The University of Adelaide

DOCTORAL THESIS

INVESTIGATING THE INTERNAL AND EXTERNAL ECOLOGY OF SIX SUBTERRANEAN DIVING BEETLE SPECIES FROM THE YILGARN REGION OF CENTRAL AUSTRALIA

Josephine Charlotte Anne Hyde

A thesis submitted in the fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Australian Centre for Evolutionary Biology and Biodiversity in the School of Biological Sciences, Faculty of Sciences, The University of Adelaide

26th June 2018

TABLE OF CONTENTS

ABSTRACT

The ecology of subterranean ecosystems and stygofauna (subterranean aquatic animals) has largely been unexplored in an Australian context. The Yilgarn region of Western Australia is known as a biodiversity hotspot in relation to stygofauna from isolated calcrete aquifers, and it is home to the most diverse assemblage of subterranean, predatory diving beetles in the world. This study used extensive grids of boreholes to access calcrete aquifers at Sturt Meadows and Laverton Downs pastoral stations to investigate how subterranean species interact with their external and internal environment, focusing on six subterranean beetle species. A mix of traditional ecological monitoring and next-generation sequencing methods were employed to examine the following specific questions: What are the types of prey available in these calcrete systems and how do they change in abundance over time? What are the natural gut microbial communities associated with these predatory beetle species? Moreover, can metagenomic analyses be used to identify trophic differences among species, including adults and larvae, and determine whether beetle species eat other beetle species?

Ecological monitoring over an 11-year period identified that rainfall and, in particular, major recharge events are important for the diversity and distribution of stygofauna within the calcrete at Sturt Meadows. Average taxon richness was highest shortly after periods of high rainfall, and when dominant taxa (i.e. amphipods and copepods) were excluded, evenness decreased after both high and low rainfall suggesting that dominant taxa are an important factor driving the system. Common taxa (i.e. amphipods and copepods) within the calcrete had broad distributions and high abundance levels, while rare taxa (oligochaete worms) had restricted distributions and low abundances. All taxon groups had lower abundances and narrower distributions after periods of intermediate and low rainfall. Over the 11-year period, the majority of boreholes sampled did not show changes in evenness, suggesting that the Sturt Meadows calcrete is a reasonably stable ecosystem with episodic fluctuations, most likely attributed to rainfall events.

The gut microbiome was investigated in six beetle species from two separate aquifers using random shotgun sequencing (metagenomic analyses). The bacterial and viral communities were investigated separately, but the investigation showed similar results as follows: In both the viral and bacterial analyses the microbial communities

varied greatly by location and there was a distinct signature in the microbial communities depending on whether samples were collected from aquifers or laboratory aquaria. There were also distinct differences among the beetle species and their stage of development (adult versus larvae), which are most likely accounted for by trophic differences among the beetles. In both the bacterial and viral analyses a large number of the sequences were novel and unable to be identified, suggesting major differences in their microbiome compared to previously studied invertebrates. The unknown sequences will once identified, provide further insights into the microbial communities of these subterranean environments. Like the bacteria, viral sequences provided evidence that calcretes had been influenced by anthropogenic activities on the surface, with a large number of vertebrate viruses infiltrating the calcrete system.

The mitochondrial (mt) genomes of four of the beetle species, *Limbodessus palmulaoides* and *P. macrosturtensis*, *P. mesosturtensis* and *P. microsturtensis,* were characterised to provide a framework for future trophic analyses of the beetle gut contents and as a basis for further assessment of the molecular evolution of mtDNA genes associated with evolution underground. The mt genomes were all consistent with both previously sequenced dytiscid beetle mt genomes and the inferred ancestral insect mt genome. All four mt genomes were circular, contained the expected 37 genes and ranged from 16,504 to 16,868 bp. The overall structure (gene number, orientation and order) was consistent with the ancestral insect mt genome, and the genome size variation resulting from length variation of intergenic regions and the CR is consistent with other surface dytiscid species sequenced.

These ecological and molecular analyses show a complex interconnected system between the surface and subterranean environments. They also demonstrate that metagenomics research can be used effectively for investigating the trophic ecology of species, particularly in taxa where traditional methods are ineffective or difficult to undertake. Preliminary analyses of the beetle metagenomes suggested that the beetle species at Sturt Meadows are not only eating other invertebrates from the calcrete but are also eating the other beetle species. This result may have been difficult to elucidate using traditional methods (e.g. metabarcoding) given how closely related the beetle species are to each other. This research also highlights that the subterranean beetles provide an excellent model system, not only for future microbiome work but also for

investigating the adaptive and regressive evolution of the genome associated with moving from surface to underground habitats.

