2 | 10.3 | 8.8 | 7.9 | 6.5 | 1.9 | 5.8 | 9.0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylobacteriaceae* | *Methylobacterium* | - |
| OTU 38 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 7.8 | 8.9 | 9.9 | 10.6 | 6.5 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Peptoniphilus* | - |
| OTU 39 | 8.2 | 4.8 | 7.1 | 10.1 | 4.7 | 1.9 | 2.0 | 1.9 | 7.6 | 10.8 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | *Ruminococcus* | - |
| OTU 40 | 4.9 | 3.5 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 11.3 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 41 | 7.7 | 2.0 | 8.0 | 9.6 | 3.4 | 1.9 | 4.3 | 1.9 | 3.4 | 11.2 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 42 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 3.5 | 11.7 | 1.9 | 3.1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Peptococcaceae* | *Peptococcus* | - |
| OTU 43 | 1.8 | 2.0 | 1.8 | 1.7 | 6.7 | 1.9 | 2.0 | 1.9 | 1.9 | 11.1 | *Planctomycetes* | *vadinHA49* | *PeHg47* | - | - | - |
| OTU 44 | 1.8 | 3.5 | 7.3 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 5.5 | 10.7 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 45 | 8.8 | 7.3 | 9.1 | 11.0 | 9.0 | 8.1 | 7.3 | 4.2 | 6.1 | 9.6 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylobacteriaceae* | *Methylobacterium* | - |
| OTU 46 | 6.6 | 4.3 | 6.9 | 1.7 | 5.7 | 1.9 | 2.0 | 1.9 | 3.4 | 10.4 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 47 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 7.4 | 9.9 | 8.8 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *ph2* | - |
| OTU 48 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 3.4 | 7.3 | 9.8 | *Synergistetes* | *Synergistia* | *Synergistales* | *Synergistaceae* | *vadinCA02* | - |

| OTU | K1 | K2 | K3 | K4 | K5 | K31 | K49 | K55 | K59 | K70 | Phylum | Class | Order | Family | Genus | Species |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| OTU 49 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 3.4 | 1.9 | 10.2 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 50 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 10.2 | 1.9 | 1.6 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Actinomycetaceae* | *Mobiluncus* | - |
| OTU 51 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 6.6 | 9.1 | 9.0 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Peptoniphilus* | - |
| OTU 52 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 9.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Veillonellaceae* | *Dialister* | - |
| OTU 53 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 9.8 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *ph2* | - |
| OTU 54 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 3.4 | 9.5 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 55 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 9.5 | 1.9 | 1.6 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Porphyromonas* | - |
| OTU 56 | 6.3 | 2.0 | 1.8 | 9.1 | 1.9 | 1.9 | 3.5 | 1.9 | 1.9 | 8.0 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 57 | 6.5 | 2.0 | 7.4 | 8.4 | 6.0 | 4.2 | 2.0 | 1.9 | 4.2 | 6.8 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylobacteriaceae* | - | - |
| OTU 58 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 9.3 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 59 | 7.3 | 2.0 | 1.8 | 6.7 | 1.9 | 1.9 | 4.8 | 1.9 | 1.9 | 9.2 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 60 | 1.8 | 2.0 | 1.8 | 9.6 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Streptococcaceae* | *Streptococcus* | - |
| OTU 61 | 5.8 | 2.0 | 3.3 | 8.8 | 6.1 | 1.9 | 4.8 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 62 | 7.1 | 5.1 | 7.3 | 7.4 | 1.9 | 8.4 | 2.0 | 1.9 | 1.9 | 7.1 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Moraxellaceae* | *Acinetobacter* | *rhizosphaerae* |
| OTU 63 | 7.5 | 2.0 | 1.8 | 9.4 | 3.4 | 1.9 | 2.0 | 1.9 | 1.9 | 7.9 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 64 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 4.3 | 1.9 | 1.9 | 9.2 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 65 | 1.8 | 2.0 | 1.8 | 1.7 | 3.4 | 8.1 | 2.0 | 9.9 | 1.9 | 1.6 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Alcaligenaceae* | *Sutterella* | - |
| OTU 66 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 8.8 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *ph2* | - |
| OTU 67 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 9.5 | 1.9 | 1.6 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Prevotellaceae* | *Prevotella* | - |
| OTU 69 | 7.6 | 2.0 | 1.8 | 8.4 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 70 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 8.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | *Oscillospira* | - |
| OTU 71 | 7.1 | 2.0 | 1.8 | 8.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 72 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 8.9 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 73 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 6.6 | 8.8 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Rikenellaceae* | - | - |
| OTU 74 | 1.8 | 2.0 | 1.8 | 8.4 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 75 | 7.3 | 2.0 | 1.8 | 9.0 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 76 | 5.6 | 2.0 | 6.8 | 8.2 | 3.4 | 1.9 | 3.5 | 3.4 | 1.9 | 7.3 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Pseudomonadaceae* | *Pseudomonas* | - |
| OTU 77 | 5.6 | 4.3 | 6.0 | 7.0 | 7.7 | 6.1 | 5.6 | 1.9 | 1.9 | 7.9 | *Proteobacteria* | *Alphaproteobacteria* | *Sphingomonadales* | *Sphingomonadaceae* | *Sphingomonas* | *yabuuchiae* |
| OTU 78 | 4.1 | 3.5 | 5.9 | 9.0 | 5.5 | 1.9 | 3.5 | 4.2 | 3.4 | 3.1 | *Firmicutes* | *Bacilli* | *Bacillales* | *Staphylococcaceae* | *Staphylococcus* | - |
| OTU 79 | 5.4 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 4.2 | 8.3 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 80 | 1.8 | 2.0 | 1.8 | 8.2 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Hyphomicrobiaceae* | *Pedomicrobium* | - |
| OTU 81 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 8.5 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Oxalobacteraceae* | - | - |
| OTU 82 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 8.4 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 83 | 1.8 | 3.5 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 3.4 | 8.4 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | *fragilis* |
| OTU 84 | 1.8 | 2.0 | 8.3 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 86 | 7.2 | 2.0 | 8.2 | 7.6 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 87 | 1.8 | 2.0 | 1.8 | 8.1 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 88 | 1.8 | 8.3 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 89 | 1.8 | 2.0 | 1.8 | 7.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 90 | 6.0 | 2.0 | 1.8 | 1.7 | 1.9 | 7.4 | 2.0 | 8.7 | 1.9 | 1.6 | *Actinobacteria* | *Coriobacteriia* | *Coriobacteriales* | *Coriobacteriaceae* | - | - |
| OTU 91 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 8.8 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfarculales* | *Desulfarculaceae* | - | - |
| OTU 92 | 5.2 | 2.0 | 6.1 | 7.8 | 1.9 | 1.9 | 2.0 | 1.9 | 4.2 | 6.5 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Comamonadaceae* | *Comamonas* | - |
| OTU 93 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 3.4 | 8.8 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 94 | 1.8 | 2.0 | 1.8 | 8.8 | 6.5 | 1.9 | 2.0 | 3.4 | 1.9 | 1.6 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylocystaceae* | - | - |
| OTU 95 | 1.8 | 2.0 | 6.7 | 7.5 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Rhizobiaceae* | *Rhizobium* | *leguminosarum* |
| OTU 96 | 1.8 | 3.5 | 1.8 | 9.3 | 1.9 | 1.9 | 2.0 | 1.9 | 3.4 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 97 | 3.3 | 4.3 | 4.9 | 7.9 | 4.7 | 1.9 | 2.0 | 1.9 | 3.4 | 6.5 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Comamonadaceae* | - | - |
| OTU 98 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 3.4 | 8.7 | *Planctomycetes* | *vadinHA49* | *PeHg47* | - | - | - |
| OTU 99 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 8.2 | 1.9 | 1.6 | *Actinobacteria* | *Coriobacteriia* | *Coriobacteriales* | *Coriobacteriaceae* | - | - |
| OTU 100 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 8.1 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Mogibacteriaceae]* | *Mogibacterium* | - |
| OTU 101 | 1.8 | 2.0 | 4.1 | 6.2 | 6.2 | 5.5 | 4.3 | 1.9 | 1.9 | 6.5 | *Bacteroidetes* | *[Saprospirae]* | *[Saprospirales]* | *Chitinophagaceae* | *Sediminibacterium* | - |
| OTU 102 | 1.8 | 2.0 | 3.3 | 7.8 | 4.2 | 4.2 | 3.5 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | *Blautia* | *producta* |

| OTU | K1 | K2 | K3 | K4 | K5 | K31 | K49 | K55 | K59 | K70 | Phylum | Class | Order | Family | Genus | Species |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| OTU 103 | 6.4 | 2.0 | 7.7 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | Unassigned | - | - | - | - | - |
| OTU 105 | 7.1 | 7.8 | 6.7 | 7.2 | 1.9 | 8.8 | 9.8 | 9.9 | 8.6 | 6.0 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | - |
| OTU 106 | 4.1 | 2.0 | 6.1 | 1.7 | 7.7 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Burkholderiaceae* | *Burkholderia* | - |
| OTU 107 | 1.8 | 2.0 | 1.8 | 8.0 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Alphaproteobacteria* | *Sphingomonadales* | *Sphingomonadaceae* | *Sphingobium* | - |
| OTU 108 | 1.8 | 2.0 | 1.8 | 7.4 | 1.9 | 4.7 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 109 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 7.6 | 6.6 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Peptostreptococcaceae* | *Peptostreptococcus* | - |
| OTU 110 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 7.0 | *Acidobacteria* | *Acidobacteriia* | *Acidobacteriales* | *Acidobacteriaceae* | - | - |
| OTU 111 | 1.8 | 4.8 | 1.8 | 1.7 | 1.9 | 1.9 | 3.5 | 1.9 | 3.4 | 7.4 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 112 | 4.1 | 2.0 | 1.8 | 7.2 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Nocardiaceae* | *Rhodococcus* | *fascians* |
| OTU 113 | 1.8 | 2.0 | 5.2 | 6.4 | 6.0 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Rhizobiaceae* | *Agrobacterium* | - |
| OTU 114 | 1.8 | 2.0 | 1.8 | 7.5 | 6.1 | 1.9 | 2.0 | 1.9 | 3.4 | 1.6 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Oxalobacteraceae* | - | - |
| OTU 115 | 4.6 | 2.0 | 5.6 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 7.5 | *Cyanobacteria* | *4C0d-2* | *YS2* | - | - | - |
| OTU 116 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 7.4 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 117 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 7.2 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfobacterales* | *Desulfobacteraceae* | - | - |
| OTU 118 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 7.3 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 119 | 4.1 | 2.0 | 1.8 | 6.9 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 5.5 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Oxalobacteraceae* | *Ralstonia* | - |
| OTU 120 | 1.8 | 2.0 | 7.7 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *[Paraprevotellaceae]* | *Paraprevotella* | - |
| OTU 121 | 1.8 | 2.0 | 1.8 | 7.5 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Erysipelotrichi* | *Erysipelotrichales* | *Erysipelotrichaceae* | *Coprobacillus* | - |
| OTU 122 | 1.8 | 2.0 | 1.8 | 7.9 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 123 | 1.8 | 2.0 | 1.8 | 5.4 | 6.4 | 3.5 | 6.8 | 3.4 | 5.3 | 8.3 | Unassigned | - | - | - | - | - |
| OTU 124 | 7.4 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Micrococcaceae* | *Rothia* | *mucilaginosa* |
| OTU 125 | 1.8 | 2.0 | 7.9 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfarculales* | *Desulfarculaceae* | - | - |
| OTU 126 | 1.8 | 2.0 | 1.8 | 7.7 | 3.4 | 3.5 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 127 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 4.2 | 2.0 | 1.9 | 1.9 | 7.0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 128 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 7.8 | *Planctomycetes* | *vadinHA49* | *PeHg47* | - | - | - |
| OTU 129 | 1.8 | 2.0 | 6.4 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Micrococcaceae* | - | - |
| OTU 130 | 1.8 | 2.0 | 3.3 | 7.1 | 1.9 | 1.9 | 3.5 | 1.9 | 1.9 | 1.6 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 131 | 1.8 | 2.0 | 6.5 | 1.7 | 6.7 | 4.2 | 3.5 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 132 | 4.9 | 3.5 | 7.3 | 1.7 | 1.9 | 5.1 | 3.5 | 1.9 | 1.9 | 6.5 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Comamonadaceae* | - | - |
| OTU 133 | 1.8 | 2.0 | 7.0 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Bacteroidetes* | *Cytophagia* | *Cytophagales* | *Cytophagaceae* | *Hymenobacter* | - |
| OTU 134 | 1.8 | 2.0 | 1.8 | 7.0 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Eubacteriaceae* | *Anaerofustis* | - |
| OTU 135 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 6.8 | 1.6 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | *uniformis* |
| OTU 136 | 1.8 | 2.0 | 1.8 | 6.8 | 1.9 | 4.7 | 2.0 | 1.9 | 1.9 | 1.6 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 137 | 6.3 | 2.0 | 6.8 | 1.7 | 5.7 | 1.9 | 2.0 | 1.9 | 1.9 | 6.5 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 138 | 1.8 | 6.8 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | *Oscillospira* | - |
| OTU 139 | 1.8 | 2.0 | 1.8 | 6.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Micrococcaceae* | *Rothia* | *dentocariosa* |
| OTU 140 | 6.2 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 4.2 | 6.3 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Moraxellaceae* | *Acinetobacter* | - |
| OTU 141 | 1.8 | 2.0 | 1.8 | 6.9 | 4.7 | 4.2 | 2.0 | 1.9 | 1.9 | 1.6 | *Actinobacteria* | *Thermoleophilia* | *Solirubrobacterales* | - | - | - |
| OTU 142 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 7.3 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 143 | 1.8 | 2.0 | 1.8 | 6.3 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | *ovatus* |
| OTU 144 | 5.2 | 2.0 | 1.8 | 1.7 | 7.0 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 145 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 6.5 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 146 | 5.8 | 2.0 | 6.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Prevotellaceae* | *Prevotella* | - |
| OTU 147 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 6.9 | Unassigned | - | - | - | - | - |
| OTU 148 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 7.4 | 6.9 | 1.6 | Unassigned | - | - | - | - | - |
| OTU 149 | 1.8 | 2.0 | 5.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 5.6 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | *Desulfovibrio* | - |
| OTU 150 | 1.8 | 2.0 | 7.4 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Bacteroidetes* | *Cytophagia* | *Cytophagales* | *Cytophagaceae* | *Hymenobacter* | - |
| OTU 151 | 1.8 | 2.0 | 7.0 | 1.7 | 4.7 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Kineosporiaceae* | - | - |
| OTU 152 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 7.2 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 153 | 4.6 | 3.5 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 6.5 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | *Bilophila* | - |
| OTU 154 | 1.8 | 2.0 | 6.4 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 155 | 7.4 | 8.7 | 6.4 | 6.5 | 1.9 | 8.6 | 6.5 | 10.0 | 1.9 | 1.6 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Carnobacteriaceae* | *Trichococcus* | - |

| OTU | K1 | K2 | K3 | K4 | K5 | K31 | K49 | K55 | K59 | K70 | Phylum | Class | Order | Family | Genus | Species |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| OTU 156 | 5.2 | 3.5 | 1.8 | 1.7 | 6.2 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Armatimonadetes* | *[Fimbriimonadia]* | *[Fimbriimonadales]* | *[Fimbriimonadaceae]* | *Fimbriimonas* | - |
| OTU 157 | 4.9 | 2.0 | 1.8 | 1.7 | 1.9 | 5.7 | 4.8 | 1.9 | 4.2 | 3.1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 158 | 1.8 | 2.0 | 1.8 | 1.7 | 6.2 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylocystaceae* | - | - |
| OTU 159 | 1.8 | 3.5 | 4.6 | 1.7 | 1.9 | 5.3 | 3.5 | 7.7 | 6.8 | 1.6 | Unassigned | - | - | - | - | - |
| OTU 160 | 1.8 | 3.5 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 6.2 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 161 | 1.8 | 2.0 | 3.3 | 1.7 | 6.8 | 4.2 | 2.0 | 1.9 | 3.4 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | *perfringens* |
| OTU 162 | 1.8 | 3.5 | 1.8 | 6.2 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 163 | 6.4 | 2.0 | 1.8 | 1.7 | 5.0 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Gammaproteobacteria* | *Oceanospirillales* | *Halomonadaceae* | - | - |
| OTU 164 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 3.5 | 2.0 | 1.9 | 1.9 | 6.4 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Moraxellaceae* | *Acinetobacter* | - |
| OTU 165 | 6.0 | 5.1 | 7.1 | 6.2 | 1.9 | 8.0 | 5.8 | 7.1 | 5.7 | 1.6 | Unassigned | - | - | - | - | - |
| OTU 166 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 7.0 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 167 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 4.3 | 1.9 | 1.9 | 6.7 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | *Dorea* | - |
| OTU 168 | 1.8 | 2.0 | 1.8 | 6.3 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | *Oscillospira* | - |
| OTU 169 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 6.6 | 1.9 | 1.6 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Intrasporangiaceae* | *Terracoccus* | - |
| OTU 170 | 1.8 | 4.3 | 3.3 | 1.7 | 6.8 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Enterococcaceae* | - | - |
| OTU 171 | 4.1 | 2.0 | 1.8 | 5.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | *Blautia* | - |
| OTU 172 | 4.9 | 2.0 | 4.6 | 5.6 | 6.4 | 1.9 | 4.3 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Oxalobacteraceae* | *Ralstonia* | - |
| OTU 173 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 6.0 | *Deferribacteres* | *Deferribacteres* | *Deferribacterales* | *Deferribacteraceae* | *Mucispirillum* | - |
| OTU 174 | 1.8 | 2.0 | 1.8 | 1.7 | 6.2 | 4.2 | 2.0 | 1.9 | 1.9 | 6.9 | Unassigned | - | - | - | - | - |
| OTU 175 | 5.2 | 2.0 | 1.8 | 6.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Oxalobacteraceae* | *Oxalobacter* | *formigenes* |
| OTU 176 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 5.8 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 177 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 7.0 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | *Bilophila* | - |
| OTU 178 | 1.8 | 3.5 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 6.4 | Unassigned | - | - | - | - | - |
| OTU 179 | 5.4 | 2.0 | 3.3 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 180 | 5.6 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Alphaproteobacteria* | *Rhodospirillales* | *Rhodospirillaceae* | - | - |
| OTU 181 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 7.6 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 182 | 1.8 | 2.0 | 5.6 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Cyanobacteria* | *4C0d-2* | *MLE1-12* | - | - | - |
| OTU 183 | 3.3 | 5.4 | 1.8 | 3.2 | 1.9 | 1.9 | 2.0 | 3.4 | 4.7 | 7.9 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Pseudomonadaceae* | *Pseudomonas* | - |
| OTU 184 | 5.4 | 2.0 | 1.8 | 1.7 | 6.0 | 1.9 | 2.0 | 1.9 | 1.9 | 5.1 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Comamonadaceae* | - | - |
| OTU 185 | 1.8 | 2.0 | 5.4 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Alphaproteobacteria* | *Rhodospirillales* | *Acetobacteraceae* | - | - |
| OTU 186 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 5.5 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Peptoniphilus* | - |
| OTU 187 | 4.9 | 2.0 | 1.8 | 1.7 | 5.3 | 1.9 | 2.0 | 1.9 | 1.9 | 5.6 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Moraxellaceae* | *Enhydrobacter* | - |
| OTU 188 | 5.4 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Betaproteobacteria* | *Neisseriales* | *Neisseriaceae* | *Neisseria* | *subflava* |
| OTU 189 | 1.8 | 2.0 | 1.8 | 1.7 | 5.3 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Alphaproteobacteria* | *Rhodospirillales* | *Acetobacteraceae* | - | - |
| OTU 190 | 1.8 | 2.0 | 1.8 | 1.7 | 5.5 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Nocardioidaceae* | - | - |
| OTU 191 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 7.4 | *Planctomycetes* | *vadinHA49* | *PeHg47* | - | - | - |
| OTU 192 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 6.0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | *ovatus* |
| OTU 193 | 1.8 | 2.0 | 5.9 | 6.3 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 194 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 7.0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Rikenellaceae* | - | - |
| OTU 195 | 6.7 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 196 | 1.8 | 3.5 | 3.3 | 4.6 | 5.0 | 5.3 | 4.8 | 3.4 | 3.4 | 3.1 | *Proteobacteria* | *Gammaproteobacteria* | *Enterobacteriales* | *Enterobacteriaceae* | *Plesiomonas* | *shigelloides* |
| OTU 197 | 1.8 | 2.0 | 1.8 | 1.7 | 5.0 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Acidobacteria* | *Acidobacteriia* | *Acidobacteriales* | *Acidobacteriaceae* | *Terriglobus* | - |
| OTU 199 | 1.8 | 2.0 | 5.9 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 200 | 4.1 | 6.8 | 1.8 | 1.7 | 1.9 | 5.7 | 2.0 | 8.1 | 1.9 | 1.6 | Unassigned | - | - | - | - | - |
| OTU 201 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 8.3 | 1.9 | 1.6 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | - |
| OTU 202 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 6.1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | *Coprococcus* | - |
| OTU 204 | 4.9 | 2.0 | 1.8 | 3.2 | 5.0 | 4.2 | 2.0 | 4.2 | 1.9 | 3.1 | *Fusobacteria* | *Fusobacteriia* | *Fusobacteriales* | *Leptotrichiaceae* | - | - |
| OTU 205 | 1.8 | 4.3 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 4.4 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 206 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 6.5 | *Proteobacteria* | *Gammaproteobacteria* | *Pasteurellales* | *Pasteurellaceae* | *Aggregatibacter* | - |
| OTU 207 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 5.3 | Unassigned | - | - | - | - | - |
| OTU 209 | 1.8 | 3.5 | 4.1 | 3.2 | 5.0 | 4.2 | 3.5 | 3.4 | 3.4 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Peptostreptococcaceae* | - | - |
| OTU 210 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 5.6 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Rikenellaceae* | - | - |

| OTU | K1 | K2 | K3 | K4 | K5 | K31 | K49 | K55 | K59 | K70 | Phylum | Class | Order | Family | Genus | Species |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| OTU 212 | 1.8 | 2.0 | 1.8 | 5.9 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Dehalobacteriaceae* | - | - |
| OTU 213 | 1.8 | 6.9 | 1.8 | 1.7 | 3.4 | 1.9 | 2.0 | 1.9 | 1.9 | 6.7 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 214 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 5.3 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Moraxellaceae* | *Acinetobacter* | - |
| OTU 215 | 5.6 | 2.0 | 1.8 | 5.2 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Veillonellaceae* | *Veillonella* | *dispar* |
| OTU 216 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 6.9 | 1.9 | 1.6 | Unassigned | - | - | - | - | - |
| OTU 217 | 5.6 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Bradyrhizobiaceae* | - | - |
| OTU 218 | 1.8 | 2.0 | 1.8 | 1.7 | 5.5 | 1.9 | 4.3 | 1.9 | 4.2 | 1.6 | *Firmicutes* | *Bacilli* | *Bacillales* | - | - | - |
| OTU 219 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 5.3 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Alphaproteobacteria* | *Rhodobacterales* | *Rhodobacteraceae* | *Amaricoccus* | - |
| OTU 220 | 1.8 | 5.9 | 1.8 | 3.2 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 7.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 221 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 6.1 | Unassigned | - | - | - | - | - |
| OTU 222 | 3.3 | 6.2 | 3.3 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | Unassigned | - | - | - | - | - |
| OTU 223 | 6.3 | 2.0 | 1.8 | 1.7 | 5.3 | 1.9 | 2.0 | 4.7 | 1.9 | 1.6 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Streptococcaceae* | *Streptococcus* | - |
| OTU 224 | 1.8 | 3.5 | 1.8 | 1.7 | 1.9 | 5.7 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Alphaproteobacteria* | *Sphingomonadales* | *Sphingomonadaceae* | - | - |
| OTU 225 | 1.8 | 2.0 | 1.8 | 1.7 | 5.0 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Betaproteobacteria* | *Gallionellales* | *Gallionellaceae* | *Gallionella* | - |
| OTU 226 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 6.1 | 1.9 | 1.6 | Unassigned | - | - | - | - | - |
| OTU 227 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 10.1 | 1.9 | 3.1 | Unassigned | - | - | - | - | - |
| OTU 228 | 5.2 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | *Ruminococcus* | - |
| OTU 229 | 1.8 | 5.4 | 1.8 | 4.9 | 1.9 | 5.3 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | - | - |
| OTU 230 | 1.8 | 2.0 | 1.8 | 1.7 | 3.4 | 1.9 | 2.0 | 1.9 | 1.9 | 4.4 | Unassigned | - | - | - | - | - |
| OTU 231 | 1.8 | 2.0 | 5.9 | 1.7 | 5.3 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylocystaceae* | - | - |
| OTU 232 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 4.2 | 1.6 | *Proteobacteria* | *Alphaproteobacteria* | *Rhodospirillales* | *Acetobacteraceae* | - | - |
| OTU 233 | 1.8 | 2.0 | 1.8 | 1.7 | 4.2 | 1.9 | 2.0 | 1.9 | 1.9 | 3.9 | Unassigned | - | - | - | - | - |
| OTU 234 | 1.8 | 2.0 | 1.8 | 1.7 | 5.0 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 235 | 1.8 | 2.0 | 1.8 | 1.7 | 5.5 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 236 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 3.9 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Rikenellaceae* | - | - |
| OTU 237 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 5.1 | 2.0 | 1.9 | 1.9 | 1.6 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 238 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 8.8 | 1.9 | 1.6 | Unassigned | - | - | - | - | - |
| OTU 239 | 1.8 | 4.8 | 4.1 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 4.2 | 1.6 | *Fusobacteria* | *Fusobacteriia* | *Fusobacteriales* | *Fusobacteriaceae* | *Fusobacterium* | - |
| OTU 240 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 5.0 | 10.3 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 241 | 5.4 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Anaerococcus* | - |
| OTU 242 | 3.3 | 2.0 | 5.4 | 5.2 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Comamonadaceae* | - | - |
| OTU 243 | 3.3 | 4.3 | 4.1 | 1.7 | 5.3 | 3.5 | 2.0 | 3.4 | 3.4 | 1.6 | *Fusobacteria* | *Fusobacteriia* | *Fusobacteriales* | *Fusobacteriaceae* | *Cetobacterium* | *somerae* |
| OTU 244 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 6.5 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 245 | 1.8 | 3.5 | 4.1 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 4.8 | *Synergistetes* | *Synergistia* | *Synergistales* | *Synergistaceae* | - | - |
| OTU 246 | 1.8 | 2.0 | 5.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Actinobacteria* | *Coriobacteriia* | *Coriobacteriales* | *Coriobacteriaceae* | - | - |
| OTU 247 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 5.5 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Rikenellaceae* | - | - |
| OTU 248 | 4.6 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 249 | 1.8 | 5.6 | 1.8 | 1.7 | 1.9 | 3.5 | 3.5 | 5.3 | 5.7 | 1.6 | Unassigned | - | - | - | - | - |
| OTU 250 | 1.8 | 2.0 | 4.1 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 6.0 | 7.6 | *Planctomycetes* | *vadinHA49* | *PeHg47* | - | - | - |
| OTU 251 | 1.8 | 2.0 | 1.8 | 7.5 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 252 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 6.7 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Veillonellaceae* | - | - |
| OTU 253 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 4.2 | 1.9 | 1.6 | Unassigned | - | - | - | - | - |
| OTU 254 | 3.3 | 2.0 | 5.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 1.6 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Actinomycetaceae* | *Actinomyces* | - |
| OTU 255 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 1.9 | 5.1 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 256 | 1.8 | 2.0 | 1.8 | 1.7 | 5.3 | 1.9 | 3.5 | 1.9 | 1.9 | 1.6 | *Proteobacteria* | *Alphaproteobacteria* | *Rhodospirillales* | *Rhodospirillaceae* | - | - |
| OTU 257 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 4.7 | 1.9 | 1.6 | Unassigned | - | - | - | - | - |
| OTU 258 | 3.3 | 2.0 | 3.3 | 1.7 | 1.9 | 1.9 | 5.8 | 1.9 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | *Roseburia* | - |
| OTU 259 | 1.8 | 5.6 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 3.4 | 1.9 | 1.6 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 260 | 1.8 | 2.0 | 4.9 | 1.7 | 1.9 | 1.9 | 2.0 | 1.9 | 4.7 | 1.6 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Micrococcaceae* | *Kocuria* | *palustris* |
| OTU 261 | 1.8 | 2.0 | 1.8 | 1.7 | 1.9 | 1.9 | 2.0 | 5.7 | 6.6 | 1.6 | Unassigned | - | - | - | - | - |

^ Classification could not be made to this level

**Appendix 8.** Statistical values of differential abundance comparisons between DESeq2 normalised reads (Love *et al.* 2014) in koalas with (K1 – K5) and without (K31 – K70) wet bottom. Benjamini and Hochberg (1995) (BH) adjusted $P$ values < 0.05 are in bold.

| OTU | Base Mean | LFC[*] | LFC standard error | Wald test Z statistic | BH Adjusted *P* value | Higher abundance group | Present in samples (n/5) | | Taxonomic classification | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | WB absent | WB present | Phylum | Class | Order | Family | Genus | Species |
| OTU 1 | 71520.47 | -1.98 | 0.80 | -2.46 | **0.045** | WB present | 5 | 5 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | -[^] |
| OTU 2 | 61538.06 | 2.79 | 0.84 | 3.31 | **0.006** | WB absent | 5 | 5 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | - |
| OTU 3 | 15469.07 | -4.63 | 1.64 | -2.82 | - | WB present | 5 | 5 | *Proteobacteria* | *Gammaproteobacteria* | *Enterobacteriales* | *Enterobacteriaceae* | - | - |
| OTU 4 | 17147.87 | 3.88 | 1.10 | 3.52 | **0.004** | WB absent | 5 | 5 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | - |
| OTU 5 | 8503.18 | -2.60 | 1.57 | -1.66 | - | WB absent | 5 | 4 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 6 | 10981.84 | 0.52 | 1.16 | 0.45 | 0.729 | WB absent | 5 | 5 | Unassigned | - | - | - | - | - |
| OTU 7 | 7479.20 | 1.70 | 1.23 | 1.39 | 0.236 | WB absent | 5 | 5 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 8 | 3575.92 | -4.32 | 1.53 | -2.81 | - | WB absent | 4 | 5 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 9 | 4133.53 | 2.17 | 1.44 | 1.51 | 0.208 | WB present | 4 | 5 | *Actinobacteria* | *Coriobacteriia* | *Coriobacteriales* | *Coriobacteriaceae* | - | - |
| OTU 10 | 15826.26 | 0.29 | 1.02 | 0.28 | 0.844 | WB absent | 5 | 5 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Facklamia* | - |
| OTU 11 | 3525.98 | -3.29 | 1.63 | -2.02 | - | WB absent | 4 | 4 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 12 | 2503.43 | -2.33 | 1.56 | -1.49 | - | WB absent | 5 | 4 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 13 | 1299.60 | -3.31 | 1.51 | -2.19 | - | WB absent | 5 | 3 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 14 | 1603.98 | 0.18 | 1.44 | 0.13 | 0.911 | WB present | 3 | 5 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 15 | 1406.37 | -2.27 | 1.48 | -1.53 | 0.202 | WB present | 4 | 5 | *TM7* | *TM7-3* | *I025* | *Rs-045* | - | - |
| OTU 16 | 2807.38 | -9.41 | 1.37 | -6.87 | - | WB present | 3 | 5 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Porphyromonas* | - |
| OTU 17 | 1008.79 | -8.43 | 1.53 | -5.51 | - | WB present | 0 | 3 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 18 | 1646.50 | -4.05 | 1.59 | -2.56 | - | WB absent | 4 | 3 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Veillonellaceae* | *Phascolarctobacterium* | - |
| OTU 19 | 25092.18 | 0.80 | 0.86 | 0.92 | 0.433 | WB absent | 5 | 5 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | - |
| OTU 20 | 1481.12 | -2.32 | 1.56 | -1.49 | - | WB absent | 5 | 3 | *Synergistetes* | *Synergistia* | *Synergistales* | *Synergistaceae* | - | - |
| OTU 21 | 503.90 | -7.94 | 1.37 | -5.80 | **< 0.001** | WB present | 1 | 5 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Peptoniphilus* | - |
| OTU 22 | 473.18 | -4.04 | 1.58 | -2.56 | - | WB absent | 3 | 2 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 23 | 518.24 | -4.82 | 1.56 | -3.09 | - | WB present | 2 | 4 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 24 | 645.43 | -4.97 | 1.56 | -3.18 | - | WB present | 3 | 3 | *Proteobacteria* | *Gammaproteobacteria* | *Enterobacteriales* | *Enterobacteriaceae* | - | - |
| OTU 25 | 392.00 | -7.31 | 1.55 | -4.71 | - | WB present | 0 | 2 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 26 | 444.21 | 1.68 | 1.31 | 1.28 | 0.273 | WB absent | 5 | 5 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Propionibacteriaceae* | - | - |
| OTU 27 | 260.40 | -5.10 | 1.53 | -3.34 | - | WB present | 1 | 3 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 28 | 367.85 | -4.19 | 1.51 | -2.77 | - | WB present | 3 | 3 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 29 | 370.15 | -7.55 | 1.38 | -5.49 | - | WB present | 1 | 4 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Gallicola* | - |
| OTU 30 | 280.22 | -6.84 | 1.52 | -4.49 | - | WB present | 1 | 2 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Dysgonomonas* | - |
| OTU 31 | 209.85 | -5.11 | 1.52 | -3.37 | - | WB present | 1 | 2 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 32 | 368.93 | -3.03 | 1.53 | -1.99 | - | WB absent | 5 | 2 | *Fusobacteria* | *Fusobacteriia* | *Fusobacteriales* | *Fusobacteriaceae* | *Fusobacterium* | - |
| OTU 33 | 252.58 | -1.20 | 1.41 | -0.85 | - | WB absent | 5 | 3 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | *Desulfovibrio* | - |
| OTU 34 | 225.96 | -4.14 | 1.46 | -2.83 | - | WB present | 4 | 3 | *Verrucomicrobia* | *Verrucomicrobiae* | *Verrucomicrobiales* | *Verrucomicrobiaceae* | *Akkermansia* | *muciniphila* |
| OTU 35 | 225.72 | -5.58 | 1.53 | -3.65 | - | WB present | 1 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 36 | 453.43 | -1.94 | 1.62 | -1.20 | - | WB absent | 3 | 3 | *Proteobacteria* | *Gammaproteobacteria* | *Pasteurellales* | *Pasteurellaceae* | *Lonepinella* | *koalarum* |
| OTU 37 | 184.58 | 2.00 | 1.21 | 1.65 | 0.165 | WB absent | 5 | 4 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylobacteriaceae* | *Methylobacterium* | - |
| OTU 38 | 214.73 | -7.38 | 1.23 | -6.01 | **< 0.001** | WB present | 0 | 5 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Peptoniphilus* | - |
| OTU 39 | 241.01 | -0.63 | 1.50 | -0.42 | - | WB absent | 5 | 2 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | *Ruminococcus* | - |
| OTU 40 | 257.43 | -6.17 | 1.52 | -4.05 | - | WB present | 2 | 1 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 41 | 285.26 | -1.67 | 1.54 | -1.08 | - | WB absent | 4 | 3 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 42 | 348.71 | -7.24 | 1.54 | -4.71 | - | WB present | 0 | 3 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Peptococcaceae* | *Peptococcus* | - |
| OTU 43 | 216.18 | -5.00 | 1.54 | -3.26 | - | WB present | 1 | 1 | *Planctomycetes* | *vadinHA49* | *PeHg47* | - | - | - |
| OTU 44 | 144.17 | -4.06 | 1.48 | -2.75 | - | WB present | 2 | 2 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 45 | 313.21 | 1.93 | 1.26 | 1.53 | 0.202 | WB absent | 5 | 5 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylobacteriaceae* | *Methylobacterium* | - |
| OTU 46 | 104.25 | -3.26 | 1.42 | -2.30 | - | WB absent | 4 | 2 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 47 | 83.56 | -6.04 | 1.31 | -4.60 | **< 0.001** | WB present | 0 | 3 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *ph2* | - |