THESIS DECLARATION

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed: **Date:** 7/11/18

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my three supervisors. Prof. Steve Cooper, Prof. Andy Austin and Assoc. Prof. Pablo Munguia for their continuous support of my PhD study, for their patience, motivation, time, funding and immense knowledge. I want to thank all three of them for staying with me when this project took such a turn outside their areas of expertise. I would also like to thank Steve for giving me such an excellent assignment in the third year of my undergraduate study and subsequently introducing me to the world of stygofauna.

Besides my supervisors, I would like to thank the rest of my co-authors: Prof. Bill Humphreys, Prof. Jim Mitchell, and Dr Lisa Dann, for their insightful comments and encouragement, but also for asking the hard questions, which led me to widen my research out from my comfort zone.

The members of the Austin group have contributed immensely to my time at Adelaide University both personally and professionally. The group has been a source of friendships as well as good advice and collaboration. I am especially grateful for the fun group of lab members who have stuck out this PhD with me: Danielle Stringer, Amelia Lewis, Erinn Fagan-Jeffries, Barbara Langille, Michelle Guzik and Gary Taylor. Other past and present group members that I have had the pleasure to work with or alongside, positively contributed to my time here and I look forward to the next conference when I can see you again. I would like to acknowledge friends of the lab Ray Chatterji, Jaimi Grey and Marc Jones for your support and for showing me the wonders of R. I would also particularly like to thank Ben Parslow; your support and faith in me as a scientist has been invaluable and I only hope one day to be as good a scientist as you already think I am. I would also like to thank my fellow lab mates in the Munguia and Mitchell labs, past and present. Without their support and encouragement, it would not have been possible for me to conduct this research.

I gratefully acknowledge the funding sources that made my PhD work possible. I was funded by an Australian Postgraduate Award for the first 3 ½ years. My work was also supported by the Australian Research Council, with industry partners Newmont Australia, Placer Dome Asia Pacific, Minara Resources Limited, the Western Australian Museum, the South Australian Museum, Rio Tinto, Biota Environmental Sciences, Bennelongia, Department of Biodiversity, Conservation and Attractions (WA).

My time at Adelaide was made enjoyable in large part due to the many friends and groups that became a part of my life. I am grateful for all the people I have spent the last four years with, for my long-suffering housemates Vi and Susie, thank you for the wine, housemate dinners and introducing me to the wonders of The Bachelor. I am also grateful for my friends and gymnastics and rugby teammates and countless other people and memories.

I want to thank Sandra Eustace and Alison Bullock for introducing me to science and particularly genetics all those years ago. Thank you both so much for showing me what I could be and for always talking about my future in terms of when and never if. If I could be half as inspirational as you both were I think I will have achieved greatness.

I would like to thank Bonnie, for being unfailingly the only one who is always happy to see me, and for keeping my feet warm during long days writing.

Lastly, I would like to thank my family for all their support, love and encouragement. For my parents, sister and brother who have supported me in all my pursuits and my parents especially for their faithful support during the final stages of this PhD it is so appreciated.

Thank you, everyone, for giving me shoulders to stand on

CHAPTER 1: GENERAL INTRODUCTION

Since the 1980s there has been a widespread exploration of obligate subterranean aquatic animals (stygofauna) in Australia with a specific focus on the discovery and identification of new species (Humphreys, 2001; Boulton *et al*., 2003). However, the ecology of these subterranean ecosystems and how their species interact with each other and their environment is still under-examined and poorly understood. This lack of research is significant because as stygofauna live permanently in groundwater, they can be used as indicator species for the health of the ecosystem. However, their use as bioindicators is limited if we do not understand these organisms or how they fully interact with their environments. Additionally, recent studies suggest that animals should not be considered in isolation, but within the context of their microbiome if you want to completely understand the total organism (Bäckhed *et al*., 2004; Gill *et al*., 2006; Prosdocimi *et al*., 2015). To address these knowledge gaps, a study was conducted investigating the external and internal ecology of six stygobiontic diving beetles from two groundwater ecosystems in the arid region of central Australia. This introductory chapter outlines the broad placement of stygofaunal beetles from the genera *Limbodessus* and *Paroster* within the context of Australian groundwater dependent ecosystems and the current state of microbiome research and sets out the aims of the study.