| OTU | Base Mean | LFC* | LFC standard error | Wald test Z statistic | BH Adjusted *P* value | Higher abundance group | Present in samples (n/5) | | Taxonomic classification | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | WB absent | WB present | Phylum | Class | Order | Family | Genus | Species |
| OTU 48 | 55.83 | -5.48 | 1.35 | -4.06 | - | WB present | 0 | 3 | *Synergistetes* | *Synergistia* | *Synergistales* | *Synergistaceae* | *vadinCA02* | - |
| OTU 49 | 78.07 | -5.77 | 1.43 | -4.04 | - | WB present | 0 | 2 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 50 | 78.75 | -5.75 | 1.44 | -4.00 | - | WB present | 0 | 1 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Actinomycetaceae* | *Mobiluncus* | - |
| OTU 51 | 48.47 | -5.39 | 1.28 | -4.20 | **< 0.001** | WB present | 0 | 3 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Peptoniphilus* | - |
| OTU 52 | 62.20 | -5.49 | 1.42 | -3.85 | - | WB present | 0 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Veillonellaceae* | *Dialister* | - |
| OTU 53 | 55.98 | -5.37 | 1.42 | -3.79 | - | WB present | 0 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *ph2* | - |
| OTU 54 | 38.95 | -4.99 | 1.38 | -3.62 | - | WB present | 0 | 2 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 55 | 38.30 | -4.95 | 1.39 | -3.56 | - | WB present | 0 | 1 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Porphyromonas* | - |
| OTU 56 | 36.05 | 1.33 | 1.38 | 0.96 | - | WB absent | 2 | 2 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 57 | 24.55 | 2.17 | 1.13 | 1.92 | 0.099 | WB absent | 4 | 3 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylobacteriaceae* | - | - |
| OTU 58 | 32.22 | -4.75 | 1.38 | -3.46 | - | WB present | 0 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 59 | 36.44 | -1.73 | 1.36 | -1.27 | - | WB absent | 2 | 2 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 60 | 43.08 | 5.07 | 1.40 | 3.62 | - | WB absent | 1 | 0 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Streptococcaceae* | *Streptococcus* | - |
| OTU 61 | 22.31 | 3.79 | 1.21 | 3.14 | - | WB absent | 4 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 62 | 29.98 | -0.30 | 1.23 | -0.25 | 0.852 | WB absent | 4 | 2 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Moraxellaceae* | *Acinetobacter* | *rhizosphaerae* |
| OTU 63 | 47.59 | 1.93 | 1.39 | 1.39 | 0.236 | WB absent | 3 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 64 | 29.51 | -4.68 | 1.35 | -3.48 | - | WB present | 0 | 2 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 65 | 67.82 | -5.47 | 1.36 | -4.02 | **0.001** | WB present | 1 | 2 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Alcaligenaceae* | *Sutterella* | - |
| OTU 66 | 19.64 | -4.19 | 1.32 | -3.18 | - | WB present | 0 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *ph2* | - |
| OTU 67 | 39.01 | -4.97 | 1.39 | -3.57 | - | WB present | 0 | 1 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Prevotellaceae* | *Prevotella* | - |
| OTU 69 | 18.56 | 4.18 | 1.24 | 3.37 | **0.005** | WB absent | 2 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 70 | 15.48 | -3.91 | 1.28 | -3.05 | - | WB present | 0 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | *Oscillospira* | - |
| OTU 71 | 21.40 | 4.34 | 1.27 | 3.42 | **0.005** | WB absent | 2 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 72 | 20.34 | -4.23 | 1.32 | -3.21 | - | WB present | 0 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 73 | 21.03 | -4.33 | 1.27 | -3.41 | **0.005** | WB present | 0 | 2 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Rikenellaceae* | - | - |
| OTU 74 | 12.93 | 3.67 | 1.26 | 2.91 | - | WB absent | 1 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 75 | 27.32 | 4.62 | 1.30 | 3.56 | **0.004** | WB absent | 2 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 76 | 19.92 | 1.34 | 1.19 | 1.12 | 0.335 | WB absent | 4 | 3 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Pseudomonadaceae* | *Pseudomonas* | - |
| OTU 77 | 23.15 | 0.19 | 1.11 | 0.17 | 0.886 | WB absent | 5 | 3 | *Proteobacteria* | *Alphaproteobacteria* | *Sphingomonadales* | *Sphingomonadaceae* | *Sphingomonas* | *yabuuchiae* |
| OTU 78 | 26.20 | 3.76 | 1.18 | 3.17 | - | WB absent | 5 | 4 | *Firmicutes* | *Bacilli* | *Bacillales* | *Staphylococcaceae* | *Staphylococcus* | - |
| OTU 79 | 13.90 | -2.74 | 1.22 | -2.24 | - | WB present | 1 | 2 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 80 | 11.75 | 3.55 | 1.25 | 2.83 | - | WB absent | 1 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Hyphomicrobiaceae* | *Pedomicrobium* | - |
| OTU 81 | 14.51 | -3.84 | 1.27 | -3.01 | - | WB present | 0 | 1 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Oxalobacteraceae* | - | - |
| OTU 82 | 12.93 | -3.69 | 1.26 | -2.93 | - | WB present | 0 | 1 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 83 | 13.59 | -3.52 | 1.24 | -2.84 | - | WB present | 1 | 2 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | *fragilis* |
| OTU 84 | 12.10 | 3.59 | 1.26 | 2.85 | - | WB absent | 1 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 86 | 20.70 | 4.37 | 1.18 | 3.71 | **0.003** | WB absent | 3 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 87 | 10.18 | 3.36 | 1.24 | 2.72 | - | WB absent | 1 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 88 | 11.83 | 3.55 | 1.26 | 2.82 | - | WB absent | 1 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 89 | 7.42 | 2.95 | 1.19 | 2.48 | - | WB absent | 1 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 90 | 24.08 | -3.05 | 1.27 | -2.40 | **0.048** | WB present | 1 | 2 | *Actinobacteria* | *Coriobacteriia* | *Coriobacteriales* | *Coriobacteriaceae* | - | - |
| OTU 91 | 19.98 | -4.21 | 1.32 | -3.20 | - | WB present | 0 | 1 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfarculales* | *Desulfarculaceae* | - | - |
| OTU 92 | 12.44 | 1.48 | 1.13 | 1.31 | 0.265 | WB absent | 3 | 2 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Comamonadaceae* | *Comamonas* | - |
| OTU 93 | 19.03 | -4.17 | 1.30 | -3.22 | - | WB present | 0 | 2 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 94 | 20.71 | 4.06 | 1.26 | 3.22 | **0.007** | WB absent | 2 | 1 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylocystaceae* | - | - |
| OTU 95 | 8.91 | 3.23 | 1.16 | 2.79 | **0.022** | WB absent | 2 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Rhizobiaceae* | *Rhizobium* | *leguminosarum* |
| OTU 96 | 33.02 | 4.56 | 1.36 | 3.36 | - | WB absent | 2 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 97 | 12.21 | 1.53 | 1.12 | 1.37 | - | WB absent | 5 | 2 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Comamonadaceae* | - | - |
| OTU 98 | 18.54 | -4.14 | 1.29 | -3.20 | - | WB present | 0 | 2 | *Planctomycetes* | *vadinHA49* | *PeHg47* | - | - | - |
| OTU 99 | 11.43 | -3.53 | 1.25 | -2.82 | - | WB present | 0 | 1 | *Actinobacteria* | *Coriobacteriia* | *Coriobacteriales* | *Coriobacteriaceae* | - | - |
| OTU 100 | 10.44 | -3.41 | 1.24 | -2.75 | - | WB present | 0 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Mogibacteriaceae]* | *Mogibacterium* | - |

| OTU | Base Mean | LFC* | LFC standard error | Wald test Z statistic | BH Adjusted *P* value | Higher abundance group | Present in samples (n/5) | | Taxonomic classification | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | WB absent | WB present | Phylum | Class | Order | Family | Genus | Species |
| OTU 101 | 8.05 | 0.05 | 1.02 | 0.05 | 0.962 | WB absent | 3 | 3 | *Bacteroidetes* | *[Saprospirae]* | *[Saprospirales]* | *Chitinophagaceae* | *Sediminibacterium* | - |
| OTU 102 | 9.07 | 2.56 | 1.15 | 2.24 | - | WB absent | 3 | 2 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | *Blautia* | *producta* |
| OTU 103 | 9.30 | 3.29 | 1.17 | 2.81 | **0.022** | WB absent | 2 | 0 | Unassigned | - | - | - | - | - |
| OTU 105 | 159.80 | -2.77 | 1.13 | -2.46 | **0.045** | WB present | 4 | 5 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Aerococcaceae* | *Aerococcus* | - |
| OTU 106 | 9.01 | 3.25 | 1.14 | 2.85 | **0.022** | WB absent | 3 | 0 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Burkholderiaceae* | *Burkholderia* | - |
| OTU 107 | 9.65 | 3.30 | 1.23 | 2.67 | - | WB absent | 1 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Sphingomonadales* | *Sphingomonadaceae* | *Sphingobium* | - |
| OTU 108 | 6.29 | 2.01 | 1.12 | 1.79 | - | WB absent | 1 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 109 | 8.93 | -3.24 | 1.16 | -2.79 | **0.022** | WB present | 0 | 2 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Peptostreptococcaceae* | *Peptostreptococcus* | - |
| OTU 110 | 4.67 | -2.35 | 1.08 | -2.17 | - | WB present | 0 | 1 | *Acidobacteria* | *Acidobacteriia* | *Acidobacteriales* | *Acidobacteriaceae* | - | - |
| OTU 111 | 6.63 | -2.11 | 1.10 | -1.93 | - | WB present | 1 | 3 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 112 | 5.59 | 2.59 | 1.10 | 2.35 | - | WB absent | 2 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Nocardiaceae* | *Rhodococcus* | *fascians* |
| OTU 113 | 5.19 | 2.51 | 1.00 | 2.50 | **0.044** | WB absent | 3 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Rhizobiaceae* | *Agrobacterium* | - |
| OTU 114 | 8.21 | 2.86 | 1.14 | 2.52 | **0.043** | WB absent | 2 | 1 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Oxalobacteraceae* | - | - |
| OTU 115 | 7.96 | -1.42 | 1.14 | -1.24 | - | WB absent | 2 | 1 | *Cyanobacteria* | *4C0d-2* | *YS2* | - | - | - |
| OTU 116 | 6.13 | -2.73 | 1.15 | -2.37 | - | WB present | 0 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 117 | 5.28 | -2.52 | 1.11 | -2.27 | - | WB present | 0 | 1 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfobacterales* | *Desulfobacteraceae* | - | - |
| OTU 118 | 5.77 | -2.64 | 1.13 | -2.33 | - | WB present | 0 | 1 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 119 | 5.40 | 1.20 | 1.04 | 1.15 | 0.331 | WB absent | 2 | 1 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Oxalobacteraceae* | *Ralstonia* | - |
| OTU 120 | 7.55 | 2.97 | 1.19 | 2.49 | - | WB absent | 1 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *[Paraprevotellaceae]* | *Paraprevotella* | - |
| OTU 121 | 6.24 | 2.72 | 1.15 | 2.36 | - | WB absent | 1 | 0 | *Firmicutes* | *Erysipelotrichi* | *Erysipelotrichales* | *Erysipelotrichaceae* | *Coprobacillus* | - |
| OTU 122 | 9.13 | 3.22 | 1.22 | 2.63 | - | WB absent | 1 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 123 | 18.77 | -2.02 | 1.14 | -1.76 | 0.134 | WB present | 2 | 5 | Unassigned | - | - | - | - | - |
| OTU 124 | 5.82 | 2.63 | 1.14 | 2.30 | - | WB absent | 1 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Micrococcaceae* | *Rothia* | *mucilaginosa* |
| OTU 125 | 8.62 | 3.15 | 1.22 | 2.59 | - | WB absent | 1 | 0 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfarculales* | *Desulfarculaceae* | - | - |
| OTU 126 | 7.58 | 2.73 | 1.16 | 2.35 | - | WB absent | 2 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 127 | 4.72 | -2.38 | 1.06 | -2.25 | - | WB present | 0 | 2 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 128 | 7.83 | -3.05 | 1.20 | -2.55 | - | WB present | 0 | 1 | *Planctomycetes* | *vadinHA49* | *PeHg47* | - | - | - |
| OTU 129 | 3.27 | 1.77 | 0.98 | 1.80 | - | WB absent | 1 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Micrococcaceae* | - | - |
| OTU 130 | 5.09 | 2.16 | 1.07 | 2.02 | - | WB absent | 2 | 1 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 131 | 6.28 | 2.02 | 1.04 | 1.94 | 0.097 | WB absent | 2 | 2 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 132 | 8.90 | 0.65 | 1.09 | 0.59 | 0.630 | WB absent | 3 | 3 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Comamonadaceae* | - | - |
| OTU 133 | 4.47 | 2.25 | 1.07 | 2.10 | - | WB absent | 1 | 0 | *Bacteroidetes* | *Cytophagia* | *Cytophagales* | *Cytophagaceae* | *Hymenobacter* | - |
| OTU 134 | 4.67 | 2.32 | 1.08 | 2.14 | - | WB absent | 1 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Eubacteriaceae* | *Anaerofustis* | - |
| OTU 135 | 3.94 | -2.07 | 1.04 | -2.00 | 0.092 | WB present | 0 | 1 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | *uniformis* |
| OTU 136 | 4.45 | 1.46 | 1.02 | 1.42 | 0.232 | WB absent | 1 | 1 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 137 | 8.75 | 1.01 | 1.08 | 0.93 | 0.433 | WB absent | 3 | 1 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 138 | 3.93 | 2.03 | 1.04 | 1.96 | 0.096 | WB absent | 1 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | *Oscillospira* | - |
| OTU 139 | 3.75 | 1.99 | 1.02 | 1.95 | - | WB absent | 1 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Micrococcaceae* | *Rothia* | *dentocariosa* |
| OTU 140 | 4.85 | -0.30 | 1.00 | -0.30 | 0.842 | WB present | 1 | 2 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Moraxellaceae* | *Acinetobacter* | - |
| OTU 141 | 4.87 | 1.85 | 1.02 | 1.81 | 0.124 | WB absent | 2 | 1 | *Actinobacteria* | *Thermoleophilia* | *Solirubrobacterales* | - | - | - |
| OTU 142 | 5.61 | -2.58 | 1.13 | -2.28 | - | WB present | 0 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 143 | 3.10 | 1.69 | 0.97 | 1.74 | - | WB absent | 1 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | *ovatus* |
| OTU 144 | 5.36 | 2.52 | 1.08 | 2.35 | 0.054 | WB absent | 2 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Corynebacteriaceae* | *Corynebacterium* | - |
| OTU 145 | 3.46 | -1.90 | 1.00 | -1.91 | - | WB present | 0 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | *Clostridium* | - |
| OTU 146 | 5.01 | 2.44 | 1.05 | 2.33 | 0.054 | WB absent | 2 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Prevotellaceae* | *Prevotella* | - |
| OTU 147 | 4.19 | -2.19 | 1.05 | -2.08 | - | WB present | 0 | 1 | Unassigned | - | - | - | - | - |
| OTU 148 | 8.75 | -3.22 | 1.15 | -2.79 | **0.022** | WB present | 0 | 2 | Unassigned | - | - | - | - | - |
| OTU 149 | 3.43 | 0.04 | 0.92 | 0.05 | - | WB present | 1 | 1 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | *Desulfovibrio* | - |
| OTU 150 | 5.94 | 2.65 | 1.14 | 2.32 | - | WB absent | 1 | 0 | *Bacteroidetes* | *Cytophagia* | *Cytophagales* | *Cytophagaceae* | *Hymenobacter* | - |
| OTU 151 | 4.90 | 2.40 | 1.06 | 2.26 | 0.061 | WB absent | 2 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Kineosporiaceae* | - | - |
| OTU 152 | 5.16 | -2.49 | 1.11 | -2.25 | - | WB present | 0 | 1 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |

| OTU | Base Mean | LFC* | LFC standard error | Wald test Z statistic | BH Adjusted $P$ value | Higher abundance group | Present in samples (n/5) | | Taxonomic classification | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | WB absent | WB present | Phylum | Class | Order | Family | Genus | Species |
| OTU 153 | 4.01 | -1.11 | 0.97 | -1.15 | 0.331 | WB present | 2 | 1 | Proteobacteria | Deltaproteobacteria | Desulfovibrionales | Desulfovibrionaceae | Bilophila | - |
| OTU 154 | 3.27 | 1.77 | 0.98 | 1.80 | - | WB absent | 1 | 0 | Proteobacteria | Deltaproteobacteria | Desulfovibrionales | Desulfovibrionaceae | - | - |
| OTU 155 | 104.48 | -1.49 | 1.32 | -1.13 | 0.335 | WB absent | 4 | 3 | Firmicutes | Bacilli | Lactobacillales | Carnobacteriaceae | Trichococcus | - |
| OTU 156 | 3.79 | 2.01 | 0.96 | 2.09 | - | WB absent | 3 | 0 | Armatimonadetes | [Fimbriimonadia] | [Fimbriimonadales] | [Fimbriimonadaceae] | Fimbriimonas | - |
| OTU 157 | 3.80 | -1.00 | 0.88 | -1.14 | - | WB present | 1 | 4 | Firmicutes | Clostridia | Clostridiales | Clostridiaceae | Clostridium | - |
| OTU 158 | 2.97 | 1.59 | 0.96 | 1.66 | - | WB absent | 1 | 0 | Proteobacteria | Alphaproteobacteria | Rhizobiales | Methylocystaceae | - | - |
| OTU 159 | 11.73 | -2.84 | 1.08 | -2.62 | **0.035** | WB present | 2 | 4 | Unassigned | - | - | - | - | - |
| OTU 160 | 3.12 | -1.41 | 0.94 | -1.50 | - | WB present | 1 | 1 | Firmicutes | Clostridia | Clostridiales | - | - | - |
| OTU 161 | 4.54 | 1.49 | 1.00 | 1.48 | 0.214 | WB absent | 2 | 2 | Firmicutes | Clostridia | Clostridiales | Clostridiaceae | Clostridium | perfringens |
| OTU 162 | 3.11 | 1.70 | 0.95 | 1.78 | - | WB absent | 2 | 0 | Firmicutes | Clostridia | Clostridiales | Ruminococcaceae | - | - |
| OTU 163 | 3.84 | 2.04 | 0.99 | 2.06 | 0.081 | WB absent | 2 | 0 | Proteobacteria | Gammaproteobacteria | Oceanospirillales | Halomonadaceae | - | - |
| OTU 164 | 3.36 | -1.86 | 0.97 | -1.92 | - | WB present | 0 | 2 | Proteobacteria | Gammaproteobacteria | Pseudomonadales | Moraxellaceae | Acinetobacter | - |
| OTU 165 | 22.27 | -0.89 | 1.07 | -0.83 | 0.486 | WB present | 4 | 4 | Unassigned | - | - | - | - | - |
| OTU 166 | 4.50 | -2.27 | 1.08 | -2.11 | - | WB present | 0 | 1 | Firmicutes | Clostridia | Clostridiales | Lachnospiraceae | - | - |
| OTU 167 | 4.12 | -2.18 | 1.02 | -2.14 | 0.071 | WB present | 0 | 2 | Firmicutes | Clostridia | Clostridiales | Lachnospiraceae | Dorea | - |
| OTU 168 | 3.10 | 1.69 | 0.97 | 1.74 | - | WB absent | 1 | 0 | Firmicutes | Clostridia | Clostridiales | Ruminococcaceae | Oscillospira | - |
| OTU 169 | 3.51 | -1.89 | 1.01 | -1.88 | - | WB present | 0 | 1 | Actinobacteria | Actinobacteria | Actinomycetales | Intrasporangiaceae | Terracoccus | - |
| OTU 170 | 4.40 | 2.23 | 1.02 | 2.18 | 0.068 | WB absent | 3 | 0 | Firmicutes | Bacilli | Lactobacillales | Enterococcaceae | - | - |
| OTU 171 | 2.71 | 1.46 | 0.91 | 1.62 | - | WB absent | 2 | 0 | Firmicutes | Clostridia | Clostridiales | Lachnospiraceae | Blautia | - |
| OTU 172 | 5.41 | 2.03 | 0.94 | 2.16 | 0.070 | WB absent | 4 | 1 | Proteobacteria | Betaproteobacteria | Burkholderiales | Oxalobacteraceae | Ralstonia | - |
| OTU 173 | 2.73 | -1.52 | 0.93 | -1.63 | - | WB present | 0 | 1 | Deferribacteres | Deferribacteres | Deferribacterales | Deferribacteraceae | Mucispirillum | - |
| OTU 174 | 6.05 | -0.70 | 1.07 | -0.66 | 0.592 | WB present | 1 | 2 | Unassigned | - | - | - | - | - |
| OTU 175 | 4.55 | 2.30 | 1.03 | 2.23 | 0.063 | WB absent | 2 | 0 | Proteobacteria | Betaproteobacteria | Burkholderiales | Oxalobacteraceae | Oxalobacter | formigenes |
| OTU 176 | 2.49 | -1.36 | 0.91 | -1.50 | - | WB present | 0 | 1 | Firmicutes | Clostridia | Clostridiales | Lachnospiraceae | - | - |
| OTU 177 | 4.67 | -2.35 | 1.08 | -2.17 | - | WB present | 0 | 1 | Proteobacteria | Deltaproteobacteria | Desulfovibrionales | Desulfovibrionaceae | Bilophila | - |
| OTU 178 | 3.36 | -1.54 | 0.96 | -1.60 | - | WB present | 1 | 1 | Unassigned | - | - | - | - | - |
| OTU 179 | 2.33 | 1.18 | 0.88 | 1.33 | - | WB absent | 2 | 0 | Actinobacteria | Actinobacteria | Actinomycetales | Corynebacteriaceae | Corynebacterium | - |
| OTU 180 | 2.33 | 1.17 | 0.90 | 1.31 | - | WB absent | 1 | 0 | Proteobacteria | Alphaproteobacteria | Rhodospirillales | Rhodospirillaceae | - | - |
| OTU 181 | 7.22 | -2.94 | 1.18 | -2.49 | - | WB present | 0 | 1 | Bacteroidetes | Bacteroidia | Bacteroidales | Porphyromonadaceae | Parabacteroides | - |
| OTU 182 | 2.33 | 1.17 | 0.90 | 1.31 | - | WB absent | 1 | 0 | Cyanobacteria | 4C0d-2 | MLE1-12 | - | - | - |
| OTU 183 | 10.62 | -2.20 | 1.14 | -1.93 | - | WB present | 3 | 3 | Proteobacteria | Gammaproteobacteria | Pseudomonadales | Pseudomonadaceae | Pseudomonas | - |
| OTU 184 | 4.09 | 1.00 | 0.93 | 1.06 | 0.360 | WB absent | 2 | 1 | Proteobacteria | Betaproteobacteria | Burkholderiales | Comamonadaceae | - | - |
| OTU 185 | 2.20 | 1.06 | 0.88 | 1.20 | - | WB absent | 1 | 0 | Proteobacteria | Alphaproteobacteria | Rhodospirillales | Acetobacteraceae | - | - |
| OTU 186 | 2.24 | -1.10 | 0.89 | -1.24 | - | WB present | 0 | 1 | Firmicutes | Clostridia | Clostridiales | [Tissierellaceae] | Peptoniphilus | - |
| OTU 187 | 3.62 | 0.17 | 0.90 | 0.18 | - | WB absent | 2 | 1 | Proteobacteria | Gammaproteobacteria | Pseudomonadales | Moraxellaceae | Enhydrobacter | - |
| OTU 188 | 2.20 | 1.06 | 0.88 | 1.20 | - | WB absent | 1 | 0 | Proteobacteria | Betaproteobacteria | Neisseriales | Neisseriaceae | Neisseria | subflava |
| OTU 189 | 2.11 | 0.95 | 0.88 | 1.07 | - | WB absent | 1 | 0 | Proteobacteria | Alphaproteobacteria | Rhodospirillales | Acetobacteraceae | - | - |
| OTU 190 | 2.25 | 1.08 | 0.89 | 1.21 | - | WB absent | 1 | 0 | Actinobacteria | Actinobacteria | Actinomycetales | Nocardioidaceae | - | - |
| OTU 191 | 6.13 | -2.73 | 1.15 | -2.37 | - | WB present | 0 | 1 | Planctomycetes | vadinHA49 | PeHg47 | - | - | - |
| OTU 192 | 2.73 | -1.52 | 0.93 | -1.63 | - | WB present | 0 | 1 | Bacteroidetes | Bacteroidia | Bacteroidales | Bacteroidaceae | Bacteroides | ovatus |
| OTU 193 | 4.30 | 2.22 | 1.00 | 2.21 | 0.065 | WB absent | 2 | 0 | Firmicutes | Clostridia | Clostridiales | Lachnospiraceae | - | - |
| OTU 194 | 4.43 | -2.27 | 1.07 | -2.13 | - | WB present | 0 | 1 | Bacteroidetes | Bacteroidia | Bacteroidales | Rikenellaceae | - | - |
| OTU 195 | 3.81 | 2.01 | 1.03 | 1.95 | 0.096 | WB absent | 1 | 0 | Firmicutes | Clostridia | Clostridiales | Ruminococcaceae | - | - |
| OTU 196 | 4.23 | -0.22 | 0.83 | -0.26 | 0.852 | WB present | 4 | 5 | Proteobacteria | Gammaproteobacteria | Enterobacteriales | Enterobacteriaceae | Plesiomonas | shigelloides |
| OTU 197 | 1.97 | 0.80 | 0.88 | 0.91 | - | WB absent | 1 | 0 | Acidobacteria | Acidobacteriia | Acidobacteriales | Acidobacteriaceae | Terriglobus | - |
| OTU 199 | 2.60 | 1.38 | 0.92 | 1.49 | - | WB absent | 1 | 0 | Actinobacteria | Actinobacteria | Actinomycetales | Corynebacteriaceae | Corynebacterium | - |
| OTU 200 | 14.70 | -1.45 | 1.22 | -1.19 | - | WB present | 2 | 2 | Unassigned | - | - | - | - | - |
| OTU 201 | 12.56 | -3.65 | 1.26 | -2.89 | - | WB present | 0 | 1 | Firmicutes | Bacilli | Lactobacillales | Aerococcaceae | Aerococcus | - |
| OTU 202 | 2.85 | -1.59 | 0.94 | -1.69 | - | WB present | 0 | 1 | Firmicutes | Clostridia | Clostridiales | Lachnospiraceae | Coprococcus | - |
| OTU 204 | 3.33 | 0.45 | 0.83 | 0.54 | - | WB absent | 3 | 3 | Fusobacteria | Fusobacteriia | Fusobacteriales | Leptotrichiaceae | - | - |
| OTU 205 | 2.05 | -0.17 | 0.84 | -0.20 | - | WB present | 1 | 1 | Firmicutes | Clostridia | Clostridiales | Lachnospiraceae | - | - |

| OTU | Base Mean | LFC* | LFC standard error | Wald test Z statistic | BH Adjusted *P* value | Higher abundance group | Present in samples (n/5) | | Taxonomic classification | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | WB absent | WB present | Phylum | Class | Order | Family | Genus | Species |
| OTU 206 | 3.34 | -1.85 | 0.99 | -1.87 | - | WB present | 0 | 1 | *Proteobacteria* | *Gammaproteobacteria* | *Pasteurellales* | *Pasteurellaceae* | *Aggregatibacter* | - |
| OTU 207 | 2.12 | -1.06 | 0.88 | -1.20 | - | WB present | 0 | 1 | Unassigned | - | - | - | - | - |
| OTU 209 | 3.24 | 0.34 | 0.81 | 0.42 | - | WB absent | 4 | 4 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Peptostreptococcaceae* | - | - |
| OTU 210 | 2.36 | -1.27 | 0.90 | -1.41 | - | WB present | 0 | 1 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Rikenellaceae* | - | - |
| OTU 212 | 2.57 | 1.37 | 0.92 | 1.49 | - | WB absent | 1 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Dehalobacteriaceae* | - | - |
| OTU 213 | 6.68 | 0.24 | 1.12 | 0.21 | 0.867 | WB absent | 2 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 214 | 2.12 | -1.06 | 0.88 | -1.20 | - | WB present | 0 | 1 | *Proteobacteria* | *Gammaproteobacteria* | *Pseudomonadales* | *Moraxellaceae* | *Acinetobacter* | - |
| OTU 215 | 2.99 | 1.63 | 0.92 | 1.78 | - | WB absent | 2 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Veillonellaceae* | *Veillonella* | *dispar* |
| OTU 216 | 4.22 | -2.18 | 1.06 | -2.06 | 0.081 | WB present | 0 | 1 | Unassigned | - | - | - | - | - |
| OTU 217 | 2.33 | 1.17 | 0.90 | 1.31 | - | WB absent | 1 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Bradyrhizobiaceae* | - | - |
| OTU 218 | 2.84 | 0.26 | 0.86 | 0.31 | - | WB present | 1 | 2 | *Firmicutes* | *Bacilli* | *Bacillales* | - | - | - |
| OTU 219 | 2.12 | -0.98 | 0.88 | -1.11 | - | WB present | 0 | 1 | *Proteobacteria* | *Alphaproteobacteria* | *Rhodobacterales* | *Rhodobacteraceae* | *Amaricoccus* | - |
| OTU 220 | 8.18 | -1.49 | 1.16 | -1.28 | - | WB present | 2 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 221 | 2.85 | -1.59 | 0.94 | -1.69 | - | WB present | 0 | 1 | Unassigned | - | - | - | - | - |
| OTU 222 | 3.15 | 1.69 | 0.94 | 1.79 | - | WB absent | 3 | 0 | Unassigned | - | - | - | - | - |
| OTU 223 | 4.14 | 1.35 | 0.95 | 1.42 | 0.232 | WB absent | 2 | 1 | *Firmicutes* | *Bacilli* | *Lactobacillales* | *Streptococcaceae* | *Streptococcus* | - |
| OTU 224 | 2.56 | -0.98 | 0.89 | -1.10 | - | WB present | 1 | 1 | *Proteobacteria* | *Alphaproteobacteria* | *Sphingomonadales* | *Sphingomonadaceae* | - | - |
| OTU 225 | 1.97 | 0.80 | 0.88 | 0.91 | - | WB absent | 1 | 0 | *Proteobacteria* | *Betaproteobacteria* | *Gallionellales* | *Gallionellaceae* | *Gallionella* | - |
| OTU 226 | 2.81 | -1.52 | 0.94 | -1.61 | - | WB present | 0 | 1 | Unassigned | - | - | - | - | - |
| OTU 227 | 71.09 | -5.66 | 1.42 | -3.98 | - | WB present | 0 | 2 | Unassigned | - | - | - | - | - |
| OTU 228 | 2.06 | 0.93 | 0.88 | 1.06 | - | WB absent | 1 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | *Ruminococcus* | - |
| OTU 229 | 3.39 | 0.44 | 0.89 | 0.50 | - | WB absent | 2 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Clostridiaceae* | - | - |
| OTU 230 | 1.90 | -0.39 | 0.85 | -0.46 | - | WB present | 1 | 1 | Unassigned | - | - | - | - | - |
| OTU 231 | 3.31 | 1.80 | 0.94 | 1.91 | - | WB absent | 2 | 0 | *Proteobacteria* | *Alphaproteobacteria* | *Rhizobiales* | *Methylocystaceae* | - | - |
| OTU 232 | 1.67 | -0.48 | 0.88 | -0.54 | - | WB present | 0 | 1 | *Proteobacteria* | *Alphaproteobacteria* | *Rhodospirillales* | *Acetobacteraceae* | - | - |
| OTU 233 | 1.92 | 0.00 | 0.84 | 0.00 | - | WB present | 1 | 1 | Unassigned | - | - | - | - | - |
| OTU 234 | 1.97 | 0.80 | 0.88 | 0.91 | - | WB absent | 1 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 235 | 2.25 | 1.08 | 0.89 | 1.21 | - | WB absent | 1 | 0 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 236 | 1.64 | -0.47 | 0.88 | -0.53 | - | WB present | 0 | 1 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Rikenellaceae* | - | - |
| OTU 237 | 1.97 | -0.83 | 0.88 | -0.95 | - | WB present | 0 | 1 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Bacteroidaceae* | *Bacteroides* | - |
| OTU 238 | 18.50 | -4.12 | 1.31 | -3.15 | - | WB present | 0 | 1 | Unassigned | - | - | - | - | - |
| OTU 239 | 2.39 | 0.48 | 0.84 | 0.58 | - | WB absent | 2 | 1 | *Fusobacteria* | *Fusobacteriia* | *Fusobacteriales* | *Fusobacteriaceae* | *Fusobacterium* | - |
| OTU 240 | 93.07 | -5.98 | 1.43 | -4.19 | - | WB present | 0 | 2 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Porphyromonadaceae* | *Parabacteroides* | - |
| OTU 241 | 2.20 | 1.06 | 0.88 | 1.20 | - | WB absent | 1 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *[Tissierellaceae]* | *Anaerococcus* | - |
| OTU 242 | 2.99 | 1.64 | 0.90 | 1.82 | - | WB absent | 3 | 0 | *Proteobacteria* | *Betaproteobacteria* | *Burkholderiales* | *Comamonadaceae* | - | - |
| OTU 243 | 3.24 | 0.86 | 0.83 | 1.03 | - | WB absent | 4 | 3 | *Fusobacteria* | *Fusobacteriia* | *Fusobacteriales* | *Fusobacteriaceae* | *Cetobacterium* | *somerae* |
| OTU 244 | 3.46 | -1.90 | 1.00 | -1.91 | - | WB present | 0 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 245 | 2.29 | -0.15 | 0.83 | -0.18 | - | WB absent | 2 | 1 | *Synergistetes* | *Synergistia* | *Synergistales* | *Synergistaceae* | - | - |
| OTU 246 | 2.46 | 1.28 | 0.91 | 1.41 | - | WB absent | 1 | 0 | *Actinobacteria* | *Coriobacteriia* | *Coriobacteriales* | *Coriobacteriaceae* | - | - |
| OTU 247 | 2.24 | -1.17 | 0.89 | -1.32 | - | WB present | 0 | 1 | *Bacteroidetes* | *Bacteroidia* | *Bacteroidales* | *Rikenellaceae* | - | - |
| OTU 248 | 1.79 | 0.63 | 0.87 | 0.72 | - | WB absent | 1 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | - | - |
| OTU 249 | 4.28 | -0.74 | 0.91 | -0.81 | 0.491 | WB present | 1 | 4 | Unassigned | - | - | - | - | - |
| OTU 250 | 8.64 | -2.75 | 1.13 | -2.42 | **0.048** | WB present | 1 | 2 | *Planctomycetes* | *vadinHA49* | *PeHg47* | - | - | - |
| OTU 251 | 6.64 | 2.81 | 1.17 | 2.40 | - | WB absent | 1 | 0 | *Firmicutes* | *Clostridia* | *Clostridiales* | - | - | - |
| OTU 252 | 3.70 | -2.01 | 1.02 | -1.97 | - | WB present | 0 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Veillonellaceae* | - | - |
| OTU 253 | 1.67 | -0.48 | 0.88 | -0.54 | - | WB present | 0 | 1 | Unassigned | - | - | - | - | - |
| OTU 254 | 2.60 | 1.38 | 0.91 | 1.52 | - | WB absent | 2 | 0 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Actinomycetaceae* | *Actinomyces* | - |
| OTU 255 | 2.00 | -0.93 | 0.87 | -1.07 | - | WB present | 0 | 1 | *Proteobacteria* | *Deltaproteobacteria* | *Desulfovibrionales* | *Desulfovibrionaceae* | - | - |
| OTU 256 | 2.26 | 0.69 | 0.86 | 0.80 | - | WB absent | 1 | 1 | *Proteobacteria* | *Alphaproteobacteria* | *Rhodospirillales* | *Rhodospirillaceae* | - | - |
| OTU 257 | 1.82 | -0.66 | 0.87 | -0.76 | - | WB present | 0 | 1 | Unassigned | - | - | - | - | - |
| OTU 258 | 2.71 | -0.79 | 0.88 | -0.90 | - | WB present | 2 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Lachnospiraceae* | *Roseburia* | - |