Groundwater Dependent Ecosystems (GDEs)

With 97% of the world's unfrozen freshwater located underground, groundwater and its dependent ecosystems are valuable natural resources (Murray *et al*., 2003). Groundwater is vital in that it maintains complex communities of plants and animals and the services and ecological processes that they support and provide (Clifton *et al*., 2007; Tomlinson, 2011). Groundwater itself also typically contains active communities comprised of microorganisms, invertebrates, and in some systems vertebrates (Gibert *et al*., 1994; Humphreys, 2006). GDEs are considered to be open systems as both water and energy flow through them (Boulton, 2005).

In Australia, GDEs can be classified broadly into three types: subterranean, surface aquatic and terrestrial ecosystems. Subterranean GDEs include aquifers, caves (including those that are submerged, wet or containing streams), wet passages in karst, pseudokarst, calcretes, and fractured rock, as well as water that fills the interstitial

spaces between sediment in alluvial aquifers (Hancock *et al*., 2010). These ecosystems are entirely dependent on groundwater, and some have called them the ultimate GDE (Humphreys, 2006). Surface aquatic ecosystems require a connection to the aquifer to maintain their water supply, for at least part if not all of the year. Surface aquatic ecosystems include spring lakes and other wetlands such as the mound springs of the Australian Great Artesian Basin (Ponder, 1986; Murray *et al*., 2003; Hancock *et al*., 2010). Terrestrial ecosystems are where plant roots extend deep enough to extract the groundwater (Hancock *et al*., 2010). Terrestrial ecosystems can have a seasonal dependence on groundwater or reliance that is episodic (Murray *et al*., 2003). Additionally, marine ecosystems that are near the shore and experience regular groundwater discharge or estuarine ecosystems are also dependent on groundwater (Murray *et al*., 2003).

Cave and aquifer ecosystems are mainly inhabited by stygobites (obligate aquatic and subterranean organisms) and are examples of GDEs that have an obligate dependency on the groundwater. Due to this dependency on groundwater, the fauna often remains *in situ* and are endemic to a particular aquifer (Humphreys, 2006). The biological characteristics that contribute to this restricted distribution are related to morphologies exhibited due to the perpetual darkness that is present in subterranean habitats. These morphologies include a loss or reduction of eyes, loss of pigmentation and, in insects, a loss of wings (Christiansen, 2012). Stygobites are highly adapted to their environment and often have no dispersal or resting stages (Humphreys, 2006). The fact that they generally produce few offspring, further decreases their dispersal potential (Humphreys, 2006).

Subterranean groundwater systems in Australia

Prior to the 1990s, it was believed that Australia lacked any subterranean diversity. It has since been discovered that the arid zone and the wet tropics in Australia contain a rich diversity of subterranean fauna (Culver and Sket, 2000; Humphreys, 2012). While the initial research was focused on finding fauna in karstic areas, the discovery of subterranean fauna in non-karstic substrates has led to the rapid expansion and discovery of many new species. However, detailed examination of the subterranean fauna in Australia has been sparse, and it has been restricted, to a great extent, to faunal surveys, mostly in Western Australia (Guzik *et al*., 2010). These faunal surveys have been driven largely by the mining boom and regulations requiring the inclusion of

subterranean fauna in any environmental review processes carried out for major resource projects in Western Australia (EPA, 2003).

The Yilgarn Region of Western Australia

In particular, the Yilgarn Region of Western Australia was found to contain a diverse range of stygofauna, which was first discovered in 1998 (Humphreys *et al*., 2009). The region contains over 200 isolated calcretes that are thought to have formed from groundwater as it approaches base level close to playas (salt lakes) in palaeodrainage systems (Humphreys, 2001) between 10 – 30 MYA (Morgan, 1993); the major calcrete bodies often have an area of greater than 100 km^2. There are also 100s of smaller calcrete bodies, some with an area of less than a few square kilometres in size (Bradford *et al*., 2010). The entire area resembles a subterranean archipelago and each calcrete, to date, that contains fauna has been found to harbour a unique combination of aquatic invertebrate species including predaceous dytiscid diving beetles (Coleoptera), crustaceans (Syncarida, Isopoda, Amphipoda, Copepoda, Ostracoda and Bathynellacea) and worms (Oligochaeta) (Watts and Humphreys, 1999; Humphreys, 2001; Karanovic, 2004; Guzik *et al*., 2008; Guzik *et al*., 2010). Multiple morphological and phylogeographical studies have shown that stygofaunal species have a distribution that is restricted to a single calcrete and do not occur in the surrounding matrix (Cooper *et al*., 2002). These studies suggest that individual calcretes are biologically isolated with respect to their stygofauna (Humphreys *et al*., 2009). The calcretes are also known to contain the most diverse group of subterranean diving beetles in the world (by a factor of 10) with over 100 species described so far (Balke *et al*., 2004).

Exemplar calcrete aquifers: Sturt Meadows and Laverton Downs

Sturt Meadows calcrete (SMC) is located on the Sturt Meadows pastoral station within the Yilgarn region, located around 50 km from Leonora, in central Western Australia (chapter 2, Fig. 2.1). SMC is approximately 43 km^2, and the surface habitat contains open Acacia woodland with lowland shrubs. In places, the calcrete is exposed on the surface while at other points the top of this calcrete is up to 2 m below the surface (and it is up to 10 m thick) (Allford *et al*., 2008; Bradford *et al*., 2013). The calcrete formation is typical of the area, in that it has formed within a palaeodrainage channel (Morgan, 1993), and the calcrete is close to (7 km north) a salt lake, Lake Raeside (Allford *et al*., 2008). SMC has previously been used for mineral exploration and, as a result, an area (3.5 km^2), consisting of two continuous grids containing

boreholes, was established. This site allows for extensive sampling of the subterranean ecosystem below (Allford *et al*., 2008; Bradford *et al*., 2013). The bores were drilled in 2001 to an average depth of 10.3 m and a diameter of 100 mm. Due to the condition of the bores, the accessible amount of water table varies from 1.9 to 4.0 m in depth below the surface, and the water depth is between 0.4 and 8 m (Allford *et al*., 2008).

The initial ecological surveys identified SMC as a simple ecosystem containing seven stygobiont macro-invertebrate species (Bradford *et al*., 2010). However, a mitochondrial Cytochrome C Oxidase I (*COI*) DNA analysis of the SMC identified a slightly more complex system (Bradford *et al*., 2010). In this calcrete, there are a triplet of sister species of dytiscid diving beetles *Paroster macrosturtensis* (large), *Paroster mesosturtensis* (medium), and *Paroster microsturtensis* (small) (3.3, 2.4 and 1.8 mm, respectively) (Watts and Humphreys, 2006; Leijs *et al*., 2012). While it was first assumed that there was only one species of amphipod, subsequent DNA and morphological analyses have now shown that there are at least three different species of chiltoniid amphipods (*Yilgarniella sturtensis*, *Scutachiltonia axfordi* and *Stygochiltonia bradfordae* (Bradford *et al*., 2010; King *et al*., 2012b). There are also at least two harpactacoid and one cyclopoid species of copepod and four divergent lineages of oligochaete worms (Bradford *et al*., 2010). In addition, there are troglobionts present above the water table: at least one mite species (Acari), one centipede species (Chilopoda), one palpigrade species (*Eukoenenia guzikae*) and one springtail (Collembola) species and multiple species of troglobiont isopods (Barranco and Harvey, 2008; Bradford *et al*., 2010; Javidkar *et al*., 2016).

The Laverton Downs calcrete (LDC) is located 135 km north-east of the SMC, near the town of Laverton, and the site is approximately 90 km^2. LDC lies on the Carey palaeodrainage channel (east) (Guzik *et al*., 2011). There are four main collection sites within the calcrete, referred to as Laverton South, Shady Well, Quandong Well and Erlistoun. Laverton South and Shady Well are approximately 9.5 km apart and are sites containing grids of ~20 mineral exploration boreholes 50-100 m apart (Guzik *et al*., 2011). Quandong is a site 6.5 km further north from Shady Well where there is an old pastoral well. A bore-hole grid (Erlistoun grid) has been recently discovered part way between Shady Well and Laverton South (Humphreys and Cooper pers. comm.).

There are three dytiscid beetle species present in the LDC. However, there is evidence that they are not sister species, with phylogenetic analyses suggesting that

their surface ancestors were distantly related and, therefore, likely to have been reproductively isolated before they colonised the calcrete (Leys *et al*., 2003; Leijs *et al*., 2012). The beetle species are *Limbodessus lapostaae*, *L. windarraensis* and *L. palmulaoides* (Watts and Humphreys, 1999, 2006), although they still do exhibit size variation with *L. lapostaae* the smallest, *L. windarraensis* of medium size and *L. palmulaoides* the largest species (1.3, 2.2 and 4.2mm, respectively) (Guzik *et al*., 2011). There is also evidence for one chiltoniid amphipod species and several lineages of *Haloniscus* isopods in the calcrete. The *Haloniscus* isopods were studied by Cooper *et al*. (2008) using *COI* sequence data, which led to the conclusion that the calcrete contained multiple divergent (~31%) mtDNA lineages and at least two different sister clades, both of which are localised in the north, with an additional one found in the south (Cooper *et al*., 2008) (S. Cooper, pers. comm.).