| OTU | Base Mean | LFC* | LFC standard error | Wald test Z statistic | BH Adjusted *P* value | Higher abundance group | Present in samples (n/5) | | Taxonomic classification | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | WB absent | WB present | Phylum | Class | Order | Family | Genus | Species |
| OTU 259 | 2.43 | 0.84 | 0.88 | 0.96 | - | WB absent | 1 | 1 | *Firmicutes* | *Clostridia* | *Clostridiales* | *Ruminococcaceae* | - | - |
| OTU 260 | 2.35 | 0.14 | 0.84 | 0.16 | - | WB absent | 1 | 1 | *Actinobacteria* | *Actinobacteria* | *Actinomycetales* | *Micrococcaceae* | *Kocuria* | *palustris* |
| OTU 261 | 4.64 | -2.34 | 1.03 | -2.27 | 0.061 | WB present | 0 | 2 | Unassigned | - | - | - | - | - |

^ Classification could not be made to this level

* $\log_2$ fold change

**Appendix 9.** An example Bayes block, used to build all MrBayes (Huelsenbeck & Ronquist 2001) trees. Comments are between square brackets. Variables are in capital letters.

```
begin mrbayes;
	log start replace filename = DIR/FILE.mrbayes.coding.log;
	set autoclose=yes nowarn=yes Seed=123456789 Swapseed=987654321;
	lset Nst=NST; [Replace NST with model value. JC/F81=1; K80/K2P/HKY=2;
SYM/GTR=6]
	lset rates=RATES; [Equal/Gamma (for +G)/Propinv (for +I)/Invgamma (for
+I+G)/Adgamma]
	lset ngammacat=4;
	prset statefreqpr=STATEFREQPR; [fixed(equal) for JC/SYM model, else
Dirichlet(1.0,1.0,1.0,1.0)]
	mcmcp Ngen=1000000; [Chain length of 1,000,000]
	mcmcp Nruns=2;
	mcmcp Nchains=4; [Use 4 chains per run, 1 cold and 3 heated]
	mcmcp Temp=0.200000;
	mcmcp Swapfreq=1;
	mcmcp Nswaps=1;
	mcmcp Samplefreq=1000;
	mcmcp Printfreq=10000; [How often the data is printed to the output]
	mcmcp Printall=No;
	mcmcp Printmax=8;
	mcmcp Mcmcdiagn=Yes;
	mcmcp Diagnfreq=1000;
	mcmcp Diagnstat=Avgstddev;
	mcmcp Minpartfreq=0.20;
	mcmcp Allchains=No;
	mcmcp Allcomps=No;
	mcmcp Relburnin=Yes; [Use this if you want burn in to be relative to the chain
length]
	mcmcp Burninfrac=0.25; [Set to proportion of chain length to use as burn in]
	mcmcp Stoprule=No
	mcmcp Stopval=0.01;
	mcmcp Checkpoint=Yes;
	mcmcp Checkfreq=10000;
	mcmcp Filename=DIR/FILE; [Replace with output filename]
	mcmc; [Run MrBayes]
	sump filename=DIR/FILE Relburnin=Yes Burninfrac=0.25;
	sumt filename=DIR/FILE contype=halfcompat Relburnin=Yes Burninfrac=0.25;
	log stop;
end;
```

**Appendix 10.** An example codeml file (Yang 2007), used to build all evolutionary models (one model, nearly neutral model and positive selection model). Comments occur after asterisks. Variables are in capital letters.

```
seqfile = INPUTFILE    * sequence data file name
treefile = TREEFILE    * tree structure file name
outfile = OUTPUTFILE  * main result file name

noisy = 9              * 0,1,2,3,9: how much rubbish on the screen
verbose = 1            * 1: detailed output, 0: concise output
runmode = 0            * 0: user tree; 1: semi-automatic; 2: automatic
* 3: StepwiseAddition; (4,5):PerturbationNNI; -2: pairwise

seqtype = 1            * 1:codons; 2:AAs; 3:codons-->AAs
CodonFreq = 2          * 0:1/61 each, 1:F1X4, 2:F3X4, 3:codon table
clock = 0              * 0: no clock, unrooted tree, 1: clock, rooted tree
aaDist = 0             * 0:equal, +:geometric; -:linear, (1-5:G1974,Miyata,c,p,v)
model = 0

NSsites = 0 1 2        * 0:one w; 1:NearlyNeutral; 2:PositiveSelection; 3:discrete;
* 4:freqs; 5:gamma;6:2gamma;7:beta;8:beta&w;9:beta&gamma;10:3normal
icode = 0              * 0:standard genetic code; 1:mammalian mt; 2-10:see below
Mgene = 0              * 0:rates, 1:separate; 2:pi, 3:kappa, 4:all

fix_kappa = 0          * 1: kappa fixed, 0: kappa to be estimated
kappa = 1              * initial or fixed kappa
fix_omega = 0          * 1: omega or omega_1 fixed, 0: estimate
omega = 1              * initial or fixed omega, for codons or codon-based AAs
ncatG = 10             * # of categories in the dG or AdG models of rates

getSE = 0              * 0: don't want them, 1: want S.E.s of estimates
RateAncestor = 0       * (0,1,2): rates (alpha>0) or ancestral states (1 or 2)
Small_Diff = .45e-6
cleandata = 1          * remove sites with ambiguity data (1:yes, 0:no)?
fix_blength = 0                * 0: ignore, -1: random, 1: initial, 2: fixed
```

**Appendix 11.** Samples selected for genome sequencing from the University of the Sunshine Coast (USC) sample archive. Samples were selected based on a combination of geographical origin and genome copy numbers.

| Sample_ID | Region | Population | Infection site | Platform | Accession | Genome copies/µL ^ | Clinical Signs* |
|---|---|---|---|---|---|---|---|
| 201_Belvedere_S1_UGT | Not recorded | Northern | UGT | HiSeq 2000 | ERR1354881 | 620168115 | ND |
| 202_Cougar_S1_UGT | Not recorded | Northern | UGT | HiSeq 2000 | ERR1354882 | 46566520 | ND |
| 203_Bella_290513_UGT | Qld | Northern | UGT | HiSeq 2000 | ERR1354883 | 2000000 | ND |
| 204_R1_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1354884 | 87618 | None |
| 205_Mango_S1_UGT | Not recorded | Northern | UGT | HiSeq 2000 | ERR1354885 | 18000 | ND |
| 206_Bev_S1_Ocular | Not recorded | Northern | Conjunctiva | HiSeq 2000 | ERR1354886 | 14000 | ND |
| 207_Savannah_S2_UGT | Not recorded | Northern | UGT | HiSeq 2000 | ERR1354887 | 52000 | ND |
| 208_Buddy_S1_Ocular | NSW | Northern | Conjunctiva | HiSeq 2000 | ERR1354888 | 32300 | Conjunctivitis |
| USC_Eleena_UGT | Not recorded | Northern | UGT | HiSeq 2500 | ERR1821171 | 3700 | ND |
| USC_Gun_koa1_UGT | Gunnedah, NSW | Northern | UGT | HiSeq 2000 | SRR1693763 | 160000 | Urinary tract infection |
| USC_Haz_Bo_Eye | Hazeldean, Qld | Northern | Conjunctiva | MiSeq | ERR710257 | ND | Conjunctivitis |
| USC_QLD_Bobby_UGT | Hazeldean, Qld | Northern | UGT | MiSeq | ERR710258 | ND | Conjunctivitis |
| USC_Lindsay_S1_UGT | Not recorded | Northern | UGT | HiSeq 2500 | ERR1821169 | 8500 | ND |
| USC_Max_S1_UGT | Not recorded | Northern | UGT | HiSeq 2500 | ERR1821170 | 7000 | ND |
| USC_No_Heri_Eye | Noosa, Qld | Northern | Conjunctiva | MiSeq | ERR710261 | ND | Conjunctivitis |
| USC_NSW_Adelle_LE | Lismore, NSW | Northern | Conjunctiva | HiSeq 2000 | ERR1067635 | 2200163 | Conjunctivitis, ovarian cysts |
| USC_NSW_Chingee_Eye | Chingee Creek, NSW | Northern | Conjunctiva | HiSeq 2000 | ERR1067624 | 160000 | conjunctivitis, keratitis, cystitis |
| USC_NSW_Chingee_UGT | Chingee Creek, NSW | Northern | UGT | HiSeq 2000 | ERR1067625 | 155000 | conjunctivitis, keratitis, cystitis |
| USC_NSW_Dobby_UGT1 | Lismore, NSW | Northern | UGT | MiSeq | ERR710260 | ND | Cystitis |
| USC_NSW_Dobby_UGT2 | Lismore, NSW | Northern | UGT | HiSeq 2000 | ERR1067623 | 322 | Cystitis |
| USC_NSW_Elmo_Eye | NSW | Northern | Conjunctiva | HiSeq 2500 | ERR1821179 | 7052 | Conjunctivitis |
| USC_NSW_Knox_Eye | NSW | Northern | Conjunctiva | HiSeq 2500 | ERR1821176 | 2500 | Conjunctivitis |
| USC_PM_03_UGT | Port Macquarie, NSW | Northern | UGT | MiSeq | ERR710251 | ND | None |
| USC_PM_11_UGT1 | Port Macquarie, NSW | Northern | UGT | MiSeq | ERR710252 | ND | None |
| USC_PM_11_UGT2 | Port Macquarie, NSW | Northern | UGT | HiSeq 2000 | ERR1067618 | 99251 | None |
| USC_PM_13_UGT1 | Port Macquarie, NSW | Northern | UGT | MiSeq | ERR710253 | ND | None |

| Sample_ID | Region | Population | Infection site | Platform | Accession | Genome copies/µL ^ | Clinical Signs* |
|---|---|---|---|---|---|---|---|
| USC_PM_13_UGT2 | Port Macquarie, NSW | Northern | UGT | HiSeq 2000 | ERR1067619 | 66033 | None |
| USC_PM_15_UGT | Port Macquarie, NSW | Northern | UGT | MiSeq | ERR710254 | ND | None |
| USC_PM_17_UGT1 | Port Macquarie, NSW | Northern | UGT | MiSeq | ERR710255 | ND | None |
| USC_PM_17_UGT2 | Port Macquarie, NSW | Northern | UGT | HiSeq 2000 | ERR1067620 | 26031 | None |
| USC_PM_3_UGT | Port Macquarie, NSW | Northern | UGT | HiSeq 2000 | ERR1067621 | 18008 | None |
| USC_Posh_S1_Eye | Not recorded | Northern | Conjunctiva | HiSeq 2000 | ERR1067648 | 67000 | None |
| USC_Posh_S1_UGT | Not recorded | Northern | UGT | HiSeq 2000 | ERR1067647 | 70000 | None |
| USC_QLD_Amelia_Eye | Eungella, Qld | Northern | Conjunctiva | HiSeq 2000 | ERR1067634 | 59620 | Ovarian cysts |
| USC_QLD_Chestnut_LE | Tamborine, Qld | Northern | Conjunctiva | HiSeq 2000 | ERR1067627 | 245000 | Conjunctivitis, cystitis |
| USC_QLD_Chestnut_RE | Tamborine, Qld | Northern | Conjunctiva | HiSeq 2000 | ERR1067626 | 3470000 | Conjunctivitis, cystitis |
| USC_QLD_Chestnut_UGT | Tamborine, Qld | Northern | UGT | HiSeq 2000 | ERR1067628 | 2319868 | Conjunctivitis, cystitis |
| USC_QLD_Evie_UGT | Elanora, Qld | Northern | UGT | HiSeq 2000 | ERR1067632 | 264755 | Cystitis |
| USC_QLD_Helen_UGT | Helensvale, Qld | Northern | UGT | HiSeq 2000 | ERR1067636 | 4107026 | Ovarian cysts |
| USC_QLD_Jasper_LE | Currumbin Waters, Qld | Northern | Conjunctiva | HiSeq 2000 | ERR1067633 | 186446 | Conjunctivitis |
| USC_QLD_Talle_UGT | Tallebudgera, Qld | Northern | UGT | HiSeq 2000 | ERR1067629 | 1618043 | Ovarian cysts |
| USC_QLD_Travis_LE | Jimboomba, Qld | Northern | Conjunctiva | HiSeq 2000 | ERR1067630 | 734031 | Conjunctivitis, cystitis |
| USC_QLD_Travis_UGT | Qld | Northern | UGT | HiSeq 2000 | ERR1067631 | 9841135 | Conjunctivitis, cystitis |
| USC_RayIs_13-14_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2500 | ERR1821172 | 600 | None |
| USC_RayIs_18_UGT1 | Raymond Island, Vic | Southern | UGT | MiSeq | ERR710256 | ND | Cystitis |
| USC_RayIs_18_UGT2 | Raymond Island, Vic | Southern | UGT | HiSeq 2000 | ERR1067622 | 726 | Cystitis |
| USC_RayIs_6-14_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2500 | ERR1821173 | 500 | None |
| USC_RayIs_7-14_Eye | Raymond Island, Vic | Southern | Conjunctiva | HiSeq 2500 | ERR1821174 | 400 | None |
| USC_RayIs_9-14_UGT | Raymond Island, Vic | Southern | UGT | HiSeq 2500 | ERR1821175 | 400 | None |
| USC_SA_12-216_UGT | Adelaide Hills, SA | Southern | UGT | HiSeq 2000 | ERR1067650 | 1809 | None |
| USC_SA_12-217_Eye | Adelaide Hills, SA | Southern | Conjunctiva | HiSeq 2000 | ERR1067653 | 5758 | None |
| USC_SA_12-217_UGT | Adelaide Hills, SA | Southern | UGT | HiSeq 2000 | ERR1067654 | 871 | None |
| USC_SA_12-220_LE | Adelaide Hills, SA | Southern | Conjunctiva | HiSeq 2000 | ERR1067651 | 3779 | None |
| USC_SA_12-220_RE | Adelaide Hills, SA | Southern | Conjunctiva | HiSeq 2000 | ERR1067652 | 97226 | None |
| USC_SA_12-327_Eye | Adelaide Hills, SA | Southern | Conjunctiva | HiSeq 2000 | ERR1067658 | 27040 | None |
| USC_SA_12-327_UGT | Adelaide Hills, SA | Southern | UGT | HiSeq 2000 | ERR1067659 | 20453 | None |
| USC_SA_12-342_UGT | Adelaide Hills, SA | Southern | UGT | HiSeq 2000 | ERR1067656 | 1277 | None |
| USC_SA_12-414_Eye | Adelaide Hills, SA | Southern | Conjunctiva | HiSeq 2000 | ERR1067655 | 86011 | None |
| USC_SA_13-128_UGT | Adelaide Hills, SA | Southern | UGT | HiSeq 2000 | ERR1067661 | 6191 | Cystitis |

| Sample_ID | Region | Population | Infection site | Platform | Accession | Genome copies/µL ^ | Clinical Signs* |
|---|---|---|---|---|---|---|---|
| USC_SA_13-152_Eye | Adelaide Hills, SA | Southern | Conjunctiva | HiSeq 2000 | ERR1067660 | 2133 | None |
| USC_SA_13-84_UGT | Adelaide Hills, SA | Southern | UGT | HiSeq 2000 | ERR1067657 | 7506 | Urinary tract infection |
| USC_SA_13-9_UGT | Adelaide Hills, SA | Southern | UGT | HiSeq 2000 | ERR1067649 | 8597 | None |
| USC_SA_k2_UGT | SA | Southern | UGT | HiSeq 2000 | SRR1693792 | 2253 | Urinary tract infection |
| USC_Tash_S1_Eye1 | Not recorded | Northern | Conjunctiva | HiSeq 2500 | ERR1821177 | 5000 | ND |
| USC_Tash_S1_Eye2 | Not recorded | Northern | Conjunctiva | HiSeq 2500 | ERR1821178 | 5000 | ND |
| USC_Ted_Hu_UGT | Teddington, Qld | Northern | UGT | MiSeq | ERR710259 | ND | Cystitis |
| USC_Tya_Butler_LE | Tyagarah, NSW | Northern | Conjunctiva | HiSeq 2000 | ERR1067664 | 41884 | Conjunctivitis |
| USC_Tya_Kasey_UGT | Tyagarah, NSW | Northern | UGT | HiSeq 2000 | ERR1067663 | 37593 | Cystitis |
| USC_Tya_Mavis_LE | Tyagarah, NSW | Northern | Conjunctiva | HiSeq 2000 | ERR1067662 | 915 | Conjunctivitis |

* Urogenital abnormalities assessed via ultrasound or gross pathology at necropsy

ND – Not determined, or information unavailable

**Appendix 12.** Total number of Illumina paired end reads from koala *C. pecorum* samples before and after trimming with Trimmomatic (Bolger *et al.* 2014), the average percentage of guanine and cytidine bases, quality post-trimming and the total number of reads removed through the quality control process.

| Sample | Reads before trimming | | Reads after trimming | | | Total reads removed |
|---|---|---|---|---|---|---|
| | Pairs | GC % | Pairs | GC % | Mean quality | |
| 101_Gipps_2G12_F_UGT | 11217172 | 43 | 10827860 | 43 | 35.4 | 389312 |
| 102_RayIs_3D2_B_UGT | 3971590 | 43 | 3765588 | 43 | 35.2 | 206002 |
| 103_RayIs_3F3_B_UGT | 4049441 | 43 | 3863176 | 43 | 35.2 | 186265 |
| 104_WestVic_3F4_B_UGT | 4679181 | 43 | 4457884 | 43 | 35.2 | 221297 |
| 105_WestVic_3C6_B_UGT | 3977023 | 43 | 3758291 | 43 | 35.2 | 218732 |
| 106_WestVic_3G6_B_UGT | 3661999 | 44 | 3457433 | 44 | 35.2 | 204566 |
| 107_RayIs_3D7_B_UGT | 3289438 | 44 | 3112428 | 44 | 35.1 | 177010 |
| 108_RayIs_3E7_B_UGT | 4659339 | 43 | 4439453 | 42 | 35.2 | 219886 |
| 109_RayIs_3C8_B_UGT | 5224998 | 43 | 4968205 | 42 | 35.3 | 256793 |
| 110_RayIs_3D8_B_UGT | 3899532 | 43 | 3704306 | 43 | 35.2 | 195226 |
| 111_RayIs_3E8_B_UGT | 3083099 | 44 | 2909360 | 43 | 35.2 | 173739 |
| 112_RayIs_3A9_B_UGT | 14517595 | 42 | 13971715 | 42 | 35.4 | 545880 |
| 113_RayIs_3C9_B_UGT | 5918023 | 42 | 5639438 | 42 | 35.3 | 278585 |
| 114_RayIs_3D9_B_UGT | 3770737 | 43 | 3593361 | 42 | 35.2 | 177376 |
| 115_RayIs_3E9_B_UGT | 11318873 | 42 | 10878048 | 41 | 35.4 | 440825 |
| 116_SWCoast_3H9_L_UGT | 4190282 | 43 | 3995706 | 43 | 35.2 | 194576 |
| 117_RayIs_3A10_B_UGT | 3755836 | 43 | 3571233 | 43 | 35.2 | 184603 |
| 118_RayIs_3G10_B_UGT | 6216763 | 42 | 5938238 | 42 | 35.3 | 278525 |
| 119_Gipps_3D11_C_UGT | 3489777 | 43 | 3317121 | 43 | 35.2 | 172656 |
| 120_Gipps_3A12_M_UGT | 3453165 | 43 | 3275509 | 43 | 35.2 | 177656 |
| 121_MoPen_3F12_B_UGT | 3950171 | 43 | 3751137 | 43 | 35.2 | 199034 |
| 122_SWCoast_4B5_B_UGT | 5427646 | 43 | 5182968 | 42 | 35.3 | 244678 |
| 123_MoPen_4G5_C_UGT | 2109503 | 43 | 2007452 | 43 | 35.2 | 102051 |
| 124_Gipps_4E6_C_UGT | 3653225 | 43 | 3502914 | 43 | 35.3 | 150311 |
| 125_Gipps_4H6_F_UGT | 3951708 | 44 | 3784872 | 44 | 35.3 | 166836 |
| 126_MoPen_4G8_C_UGT | 2099373 | 43 | 2000541 | 43 | 35.1 | 98832 |
| 127_Gipps_4C10_C_UGT | 4534652 | 42 | 4353589 | 42 | 35.4 | 181063 |
| 128_RayIs_5F5_B_UGT | 2071866 | 43 | 1968131 | 43 | 35.2 | 103735 |
| 129_RayIs_5H6_B_UGT | 121857280 | 41 | 117997843* | 41 | 35.4 | 3859437 |
| 130_RayIs_7B7_B_UGT | 4451842 | 42 | 4236967 | 42 | 35.4 | 214875 |
| 131_WestMelb_7B8_C_UGT | 2375089 | 43 | 2263145 | 43 | 35.2 | 111944 |
| 132_Vic_7E9_B_UGT | 2498586 | 44 | 2370908 | 43 | 35.1 | 127678 |
| 133_SWCoast_7H12_L_UGT | 2042977 | 43 | 1940434 | 43 | 35.2 | 102543 |
| 134_MoPen_15B5_C_UGT | 4312809 | 42 | 4149456 | 42 | 35.4 | 163353 |
| 135_Gipps_15C7_C_UGT | 2413342 | 46 | 2301664 | 46 | 35.3 | 111678 |
| 136_FrIs_15H9_N_UGT | 2303020 | 43 | 2187578 | 43 | 35.2 | 115442 |
| 137_FrIs_16H3_N_UGT | 2293271 | 43 | 2179284 | 43 | 35.2 | 113987 |
| 138_SWCoast_18H3_L_UGT | 2307850 | 43 | 2196410 | 43 | 35.2 | 111440 |
| 139_SWCoast_18C7_L_UGT | 2246401 | 43 | 2133369 | 43 | 35.2 | 113032 |
| 140_SWCoast_18D7_L_UGT | 2501421 | 43 | 2374374 | 43 | 35.2 | 127047 |
| 141_SWCoast_18C8_L_UGT | 1911532 | 44 | 1820493 | 44 | 35.2 | 91039 |
| 201_Belvedere_S1_UGT | 71449325 | 43 | 69505623* | 43 | 35.2 | 1943702 |
| 202_Cougar_S1_UGT | 1817184 | 46 | 1740655 | 46 | 35 | 76529 |
| 203_Bella_290513_UGT | 4340390 | 47 | 4172332 | 47 | 35 | 168058 |
| 204_R1_UGT | 4942613 | 46 | 4796893 | 46 | 35.2 | 145720 |
| 205_Mango_S1_UGT | 1772365 | 46 | 1702472 | 46 | 35 | 69893 |
| 206_Bev_S1_Ocular | 1871393 | 46 | 1799329 | 45 | 35.1 | 72064 |
| 207_Savannah_S2_UGT | 92405616 | 41 | 90086572* | 41 | 35.3 | 2319044 |
| 208_Buddy_S1_Ocular | 11793333 | 42 | 11493937 | 42 | 35.2 | 299396 |
| USC_Eleena_UGT | 25166675 | 43 | 22742446* | 43 | 34.6 | 2424229 |
| USC_Gun_koa1_UGT | 6590405 | 41 | 6227289 | 41 | 35.1 | 363116 |
| USC_Haz_Bo_Eye | 1283013 | 41 | 1277132 | 41 | 37.3 | 5881 |
| USC_Lindsay_S1_UGT | 10701140 | 46 | 9687723 | 46 | 34.7 | 1013417 |

| Sample | Reads before trimming | | Reads after trimming | | | Total reads |
|---|---|---|---|---|---|---|
| | Pairs | GC % | Pairs | GC % | Mean quality | removed |
| USC_Max_S1_UGT | 41719720 | 42 | 38354150* | 42 | 34.9 | 3365570 |
| USC_No_Heri_Eye | 1198426 | 42 | 1192196 | 42 | 37.2 | 6230 |
| USC_NSW_Adelle_LE | 682717 | 42 | 644699 | 42 | 34.8 | 38018 |
| USC_NSW_Chingee_Eye | 399108 | 42 | 377320 | 42 | 34.7 | 21788 |
| USC_NSW_Chingee_UGT | 397073 | 42 | 375034 | 42 | 34.7 | 22039 |
| USC_NSW_Dobby_UGT1 | 1072179 | 42 | 1064862 | 42 | 37.1 | 7317 |
| USC_NSW_Dobby_UGT2 | 80031 | 45 | 73164 | 45 | 34.6 | 6867 |
| USC_NSW_Elmo_Eye | 24385150 | 43 | 22413826* | 42 | 34.9 | 1971324 |
| USC_NSW_Knox_Eye | 15241852 | 44 | 13916437 | 44 | 34.7 | 1325415 |
| USC_PM_03_UGT | 1382011 | 42 | 1373853 | 42 | 37.2 | 8158 |
| USC_PM_11_UGT1 | 1256595 | 42 | 1248938 | 42 | 37.1 | 7657 |
| USC_PM_11_UGT2 | 1927393 | 41 | 1827580 | 41 | 34.7 | 99813 |
| USC_PM_13_UGT1 | 1087425 | 42 | 1081113 | 42 | 37.2 | 6312 |
| USC_PM_13_UGT2 | 3672942 | 41 | 3474407 | 41 | 34.8 | 198535 |
| USC_PM_15_UGT | 1487069 | 42 | 1478554 | 42 | 37.1 | 8515 |
| USC_PM_17_UGT1 | 1217626 | 42 | 1210342 | 42 | 37.1 | 7284 |
| USC_PM_17_UGT2 | 596805 | 42 | 566011 | 42 | 34.6 | 30794 |
| USC_PM_3_UGT | 1047216 | 42 | 991553 | 42 | 34.7 | 55663 |
| USC_Posh_S1_Eye | 2170672 | 44 | 1988463 | 44 | 34.6 | 182209 |
| USC_Posh_S1_UGT | 1576224 | 46 | 1441762 | 46 | 34.5 | 134462 |
| USC_QLD_Amelia_Eye | 139702 | 43 | 131622 | 43 | 34.7 | 8080 |
| USC_QLD_Bobby_UGT | 1405272 | 41 | 1397469 | 41 | 37.2 | 7803 |
| USC_QLD_Chestnut_LE | 688595 | 42 | 650466 | 42 | 34.8 | 38129 |
| USC_QLD_Chestnut_RE | 1250672 | 41 | 1175118 | 41 | 34.8 | 75554 |
| USC_QLD_Chestnut_UGT | 1427135 | 41 | 1344065 | 41 | 34.8 | 83070 |
| USC_QLD_Evie_UGT | 203144 | 42 | 191296 | 42 | 34.7 | 11848 |
| USC_QLD_Helen_UGT | 852586 | 42 | 807302 | 41 | 34.7 | 45284 |
| USC_QLD_Jasper_LE | 956356 | 42 | 907456 | 41 | 34.7 | 48900 |
| USC_QLD_Talle_UGT | 1668680 | 41 | 1580458 | 41 | 34.8 | 88222 |
| USC_QLD_Travis_LE | 792516 | 42 | 744644 | 42 | 34.9 | 47872 |
| USC_QLD_Travis_UGT | 6317787 | 41 | 6006301 | 41 | 34.7 | 311486 |
| USC_RayIs_13-14_UGT | 15635768 | 49 | 14485549 | 49 | 34.8 | 1150219 |
| USC_RayIs_18_UGT1 | 1458962 | 42 | 1449900 | 42 | 37.1 | 9062 |
| USC_RayIs_18_UGT2 | 129537 | 44 | 119103 | 44 | 34.7 | 10434 |
| USC_RayIs_6-14_UGT | 15347420 | 49 | 13940471 | 48 | 35 | 1406949 |
| USC_RayIs_7-14_Eye | 9825288 | 46 | 8775074 | 46 | 34.8 | 1050214 |
| USC_RayIs_9-14_UGT | 8342413 | 46 | 7635297 | 46 | 34.6 | 707116 |
| USC_SA_12-216_UGT | 3389853 | 45 | 3068758 | 45 | 34.2 | 321095 |
| USC_SA_12-217_Eye | 1449888 | 46 | 1324069 | 46 | 34.5 | 125819 |
| USC_SA_12-217_UGT | 2095016 | 49 | 1915555 | 48 | 34.6 | 179461 |
| USC_SA_12-220_LE | 1859262 | 45 | 1708619 | 45 | 34.5 | 150643 |
| USC_SA_12-220_RE | 1754839 | 46 | 1601411 | 45 | 34.5 | 153428 |
| USC_SA_12-327_Eye | 1351208 | 46 | 1231096 | 46 | 34.5 | 120112 |
| USC_SA_12-327_UGT | 1372234 | 46 | 1261334 | 45 | 34.5 | 110900 |
| USC_SA_12-342_UGT | 1266008 | 47 | 1157736 | 46 | 34.5 | 108272 |
| USC_SA_12-414_Eye | 1370230 | 46 | 1243561 | 46 | 34.5 | 126669 |
| USC_SA_13-128_UGT | 1300609 | 46 | 1192635 | 46 | 34.5 | 107974 |
| USC_SA_13-152_Eye | 1296849 | 46 | 1187008 | 46 | 34.5 | 109841 |
| USC_SA_13-84_UGT | 1675009 | 48 | 1522154 | 47 | 34.6 | 152855 |
| USC_SA_13-9_UGT | 5352505 | 49 | 4945508 | 49 | 34.7 | 406997 |
| USC_SA_k2_UGT | 2293421 | 42 | 2158153 | 42 | 34.9 | 135268 |
| USC_Tash_S1_Eye1 | 9780918 | 47 | 8681207 | 46 | 34.7 | 1099711 |
| USC_Tash_S1_Eye2 | 9489697 | 46 | 8663993 | 46 | 34.8 | 825704 |
| USC_Ted_Hu_UGT | 1224416 | 42 | 1217536 | 41 | 37.2 | 6880 |
| USC_Tya_Butler_LE | 15865916 | 41 | 14942334 | 41 | 34.7 | 923582 |
| USC_Tya_Kasey_UGT | 1581248 | 46 | 1438561 | 46 | 34.5 | 142687 |
| USC_Tya_Mavis_LE | 8029783 | 42 | 7504327 | 42 | 34.6 | 525456 |

* Samples with more than 20,000,000 high quality reads were subsampled to this depth to allow for further processing.

**Appendix 13.** Coverage of E58 *C. pecorum* genome (1,106,197 bp) by Illumina paired end reads from *C. pecorum* samples. Reads were aligned using 'bwa mem' (Li & Durbin 2010) and coverage was assessed using BEDtools (Quinlan & Hall 2010).