Speciation within calcrete aquifers

Patterns of colonisation within the calcretes are not entirely clear with two main theories proposed to explain the evolution of subterranean species; the climate relict hypothesis (Barr, 1968; Banarescu, 1975; Sbordoni, 1982; Barr Jr and Holsinger, 1985; Peck and Finston, 1993) and the adaptive shift hypothesis (Rouch and Danielopol, 1987; Desutter-Grandcolas and Grandcolas, 1996). In the case of the subterranean beetles, molecular studies suggest that the climate relict hypothesis is most likely, where epigean (surface) species that are already pre-adapted to subterranean life (such as those that live under stones in a stream) may adapt rapidly to subterranean life when the surface environment becomes unstable (Leys *et al*., 2003). Phylogenetic analyses, based on mtDNA sequence data, have identified 13 cases of calcretes containing two to three sympatric sister species of beetles of non-overlapping sizes (Leijs *et al*., 2012). Leijs *et al*. (2012) has proposed several hypotheses to explain the beetles' modes of speciation, including allopatric, sympatric, parapatric and/or microallopatric speciation. Simulations of colonisation events indicate that it is unlikely that the sister species evolved following repeated colonisation by the same ancestral species (allopatric speciation) due to the high frequency of sister pairs and triplets. Acceptance of this hypothesis would mean that a large number of individual calcretes would have had two or three colonisation events by the same species. This pattern was found to be better explained with species diversification occurring within the calcretes (Leijs *et al*., 2012). One factor that might have driven intra-calcrete speciation is by the beetles' divergence

into different ecological niches. In particular, it is possible that the different sized beetles are in different trophic niches (Cooper *et al*., 2002; Leijs *et al*., 2012). The microallopatric speciation hypothesis suggests that potentially, in the case of linear calcretes, colonisation may have occurred at either end or at distant sites by the same species. It is also possible that due to fragmentation within the calcretes, beetles could become separated in space over long periods giving them time to become reproductively and genetically isolated. This separation would cause them to form different lineages when the populations were re-joined (Leijs *et al*., 2012).

Molecular clock analysis of the beetles suggests that there has been no apparent gene flow between the calcretes since the late Miocene to Pliocene (10-5 MYA) during which time there was an overall trend towards aridity in the region (Byrne *et al*., 2008). It has also been estimated, using molecular clock techniques, that amphipod populations in the Yilgarn have been isolated since the late Miocene or the Pliocene (14.6 – 4.1 MYA) (Cooper *et al*., 2007). This theory fits with both the previous existing beetle data and the geological models for the region (Morgan, 1993; Cooper *et al*., 2007). Using the rate of evolution of *COI* for subterranean isopods, the colonisation of the calcretes and subsequent isolation of isopod populations within calcretes has been estimated to have occurred during the mid to late Miocene. This date is potentially slightly earlier than the colonisation and speciation estimation events for both the amphipods and beetles, but still fits the geological history of the area with aridity driving the isolation of calcrete populations (Leys *et al*., 2003).

Diet is one of the primary drivers of population abundance, and controls critical traits such as body condition and home range size, and influences an individual's activity budget. It is constrained by both environmental and social factors (Quéméré *et al*., 2013). In determining the food web of a subterranean system, a large number of unknowns must be addressed, including where the energy that is driving the system is coming from, what is the basis of the food chain, and how different species interact with each other. A large number of these unknowns can be addressed by investigating an animal's microbiome and its genes.

Microbiome research

Gut microbiome research was first initiated by Louis Pasteur and Elie Metchnikoff more than a century ago, with the dawn of microbiology. Pasteur, while most famous for his significant contributions to germ theory of disease (Pasteur *et al*.,

1878), was the first person to demonstrate that the deregulation of the gut microbiome can lead to a disease state and that the gut microbiome can be altered by environmental changes (Pasteur, 1870, 1885). His experiments led Pasteur to suggest that microbes have an essential role in life and that life under axenic conditions would be impossible. One of Pasteur's colleagues, Metchnikoff, studied the role of microbes in nutrition and digestion, and the dual positive and negative roles that intestinal flora can have on an individual (Metchnikoff, 1901). However, with the advent of the First World War and the discovery of antibiotics, this research was discontinued and was not renewed until the 1970's (Socransky and Manganiello, 1971; Tannock and Savage, 1974; Eutick *et al*., 1978; Lee and Brey, 2013). This new interest remained mostly descriptive as a large percentage of the gut microbiome cannot be cultivated.