| Sample Name | Depth of coverage | | | | Coverage percentage (%) | | Assembled^ |
|---|---|---|---|---|---|---|---|
| | Average | Q1 | Median | Q3 | Depth > 0 | Depth > 10 | |
| 101_Gipps_2G12_F_UGT | 956.1 | 694 | 852 | 1020 | 98.46 | 98.21 | Yes |
| 102_RayIs_3D2_B_UGT | 26.12 | 18 | 24 | 32 | 98.32 | 92.9 | Yes |
| 103_RayIs_3F3_B_UGT | 75.51 | 61 | 76 | 91 | 98.56 | 98.09 | Yes |
| 104_WestVic_3F4_B_UGT | 112.3 | 91 | 112 | 134 | 98.6 | 98.18 | Yes |
| 105_WestVic_3C6_B_UGT | 21.88 | 14 | 20 | 26 | 98.26 | 87.21 | |
| 106_WestVic_3G6_B_UGT | 9.03 | 5 | 8 | 12 | 95.88 | 34.2 | |
| 107_RayIs_3D7_B_UGT | 14.67 | 9 | 13 | 19 | 97.74 | 66.27 | |
| 108_RayIs_3E7_B_UGT | 140.91 | 118 | 141 | 165 | 98.7 | 98.27 | Yes |
| 109_RayIs_3C8_B_UGT | 267.54 | 231 | 272 | 313 | 98.71 | 98.45 | Yes |
| 110_RayIs_3D8_B_UGT | 47.8 | 34 | 44 | 54 | 98.47 | 97.44 | Yes |
| 111_RayIs_3E8_B_UGT | 5.83 | 2 | 5 | 8 | 90.7 | 13.26 | |
| 112_RayIs_3A9_B_UGT | 1336.28 | 1145 | 1357 | 1574 | 98.79 | 98.65 | Yes |
| 113_RayIs_3C9_B_UGT | 283.04 | 238 | 286 | 336 | 98.69 | 98.41 | Yes |
| 114_RayIs_3D9_B_UGT | 46.71 | 36 | 46 | 56 | 98.5 | 97.63 | Yes |
| 115_RayIs_3E9_B_UGT | 924.01 | 816 | 946 | 1066 | 98.79 | 98.62 | Yes |
| 116_SWCoast_3H9_L_UGT | 83.21 | 55 | 68 | 82 | 98.89 | 98.07 | Yes |
| 117_RayIs_3A10_B_UGT | 38.55 | 27 | 37 | 46 | 98.46 | 96.74 | Yes |
| 118_RayIs_3G10_B_UGT | 325.77 | 280 | 331 | 381 | 98.73 | 98.44 | Yes |
| 119_Gipps_3D11_C_UGT | 2.64 | 0 | 2 | 4 | 69.44 | 1.27 | |
| 120_Gipps_3A12_M_UGT | 4.82 | 1 | 3 | 6 | 83.75 | 5.63 | |
| 121_MoPen_3F12_B_UGT | 38.38 | 27 | 36 | 47 | 98.22 | 96.65 | Yes |
| 122_SWCoast_4B5_B_UGT | 241.02 | 205 | 244 | 281 | 98.66 | 98.38 | Yes |
| 123_MoPen_4G5_C_UGT | 24.06 | 18 | 24 | 29 | 98.86 | 94.55 | Yes |
| 124_Gipps_4E6_C_UGT | 192.76 | 148 | 180 | 212 | 98.59 | 97.71 | Yes |
| 125_Gipps_4H6_F_UGT | 182.13 | 129 | 153 | 177 | 98.65 | 97.73 | Yes |
| 126_MoPen_4G8_C_UGT | 18.77 | 14 | 18 | 23 | 98.8 | 87.15 | |
| 127_Gipps_4C10_C_UGT | 323.74 | 268 | 321 | 371 | 99.05 | 98.41 | Yes |
| 128_RayIs_5F5_B_UGT | 10.22 | 6 | 9 | 13 | 97.89 | 40.45 | |
| 129_RayIs_5H6_B_UGT | 2506.68 | 1640 | 1975 | 2386 | 98.98 | 98.67 | Yes |
| 130_RayIs_7B7_B_UGT | 265.81 | 235 | 273 | 308 | 98.8 | 98.43 | Yes |
| 131_WestMelb_7B8_C_UGT | 18.68 | 7 | 10 | 14 | 98.32 | 46.65 | |
| 132_Vic_7E9_B_UGT | 17.66 | 12 | 17 | 21 | 98.56 | 83.89 | |
| 133_SWCoast_7H12_L_UGT | 5.61 | 2 | 4 | 6 | 92.94 | 2.24 | |
| 134_MoPen_15B5_C_UGT | 290.94 | 234 | 283 | 335 | 99.09 | 98.4 | Yes |
| 135_Gipps_15C7_C_UGT | 39.75 | 1 | 2 | 3 | 79.23 | 0.41 | |
| 136_FrIs_15H9_N_UGT | 2.67 | 1 | 2 | 4 | 84.92 | 0.38 | |
| 137_FrIs_16H3_N_UGT | 2.19 | 1 | 2 | 3 | 79.61 | 0.22 | |
| 138_SWCoast_18H3_L_UGT | 6.62 | 1 | 2 | 4 | 85.7 | 0.5 | |
| 139_SWCoast_18C7_L_UGT | 7.44 | 3 | 6 | 8 | 97.03 | 10.7 | |
| 140_SWCoast_18D7_L_UGT | 8.14 | 2 | 4 | 6 | 93.32 | 2.6 | |
| 141_SWCoast_18C8_L_UGT | 19.25 | 6 | 9 | 12 | 98.34 | 38.01 | |
| 201_Belvedere_S1_UGT | 2478.2 | 1452 | 2404 | 3162 | 99.15 | 98.41 | Yes |
| 202_Cougar_S1_UGT | 1.66 | 0 | 1 | 2 | 62.33 | 0.31 | |
| 203_Bella_290513_UGT | 167.28 | 74 | 115 | 182 | 98.8 | 97.89 | Yes |
| 204_R1_UGT | 274.51 | 134 | 160 | 186 | 99.17 | 98.34 | Yes |
| 205_Mango_S1_UGT | 9.13 | 5 | 8 | 12 | 97.55 | 34.98 | |
| 206_Bev_S1_Ocular | 7.55 | 2 | 5 | 7 | 93.1 | 9.02 | |
| 207_Savannah_S2_UGT | 2521.35 | 1839 | 2226 | 2656 | 99.18 | 98.4 | Yes |
| 208_Buddy_S1_Ocular | 1197.63 | 1019 | 1228 | 1428 | 99.21 | 98.31 | Yes |
| USC_Eleena_UGT | 1757.4 | 1539 | 1789 | 2030 | 98.73 | 98.52 | Yes |
| USC_Gun_koa1_UGT | 1076.94 | 965 | 1089 | 1216 | 100 | 99.96 | Yes |
| USC_Haz_Bo_Eye | 122.38 | 113 | 124 | 135 | 98.66 | 98.52 | Yes |
| USC_Lindsay_S1_UGT | 22.44 | 0 | 0 | 0 | 23.05 | 4.65 | |
| USC_Max_S1_UGT | 2715.95 | 2344 | 2711 | 3074 | 98.75 | 98.57 | Yes |
| USC_No_Heri_Eye | 39.74 | 35 | 40 | 46 | 96.99 | 96.68 | Yes |

| Sample Name | Depth of coverage | | | | Coverage percentage (%) | | Assembled^ |
|---|---|---|---|---|---|---|---|
| | Average | Q1 | Median | Q3 | Depth > 0 | Depth > 10 | |
| USC_NSW_Adelle_LE | 71.12 | 58 | 72 | 86 | 99.87 | 97.84 | Yes |
| USC_NSW_Chingee_Eye | 42.92 | 34 | 43 | 52 | 99.63 | 97.34 | Yes |
| USC_NSW_Chingee_UGT | 43.07 | 34 | 43 | 53 | 99.54 | 97.33 | Yes |
| USC_NSW_Dobby_UGT1 | 0.09 | 0 | 0 | 0 | 7.25 | 0.01 | |
| USC_NSW_Dobby_UGT2 | 1.53 | 0 | 1 | 2 | 74.46 | 0.01 | |
| USC_NSW_Elmo_Eye | 2777.13 | 2408 | 2823 | 3239 | 98.69 | 98.53 | Yes |
| USC_NSW_Knox_Eye | 651.69 | 554 | 660 | 763 | 98.45 | 98.28 | Yes |
| USC_PM_03_UGT | 1.28 | 0 | 1 | 2 | 71.57 | 0.01 | |
| USC_PM_11_UGT1 | 2.21 | 1 | 2 | 3 | 87.69 | 0.02 | |
| USC_PM_11_UGT2 | 226.67 | 188 | 229 | 271 | 99.97 | 99.14 | Yes |
| USC_PM_13_UGT1 | 3.62 | 2 | 3 | 5 | 96.12 | 0.28 | |
| USC_PM_13_UGT2 | 438.7 | 367 | 444 | 521 | 99.97 | 99.67 | Yes |
| USC_PM_15_UGT | 7.61 | 6 | 7 | 10 | 98.89 | 16.29 | |
| USC_PM_17_UGT1 | 0.79 | 0 | 1 | 1 | 52.66 | 0.01 | |
| USC_PM_17_UGT2 | 65.23 | 53 | 66 | 78 | 99.79 | 98.29 | Yes |
| USC_PM_3_UGT | 111.53 | 92 | 112 | 133 | 99.93 | 98.77 | Yes |
| USC_Posh_S1_Eye | 79.34 | 61 | 78 | 96 | 99.24 | 98.58 | Yes |
| USC_Posh_S1_UGT | 5.26 | 1 | 4 | 7 | 78.17 | 13.03 | |
| USC_QLD_Amelia_Eye | 9.85 | 7 | 10 | 13 | 99.16 | 41.08 | |
| USC_QLD_Bobby_UGT | 141.79 | 132 | 144 | 155 | 98.67 | 98.55 | Yes |
| USC_QLD_Chestnut_LE | 71.06 | 58 | 72 | 86 | 99.83 | 98.1 | Yes |
| USC_QLD_Chestnut_RE | 147.25 | 122 | 149 | 176 | 99.91 | 98.58 | Yes |
| USC_QLD_Chestnut_UGT | 168.96 | 141 | 171 | 201 | 99.9 | 98.69 | Yes |
| USC_QLD_Evie_UGT | 11.2 | 8 | 11 | 14 | 99.34 | 53.65 | |
| USC_QLD_Helen_UGT | 86.04 | 71 | 87 | 103 | 99.9 | 98.04 | Yes |
| USC_QLD_Jasper_LE | 101.62 | 83 | 102 | 122 | 99.92 | 98.67 | Yes |
| USC_QLD_Talle_UGT | 196.97 | 164 | 200 | 235 | 99.95 | 98.85 | Yes |
| USC_QLD_Travis_LE | 85.91 | 70 | 86 | 103 | 99.84 | 98.13 | Yes |
| USC_QLD_Travis_UGT | 754.39 | 635 | 766 | 895 | 99.96 | 99.82 | Yes |
| USC_RayIs_13-14_UGT | 962.45 | 342 | 437 | 530 | 99.06 | 98.77 | Yes |
| USC_RayIs_18_UGT1 | 0.17 | 0 | 0 | 0 | 14.22 | 0.01 | |
| USC_RayIs_18_UGT2 | 3.22 | 2 | 3 | 4 | 92.35 | 0.3 | |
| USC_RayIs_6-14_UGT | 779 | 193 | 243 | 294 | 98.88 | 98.58 | Yes |
| USC_RayIs_7-14_Eye | 81.21 | 42 | 62 | 86 | 98.49 | 96.96 | Yes |
| USC_RayIs_9-14_UGT | 43.97 | 19 | 32 | 47 | 97.29 | 88.96 | |
| USC_SA_12-216_UGT | 0.93 | 0 | 0 | 1 | 41.29 | 0.32 | |
| USC_SA_12-217_Eye | 0.63 | 0 | 0 | 0 | 20.38 | 0.3 | |
| USC_SA_12-217_UGT | 35.43 | 0 | 0 | 1 | 46.62 | 0.5 | |
| USC_SA_12-220_LE | 38.08 | 27 | 37 | 47 | 99.84 | 98.23 | Yes |
| USC_SA_12-220_RE | 4.24 | 0 | 3 | 6 | 71.68 | 8.87 | |
| USC_SA_12-327_Eye | 3.38 | 0 | 2 | 5 | 64.9 | 6.04 | |
| USC_SA_12-327_UGT | 3.86 | 0 | 3 | 6 | 71.82 | 7.47 | |
| USC_SA_12-342_UGT | 1.44 | 0 | 0 | 1 | 34.68 | 0.81 | |
| USC_SA_12-414_Eye | 0.93 | 0 | 0 | 1 | 27.52 | 0.3 | |
| USC_SA_13-128_UGT | 0.88 | 0 | 0 | 1 | 44.98 | 0.23 | |
| USC_SA_13-152_Eye | 0.69 | 0 | 0 | 1 | 38.97 | 0.17 | |
| USC_SA_13-84_UGT | 28.63 | 5 | 9 | 15 | 93.35 | 43.31 | |
| USC_SA_13-9_UGT | 256.12 | 104 | 129 | 154 | 99.25 | 98.32 | Yes |
| USC_SA_k2_UGT | 332.94 | 264 | 313 | 363 | 99.74 | 98.56 | Yes |
| USC_Tash_S1_Eye1 | 89.15 | 20 | 34 | 51 | 96.87 | 88.38 | |
| USC_Tash_S1_Eye2 | 25.58 | 5 | 13 | 24 | 84.88 | 56.96 | |
| USC_Ted_Hu_UGT | 128.32 | 119 | 130 | 141 | 98.49 | 98.32 | Yes |
| USC_Tya_Butler_LE | 1585.72 | 1403 | 1613 | 1814 | 99.35 | 98.69 | Yes |
| USC_Tya_Kasey_UGT | 4.71 | 0 | 3 | 7 | 73.29 | 10.02 | |
| USC_Tya_Mavis_LE | 624.18 | 546 | 634 | 720 | 99.26 | 98.55 | Yes |

^ Samples with a depth of coverage greater than 10 over more than 90% of the genome were used for *de novo* assembly.

**Appendix 14.** Summary of *de novo* assembly of *C. pecorum* samples, with contigs mapped to a custom BLAST database containing *Chlamydiales* reference genomes. QUAST (Gurevich *et al.* 2013) analysis was also conducted on contigs, using the type strain E58 as a reference, to determine assembly metrics and protentional misassemblies.

| Sample Name | Contigs | Contigs ≥ 1000 bp | Mean contig length (bp) (min – max) | Sum of contig lengths (≥ 1000 bp) | Average % identity (min – max) * | Genome fraction (%) ^ | N50 | Misassembled contigs |
|---|---|---|---|---|---|---|---|---|
| 101_Gipps_2G12_F_UGT | 189 | 21 | 54329 (1092 - 627812) | 1140910 | 85.6 (74.5 - 99.7) | 95.2 | 627812 | 0 |
| 102_RayIs_3D2_B_UGT | 76 | 58 | 19290 (1030 - 170207) | 1118839 | 95.2 (75.8 - 99.9) | 95.3 | 42625 | 1 |
| 103_RayIs_3F3_B_UGT | 28 | 15 | 74666 (1130 - 627342) | 1119983 | 95.3 (79 - 99.7) | 95.2 | 627342 | 1 |
| 104_WestVic_3F4_B_UGT | 41 | 14 | 80289 (1197 - 758468) | 1124045 | 92 (75.5 - 99.7) | 95.0 | 758468 | 0 |
| 108_RayIs_3E7_B_UGT | 36 | 16 | 70318 (1100 - 768609) | 1125087 | 91.3 (75.8 - 99.7) | 95.0 | 768609 | 0 |
| 109_RayIs_3C8_B_UGT | 15 | 7 | 160097 (1044 - 759543) | 1120678 | 93.6 (79.5 - 99.7) | 95.0 | 759543 | 1 |
| 110_RayIs_3D8_B_UGT | 75 | 29 | 39053 (1102 - 562783) | 1132537 | 89.3 (76.8 - 99.7) | 94.9 | 131854 | 0 |
| 112_RayIs_3A9_B_UGT | 20 | 9 | 124844 (1020 - 759679) | 1123595 | 92.4 (78.4 - 99.7) | 95.0 | 759679 | 1 |
| 113_RayIs_3C9_B_UGT | 20 | 10 | 112472 (1026 - 800044) | 1124724 | 88.8 (77.8 - 99.7) | 95.0 | 800044 | 2 |
| 114_RayIs_3D9_B_UGT | 32 | 18 | 62253 (1144 - 725851) | 1120561 | 93.1 (76.2 - 99.7) | 94.9 | 725851 | 0 |
| 115_RayIs_3E9_B_UGT | 20 | 10 | 112564 (1282 - 626732) | 1125635 | 90.7 (76.8 - 99.7) | 94.9 | 626732 | 1 |
| 116_SWCoast_3H9_L_UGT | 102 | 34 | 33747 (1053 - 709471) | 1147393 | 85.3 (74.7 - 100) | 95.3 | 709471 | 1 |
| 117_RayIs_3A10_B_UGT | 61 | 34 | 32978 (1006 - 200662) | 1121238 | 93.1 (76.4 - 99.7) | 95.0 | 114552 | 0 |
| 118_RayIs_3G10_B_UGT | 22 | 11 | 102442 (1037 - 759462) | 1126860 | 89 (75.2 - 99.8) | 95.0 | 759462 | 1 |
| 121_MoPen_3F12_B_UGT | 45 | 28 | 39720 (1248 - 319012) | 1112170 | 92.7 (74.9 - 99.8) | 95.3 | 241512 | 0 |
| 122_SWCoast_4B5_B_UGT | 25 | 10 | 112371 (1258 - 759489) | 1123710 | 90.4 (76.8 - 99.7) | 95.0 | 759489 | 1 |
| 123_MoPen_4G5_C_UGT | 47 | 33 | 33688 (1097 - 383018) | 1111707 | 93.8 (76.3 - 99.7) | 96.8 | 248721 | 0 |
| 124_Gipps_4E6_C_UGT | 56 | 19 | 59496 (1002 - 539888) | 1130431 | 87.5 (75.3 - 99.7) | 95.0 | 258078 | 1 |
| 125_Gipps_4H6_F_UGT | 118 | 26 | 43737 (1052 - 671740) | 1137162 | 87.9 (75.3 - 100) | 95.0 | 671740 | 0 |
| 127_Gipps_4C10_C_UGT | 83 | 25 | 45562 (1011 - 630485) | 1139039 | 85.2 (75 - 99.5) | 97.0 | 630485 | 0 |
| 129_RayIs_5H6_B_UGT | 47 | 18 | 63416 (1048 - 1101827) | 1141490 | 82.2 (75.1 - 99.7) | 95.0 | 1101827 | 1 |
| 130_RayIs_7B7_B_UGT | 28 | 12 | 93670 (1143 - 759454) | 1124037 | 91.3 (75.8 - 99.7) | 95.0 | 759454 | 1 |
| 134_MoPen_15B5_C_UGT | 45 | 20 | 56548 (1061 - 630564) | 1130965 | 87.6 (75 - 99.5) | 97.0 | 630564 | 0 |
| 201_Belvedere_S1_UGT | 122 | 18 | 63013 (1001 - 526264) | 1134230 | 86.7 (75.1 - 100) | 97.1 | 334075 | 0 |
| 203_Bella_290513_UGT | 45 | 12 | 93004 (1093 - 758371) | 1116047 | 96.5 (78.6 - 100) | 96.0 | 758371 | 0 |
| 204_R1_UGT | 162 | 18 | 63196 (1039 - 758512) | 1137527 | 86.1 (73.9 - 99.7) | 94.9 | 758512 | 1 |
| 207_Savannah_S2_UGT | 115 | 15 | 74767 (1146 - 443140) | 1121503 | 90.2 (76.2 - 100) | 97.0 | 315526 | 1 |
| 208_Buddy_S1_Ocular | 40 | 7 | 158449 (2302 - 630144) | 1109142 | 96.1 (81.9 - 99.6) | 97.3 | 630144 | 0 |
| USC_Eleena_UGT | 45 | 17 | 66047 (1162 - 624425) | 1122794 | 94.1 (77.1 - 100) | 96.1 | 624425 | 1 |
| USC_Gun_koa1_UGT | 99 | 27 | 41300 (1035 - 652671) | 1115101 | 93.5 (76.9 - 100) | 96.6 | 652671 | 0 |