With the advent of recent technological advances, initially 16S sequencing and more recently next-generation sequencing and multi-omics technology including transcriptomics, proteomics and metabolomics, the field of microbiology has been revolutionised (Lee and Brey, 2013). This expansion can be seen particularly in human microbiome research, including such large-scale efforts as the Human Microbiome Project launched in 2007 (Turnbaugh *et al*., 2007; Peterson *et al*., 2009). As the gut contains the densest and most diverse microbiome, animal studies involving the gut microbiome have most captivated the interest of researchers. The gut microbiome is influenced by a wide range of intrinsic and extrinsic variables and has been found vital to the health and development of organisms.

The gut microbiome is a complex community and is comprised of bacteria, archaea, fungi, protozoa, and viruses, with viruses the dominant part of the community. Microbial communities play key roles in the host's fitness including the host's metabolism, fecundity, immunity, and longevity (Dillon and Dillon, 2004; Wang *et al*., 2011). It has been suggested that organisms should no longer be considered in isolation, but as part of larger holobionts with their microbiome. The gut microbiome composition reflects the natural selection of microbes and hosts, with the gut microbiome biased towards mutual cooperation and stability within this complex ecosystem (Dillon and Dillon 2004).

In recent years, non-human animal microbiomes have started to be explored and can be broadly classified into three main categories; domestic animals, model animals and wild animals. The majority of published studies are dominated by domestic

animals, followed by model animal, with comparatively few wild animals studied to date. Studies where the microbiome or host organism is modified (perturbation studies) are used heavily in domestic and model animal studies compared to observational studies. The opposite is true of wild animal studies where perturbation studies are rarely used (Pascoe *et al*., 2017). Perturbation studies are used to explore interactions between the gut microbiome and host health (Brinkman *et al*., 2011) or to improve animal productivity, in the case of domestic animals (Ahmed *et al*., 2014). Direct observation studies, which are the norm for wild animal studies, are used in order to characterise 'natural' microbiome community structure and function (Degli Esposti and Romero, 2017).

Studies on model animals are fundamental in furthering our understanding of gut microbiome structure, function and modulation. Previously, invertebrate and vertebrate models have been used to investigate the host-virome relationship at varying levels of microbiome complexity due to varying levels of experimental control (Grover and Kashyap 2014; Lee and Brey 2013; Ma *et al*., 2015). By using these models, various important factors, such as those that enable symbiosis have been discovered and have enabled an improved understanding of the human microbiome (Kostic *et al*., 2013). Key model systems for gut microbiomes of increasing complexity include the Hawaiian bobtail squid (*Euprymna scolopes*), the fruit fly (*Drosophila melanogaster*), the Zebrafish (*Danio rerio*), and the mouse (*Mus musculus*) (Kostic *et al*., 2013). In a review of model organism microbiome studies published between November 2006 and July 2016, 93% were composed of vertebrate studies (Pascoe *et al*., 2017), which presents limitations in extrapolating results to invertebrate species, particularly beetles. Other limitations include the fact that model organisms are often highly inbred and lack the genetic diversity and experience less environmental variability than their wild counterparts. Analyses suggest that the microbiomes of laboratory-raised organisms do not necessarily reflect those in wild-caught individuals (Chandler *et al*., 2011; Rosshart *et al*., 2017). Analyses on microbiomes of subterranean animals is virtually limited to a single vertebrate species, the naked mole rat (Debebe *et al*., 2017), while research on aquatic invertebrates is limited to marine oyster (King *et al*., 2012a).

Given the importance of insects for both economical and anthropogenic reasons, more research on their microbiomes will be important going forward; this study hopes

to contribute to knowledge of the ecology and genetics of species of dytiscid diving beetles with emphasis on their microbiomes, including their viromes.

My study aimed to provide further knowledge of the groundwater ecosystems inhabited by subterranean dytiscids, with an emphasis on their associated microbial communities and trophic position in the food web of subterranean aquifers.

The specific aims were to:

Characterise the diversity patterns of the subterranean invertebrate fauna in the Sturt Meadows calcrete with a particular focus on the role of rainfall in influencing diversity patterns using observational data (Chapter 2)

Characterise the microbiome, including bacteria and viruses, of six species of subterranean diving beetles from the Sturt Meadows and Laverton calcretes using metagenomic analyses (Chapters 3 and Chapter 4)

Characterise and conduct comparative analyses of complete mitochondrial genomes from four species of subterranean diving beetles from the Sturt Meadows and Laverton calcretes using molecular data (Chapter 5).

The results chapters comprising this thesis have been formatted as journal papers, with chapters 2 and 5 already published, and chapters 3 and 4 ready to submit. For this reason, there is some repetition among the chapters, particularly the information presented in the introductions and reference lists. This could not be avoided, but an attempt has been made to limit such repetition given the focus of each chapter.

In the last section (Chapter 6) a general discussion is presented on the broader implications of the study as well as the limitations and ideas for future research.

References

Ahmed, H.A., Sirohi, S.K., Dagar, S.S., Puniya, A.K., and Singh, N. (2014) Effect of supplementation of *Selenomonas ruminantium* NDRI-PAPB 4 as direct fed microbial on rumen microbial population in Karan Fries male calves. *Indian Journal of Animal Nutrition* 31: 20-26.

Allford, A., Cooper, S.J.B., Humphreys, W.F., and Austin, A.D. (2008) Diversity and distribution of groundwater fauna in a calcrete aquifer: does sampling method influence the story? *Invertebrate Systematics* 22: 127-138.

Bäckhed, F., Ding, H., Wang, T., Hooper, L.V., Koh, G.Y., Nagy, A. *et al*. (2004) The gut microbiota as an environmental factor that regulates fat storage. *Proceedings of the National Academy of Sciences of the United States of America* 101: 15718-15723.

Balke, M., Watts, C.H.S., Cooper, S.J.B., Humphreys, W.F., and Vogler, A.P. (2004) A highly modified stygobiont diving beetle of the genus *Copelatus* (Coleoptera, Dytiscidae): taxonomy and cladistic analysis based on mitochondrial DNA sequences. *Systematic Entomology* 29: 59-67.

Banarescu, P. (1975) Principles and problems of zoogeography: National Technical Information Service. Belgrade

Barr Jr, T.C., and Holsinger, J.R. (1985) Speciation in cave faunas. *Annual Review of Ecology and Systematics* 16: 313-337.

Barr, T.C. (1968) Cave ecology and the evolution of troglobites. (eds Dobzhansky, T., Hecht, M.K., Steere, W.C.) *Evolutionary Biology* (pp. 35-102) Springer, Boston MA.

Barranco, P., and Harvey, M.S. (2008) The first indigenous palpigrade from Australia: a new species of Eukoenenia (Palpigradi: Eukoeneniidae). *Invertebrate Systematics* 22: 227-233.

Boulton, A.J. (2005) Chances and challenges in the conservation of groundwaters and their dependent ecosystems. *Aquatic Conservation: Marine and Freshwater Ecosystems* 15: 319-323.

Boulton, A.J., Humphreys, W.F., and Eberhard, S.M. (2003) Imperilled Subsurface Waters in Australia: Biodiversity, Threatening Processes and Conservation. *Aquatic Ecosystem Health and Management* 6: 41-54.

Bradford, T.M., Adams, M., Humphreys, W.F., Austin, A.D., and Cooper, S.J.B. (2010) DNA barcoding of stygofauna uncovers cryptic amphipod diversity in a calcrete aquifer in Western Australia's arid zone. *Molecular Ecology Resources* 10: 41-50.

Bradford, T.M., Adams, M., Guzik, M.T., Humphreys, W.F., Austin, A.D., and Cooper, S.J.B. (2013) Patterns of population genetic variation in sympatric chiltoniid amphipods within a calcrete aquifer reveal a dynamic subterranean environment. *Heredity* 111: 77-85.

Brinkman, B.M., Hildebrand, F., Kubica, M., Goosens, D., Del Favero, J., Declercq, W. *et al*. (2011) Caspase deficiency alters the murine gut microbiome. *Cell Death and Disease* 2: e220.

Byrne, M., Yeates, D., Joseph, L., Kearney, M., Bowler, J., Williams, M. *et al*. (2008) Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. *Molecular Ecology* 17: 4398-4417.

Christiansen, K. (2012) Morphological adaptations. (eds White W.B., Culver D.C.) *Encyclopedia of Caves* (Second Edition) (pp. 517-528) Elsevier, Washington DC.