| Sample Name | Contigs | Contigs ≥ 1000 bp | Mean contig length (bp) (min – max) | Sum of contig lengths (≥ 1000 bp) | Average % identity (min – max) * | Genome fraction (%) ^ | N50 | Misassembled contigs |
|---|---|---|---|---|---|---|---|---|
| USC_Haz_Bo_Eye | 4 | 3 | 372759 (7624 - 807478) | 1118276 | 99.4 (99.1 - 99.7) | 95.9 | 807478 | 1 |
| USC_Max_S1_UGT | 76 | 17 | 66169 (1019 - 458937) | 1124879 | 92.5 (78.1 - 99.7) | 96.0 | 166280 | 0 |
| USC_NSW_Adelle_LE | 57 | 13 | 85338 (1363 - 498553) | 1109394 | 94.5 (73.4 - 99.7) | 97.2 | 142243 | 0 |
| USC_NSW_Chingee_Eye | 40 | 16 | 69422 (1684 - 498537) | 1110746 | 95.1 (76.7 - 99.7) | 97.7 | 131872 | 0 |
| USC_NSW_Chingee_UGT | 34 | 14 | 79379 (3908 - 498549) | 1111300 | 98.3 (87.8 - 99.7) | 96.3 | 151460 | 0 |
| USC_NSW_Elmo_Eye | 30 | 8 | 139054 (1351 - 630017) | 1112431 | 94.2 (78.4 - 99.7) | 96.5 | 630017 | 0 |
| USC_NSW_Knox_Eye | 24 | 11 | 100917 (1321 - 629939) | 1110091 | 90.7 (73.8 - 99.7) | 94.7 | 629939 | 1 |
| USC_PM_11_UGT2 | 62 | 18 | 62862 (1124 - 630592) | 1131509 | 96.4 (84.6 - 100) | 96.1 | 630592 | 0 |
| USC_PM_13_UGT2 | 83 | 23 | 48971 (1031 - 506026) | 1126333 | 95.5 (76.8 - 100) | 96.1 | 281195 | 0 |
| USC_PM_17_UGT2 | 31 | 9 | 124004 (1667 - 758024) | 1116039 | 94.5 (75.1 - 99.5) | 96.3 | 758024 | 0 |
| USC_PM_3_UGT | 46 | 10 | 112359 (1316 - 630477) | 1123590 | 95.5 (73.2 - 99.7) | 96.7 | 630477 | 1 |
| USC_Posh_S1_Eye | 65 | 20 | 55669 (1284 - 517260) | 1113376 | 95 (76.7 - 99.7) | 97.6 | 176937 | 0 |
| USC_QLD_Bobby_UGT | 12 | 3 | 372771 (7624 - 807517) | 1118312 | 99.4 (99.1 - 99.7) | 95.5 | 807517 | 1 |
| USC_QLD_Chestnut_LE | 47 | 14 | 79603 (1034 - 498598) | 1114445 | 96.7 (85.7 - 99.8) | 97.3 | 133562 | 0 |
| USC_QLD_Chestnut_RE | 50 | 10 | 111683 (1177 - 630024) | 1116828 | 97.4 (91.9 - 99.8) | 96.3 | 630024 | 0 |
| USC_QLD_Chestnut_UGT | 47 | 10 | 111948 (1051 - 630013) | 1119477 | 98.3 (94 - 100) | 98.3 | 630013 | 0 |
| USC_QLD_Helen_UGT | 60 | 20 | 55835 (1169 - 498571) | 1116696 | 96.4 (83.4 - 100) | 92.5 | 165255 | 0 |
| USC_QLD_Jasper_LE | 54 | 12 | 93316 (1406 - 629958) | 1119787 | 97.1 (88.5 - 99.7) | 94.2 | 629958 | 0 |
| USC_QLD_Talle_UGT | 63 | 15 | 74919 (1090 - 498086) | 1123786 | 96.4 (83.4 - 100) | 93.8 | 179366 | 1 |
| USC_QLD_Travis_LE | 42 | 13 | 85763 (1040 - 630564) | 1114922 | 94.9 (83.7 - 99.6) | 99.3 | 630564 | 0 |
| USC_QLD_Travis_UGT | 111 | 41 | 27631 (1004 - 505998) | 1132869 | 92.3 (76.7 - 100) | 95.0 | 145384 | 0 |
| USC_RayIs_13-14_UGT | 8841 | 269 | 4092 (1000 - 31545) | 1100730 | 98.4 (76.7 - 100) | 96.0 | 5227 | 8 |
| USC_RayIs_6-14_UGT | 12566 | 142 | 7923 (1045 - 42219) | 1125028 | 97.9 (75 - 100) | 95.3 | 13672 | 2 |
| USC_RayIs_7-14_Eye | 382 | 121 | 9179 (1009 - 51369) | 1110619 | 97.3 (76.3 - 99.9) | 95.2 | 15824 | 0 |
| USC_SA_12-220_LE | 55 | 39 | 28480 (1065 - 188184) | 1110717 | 99.3 (94.5 - 100) | 95.2 | 69285 | 0 |
| USC_SA_13-9_UGT | 67 | 15 | 75039 (1095 - 630520) | 1125581 | 88.8 (76.6 - 99.7) | 95.3 | 630520 | 0 |
| USC_SA_K2_UGT | 94 | 18 | 61980 (1142 - 667335) | 1115646 | 87.4 (75.6 - 99.4) | 95.2 | 667335 | 0 |
| USC_Ted_Hu_UGT | 8 | 2 | 559032 (7624 - 1110440) | 1118064 | 99.3 (99.1 - 99.5) | 95.0 | 1110440 | 1 |
| USC_Tya_Butler_LE | 31 | 7 | 159781 (7567 - 317954) | 1118468 | 99.4 (98.9 - 99.7) | 95.0 | 180625 | 0 |
| USC_Tya_Mavis_LE | 25 | 10 | 111775 (1238 - 630479) | 1117749 | 95.6 (77.7 - 100) | 95.0 | 630479 | 0 |

[*] Identity of contigs to custom BLAST database containing nucleotide sequences of reference genomes from the *Chlamydiaceae* family, in addition to the *C. pecorum* plasmid

[^] Percentage of coverage of E58 reference genome in QUAST using contigs with a successful BLAST hit.

**Appendix 15.** Gene annotation, using Prokka (Seemann 2014), for each *C. pecorum* genome assembled *de novo*. Previously constructed genomes (DBDeUG, IpTaLE, and MC_Marsbar) were also included.

| Genome | Genome length | Genes | Coding DNA sequence | Signal peptides | Miscellaneous RNA* | rRNA | tmRNA | tRNA |
|---|---|---|---|---|---|---|---|---|
| 101_Gipps_2G12_F_UGT | 1107848 | 994 | 946 | 52 | 6 | 3 | 1 | 38 |
| 103_RayIs_3F3_B_UGT | 1110360 | 990 | 943 | 52 | 5 | 3 | 1 | 38 |
| 104_WestVic_3F4_B_UGT | 1111051 | 989 | 942 | 53 | 5 | 3 | 1 | 38 |
| 108_RayIs_3E7_B_UGT | 1111150 | 992 | 945 | 52 | 5 | 3 | 1 | 38 |
| 109_RayIs_3C8_B_UGT | 1111723 | 988 | 941 | 51 | 5 | 3 | 1 | 38 |
| 110_RayIs_3D8_B_UGT | 1109518 | 991 | 944 | 51 | 5 | 3 | 1 | 38 |
| 112_RayIs_3A9_B_UGT | 1111925 | 990 | 943 | 54 | 5 | 3 | 1 | 38 |
| 113_RayIs_3C9_B_UGT | 1111156 | 992 | 945 | 52 | 5 | 3 | 1 | 38 |
| 114_RayIs_3D9_B_UGT | 1110384 | 988 | 941 | 52 | 5 | 3 | 1 | 38 |
| 115_RayIs_3E9_B_UGT | 1111963 | 987 | 940 | 51 | 5 | 3 | 1 | 38 |
| 116_SWCoast_3H9_L_UGT | 1104801 | 989 | 942 | 53 | 5 | 3 | 1 | 38 |
| 117_RayIs_3A10_B_UGT | 1106527 | 988 | 941 | 51 | 5 | 3 | 1 | 38 |
| 118_RayIs_3G10_B_UGT | 1111606 | 991 | 944 | 51 | 5 | 3 | 1 | 38 |
| 121_MoPen_3F12_B_UGT | 1099576 | 983 | 935 | 51 | 5 | 3 | 1 | 39 |
| 122_SWCoast_4B5_B_UGT | 1110903 | 988 | 941 | 52 | 5 | 3 | 1 | 38 |
| 123_MoPen_4G5_C_UGT | 1097918 | 988 | 940 | 53 | 5 | 3 | 1 | 39 |
| 124_Gipps_4E6_C_UGT | 1108051 | 996 | 949 | 51 | 5 | 3 | 1 | 38 |
| 125_Gipps_4H6_F_UGT | 1108622 | 994 | 946 | 50 | 6 | 3 | 1 | 38 |
| 127_Gipps_4C10_C_UGT | 1106556 | 995 | 947 | 50 | 5 | 3 | 1 | 39 |
| 129_RayIs_5H6_B_UGT | 1113009 | 990 | 943 | 51 | 5 | 3 | 1 | 38 |
| 130_RayIs_7B7_B_UGT | 1111653 | 991 | 944 | 53 | 5 | 3 | 1 | 38 |
| 134_MoPen_15B5_C_UGT | 1106710 | 997 | 949 | 52 | 5 | 3 | 1 | 39 |
| 201_Belvedere_S1_UGT | 1106109 | 994 | 947 | 51 | 5 | 3 | 1 | 38 |
| 203_Bella_290513_UGT | 1101753 | 996 | 948 | 53 | 5 | 3 | 1 | 39 |
| 204_R1_UGT | 1110857 | 989 | 942 | 52 | 5 | 3 | 1 | 38 |
| 207_Savannah_S2_UGT | 1107062 | 994 | 947 | 52 | 5 | 3 | 1 | 38 |
| 208_Buddy_S1_Ocular | 1103941 | 996 | 948 | 52 | 5 | 3 | 1 | 39 |
| DBDeUG | 1092388 | 989 | 941 | 53 | 5 | 3 | 1 | 39 |
| IpTaLE | 1090473 | 986 | 938 | 53 | 5 | 3 | 1 | 39 |
| Mc_Marsbar | 1090694 | 989 | 941 | 53 | 5 | 3 | 1 | 39 |
| USC_Eleena_UGT | 1105218 | 992 | 945 | 52 | 5 | 3 | 1 | 38 |
| USC_Gun_koa1_UGT | 1102586 | 990 | 942 | 52 | 5 | 3 | 1 | 39 |
| USC_Haz_Bo_Eye | 1110477 | 996 | 948 | 52 | 5 | 3 | 1 | 39 |
| USC_Max_S1_UGT | 1104535 | 992 | 945 | 52 | 5 | 3 | 1 | 38 |
| USC_NSW_Adelle_LE | 1106363 | 996 | 949 | 53 | 5 | 3 | 1 | 38 |
| USC_NSW_Chingee_Eye | 1103238 | 991 | 943 | 52 | 5 | 3 | 1 | 39 |
| USC_NSW_Chingee_UGT | 1107075 | 998 | 950 | 51 | 5 | 3 | 1 | 39 |
| USC_NSW_Elmo_Eye | 1104984 | 991 | 944 | 50 | 5 | 3 | 1 | 38 |
| USC_NSW_Knox_Eye | 1100113 | 991 | 944 | 53 | 5 | 3 | 1 | 38 |
| USC_PM_11_UGT2 | 1111179 | 996 | 948 | 53 | 5 | 3 | 1 | 39 |
| USC_PM_13_UGT2 | 1110763 | 997 | 949 | 52 | 5 | 3 | 1 | 39 |
| USC_PM_17_UGT2 | 1109737 | 997 | 949 | 53 | 5 | 3 | 1 | 39 |
| USC_PM_3_UGT | 1106163 | 994 | 946 | 54 | 5 | 3 | 1 | 39 |
| USC_Posh_S1_Eye | 1103472 | 990 | 942 | 52 | 5 | 3 | 1 | 39 |
| USC_QLD_Bobby_UGT | 1110512 | 995 | 947 | 53 | 5 | 3 | 1 | 39 |
| USC_QLD_Chestnut_LE | 1103976 | 988 | 940 | 52 | 5 | 3 | 1 | 39 |
| USC_QLD_Chestnut_RE | 1110936 | 992 | 944 | 52 | 5 | 3 | 1 | 39 |
| USC_QLD_Chestnut_UGT | 1110967 | 990 | 942 | 50 | 5 | 3 | 1 | 39 |
| USC_QLD_Helen_UGT | 1100771 | 991 | 944 | 53 | 5 | 3 | 1 | 38 |
| USC_QLD_Jasper_LE | 1110193 | 994 | 946 | 52 | 5 | 3 | 1 | 39 |
| USC_QLD_Talle_UGT | 1110462 | 993 | 945 | 53 | 5 | 3 | 1 | 39 |
| USC_QLD_Travis_LE | 1107790 | 997 | 949 | 52 | 5 | 3 | 1 | 39 |
| USC_QLD_Travis_UGT | 1110311 | 992 | 944 | 53 | 5 | 3 | 1 | 39 |
| USC_RayIs_7-14_Eye | 1101643 | 1003 | 956 | 52 | 5 | 3 | 1 | 38 |
| USC_SA_12-220_LE | 1104912 | 988 | 940 | 52 | 5 | 3 | 1 | 39 |

| Genome | Genome length | Genes | Coding DNA sequence | Signal peptides | Miscellaneous RNA* | rRNA | tmRNA | tRNA |
|---|---|---|---|---|---|---|---|---|
| USC_SA_13-9_UGT | 1110787 | 990 | 943 | 51 | 5 | 3 | 1 | 38 |
| USC_SA_K2_UGT | 1102924 | 984 | 937 | 53 | 5 | 3 | 1 | 38 |
| USC_Ted_Hu_UGT | 1110363 | 995 | 947 | 53 | 5 | 3 | 1 | 39 |
| USC_Tya_Butler_LE | 1110933 | 993 | 945 | 53 | 5 | 3 | 1 | 39 |
| USC_Tya_Mavis_LE | 1109470 | 991 | 943 | 53 | 5 | 3 | 1 | 39 |

* Miscellaneous RNA features includes ‘small stable RNA A’

**Appendix 16.** Statistically significant results of gene presence/absence in relation to a binomial variable as assessed by Scoary (Brynildsrud *et al.* 2016). Presence or absence of genes was compared per *C. pecorum* genome in relation to geographical location (north/south), or swab sample location (urogenital tract (UGT)/ocular). Genes are clustered by Roary (Page *et al.* 2015), and assigned a cluster number if no annotation is derived (akin to 'hypothetical protein'). Scoary utilises a Fisher's exact test to compare binomial variables, and estimates population structuring by pairwise comparison.

| Gene cluster | Annotation/Comment[#] | Annotation | Binomial variable | | BH[*] adjust *P* value | Pairwise comparison *P* values[^] | |
|---|---|---|---|---|---|---|---|
| | | | North | South | | Best | Worst |
| group_138 | | hypothetical protein | 33/33 (100%) | 4/27 (15%) | < 0.001 | 1 | 1 |
| group_124 | group_270 | hypothetical protein | 0/33 (0%) | 23/27 (85%) | < 0.001 | 1 | 1 |
| group_270 | group_124 | hypothetical protein | 33/33 (100%) | 4/27 (15%) | < 0.001 | 1 | 1 |
| group_263 | group_269 | hypothetical protein | 4/33 (12%) | 23/27 (85%) | < 0.001 | 1 | 1 |
| group_269 | group_263 | hypothetical protein | 29/33 (88%) | 4/27 (15%) | < 0.001 | 1 | 1 |
| *pgk/tpi*_2 | Bifunctional PGK/TIM | Bifunctional PGK/TIM | 8/33 (24%) | 22/27 (81%) | < 0.001 | 0.5 | 1 |
| group_234 | | hypothetical protein | 15/33 (45%) | 1/27 (4%) | 0.002 | 0.5 | 1 |
| group_259 | Phospholipase D truncation | hypothetical protein | 15/33 (45%) | 1/27 (4%) | 0.002 | 0.5 | 1 |
| | | | **UGT** | **Ocular** | | | |
| group_234 | | hypothetical protein | 5/44 (11%) | 11/16 (69%) | 0.002 | 0.375 | 1 |
| group_263 | group_269 | hypothetical protein | 26/44 (59%) | 1/16 (6%) | 0.004 | 0.5 | 1 |
| group_269 | group_263 | hypothetical protein | 18/44 (41%) | 15/16 (94%) | 0.004 | 0.5 | 1 |
| group_138 | | hypothetical protein | 22/44 (50%) | 15/16 (94%) | 0.017 | 1 | 1 |
| group_124 | group_270 | hypothetical protein | 22/44 (50%) | 1/16 (6%) | 0.017 | 1 | 1 |
| group_270 | group_124 | hypothetical protein | 22/44 (50%) | 15/16 (94%) | 0.017 | 1 | 1 |

[#] 2 hypothetical proteins had <75% amino acid identity, and thus formed 4 clusters.

[*] *P* values adjusted for false discovery using the method described by Benjamini and Hochberg (1995)

[^] Best and worst case binomial *P* values should both be < 0.05 in cases where a variable's distribution is not impacted by population structure

**University Library**

**A gateway to Melbourne's research publications**

**Minerva Access is the Institutional Repository of The University of Melbourne**

**Author/s:**
Legione, Alistair Raymond

**Title:**
An investigation of two significant infectious diseases in populations of Victorian koalas (Phascolarctos cinereus)

**Date:**
2017

**Persistent Link:**
http://hdl.handle.net/11343/197532

**File Description:**
Thesis

**Terms and Conditions:**
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.