Clifton, C., Cossens, B., McAuley, C., Evans, R., Cook, P., Howe, P., and Boulton, A. (2007) A framework for assessing the environmental water requirements of groundwater dependent ecosystems. Report

(https://publications.csiro.au/rpr/download?pid=procite:f576497c-4cf3-42f9-a6b6-befa9048137e&dsid=DS1)

Cooper, S.J.B., Hinze, S., Leys, R., Watts, C.H.S., and Humphreys, W.F. (2002) Islands under the desert: molecular systematics and evolutionary origins of stygobitic water beetles (Coleoptera: Dytiscidae) from central Western Australia. *Invertebrate Systematics* 16: 589-590.

Cooper, S.J.B, Bradbury, J.H., Saint, K.M., Leys, R., Austin, A.D., and Humphreys, W.F. (2007) Subterranean archipelago in the Australian arid zone: mitochondrial

DNA phylogeography of amphipods from central Western Australia. *Molecular Ecology* 16: 1533-1544.

Cooper, S.J.B., Saint, K.M., Taiti, S., Austin, A.D., and Humphreys, W.F. (2008) Subterranean archipelago: mitochondrial DNA phylogeography of stygobitic isopods (Oniscidea: *Haloniscus*) from the Yilgarn region of Western Australia. *Invertebrate Systematics* 22: 195-203.

Culver, D.C., and Sket, B. (2000) Hotspots of subterranean biodiversity in caves and wells. *Journal of Cave and Karst Studies* 62: 11-17.

Debebe, T., Biagi, E., Soverini, M., Holtze, S., Hildebrandt, T.B., Birkemeyer, C. *et al*. (2017) Unraveling the gut microbiome of the long-lived naked mole-rat. *Scientific Reports* 7: 9590.

Degli Esposti, M., and Romero, E.M. (2017) The functional microbiome of arthropods. *PLoS One* 12: e0176573.

Desutter-Grandcolas, L., and Grandcolas, P. (1996) The evolution toward troglobitic life: a phylogenetic reappraisal of climatic relict and local habitat shift hypotheses. *Mémoires de Biospéologie* 23: 57-63.

Dillon, R., and Dillon, V. (2004) The gut bacteria of insects: nonpathogenic interactions. *Annual Reviews in Entomology* 49: 71-92.

EPA (2003) Consideration of subterranean fauna in groundwater and caves during environmental impact assessment in Western Australia. Guidance for the assessment of environmental factors (in accordance with the Environmental Protection Act 1986): 1-12. (http://www.epa.wa.gov.au/sites/default/files/Policies_and_Guidance/Technical%20Guidance-Subterranean%20fauna-Dec2016.pdf)

Eutick, M.L., O'Brien, R.W., and Slaytor, M. (1978) Bacteria from the gut of Australian termites. *Applied and Environmental Microbiology*35: 823-828.

Gibert, J., Danielopol, D., and Stanford, J.A. (Eds.). (1994) *Groundwater Ecology*: (Vol. 1). Academic Press. New York NY

Gill, S.R., Pop, M., DeBoy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S. *et al*. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* 312: 1355-1359.

Guzik, M.T., Abrams, K.M., Cooper, S.J.B., Humphreys, W.F., Cho, J.L., and Austin, A.D. (2008) Phylogeography of the ancient Parabathynellidae (Crustacea: Bathynellacea) from the Yilgarn region of Western Australia. *Invertebrate Systematics* 22: 205-216.

Guzik, M.T., Cooper, S.J.B., Humphreys, W.F., Ong, S., Kawakami, T., and Austin, A.D. (2011) Evidence for population fragmentation within a subterranean aquatic habitat in the Western Australian desert. *Heredity* 107: 215-230.

Guzik, M.T., Austin, A.D., Cooper, S.J.B., Harvey, M.S., Humphreys, W.F., Bradford, T.M. *et al*. (2010) Is the Australian subterranean fauna uniquely diverse? *Invertebrate Systematics* 24: 407-418.

Nevill, J. C., Hancock, P. J., Murray, B. R., Ponder, W. F., Humphreys, W. F., Phillips, M. L., and Groom, P. K. (2010). Groundwater-dependent ecosystems and the dangers of groundwater overdraft: a review and an Australian perspective. *Pacific Conservation Biology*, *16*(3), 187-208.

Humphreys, W.F. (2001) Groundwater calcrete aquifers in the Australian arid zone: the context to an unfolding plethora of stygal biodiversity. *Records of the Western Australian Museum*, Supplement 64: 63-83.

Humphreys, W.F. (2012) Diversity patterns in Australia. (eds White W.B., Culver D.C.) *Encyclopedia of Caves* (Second Edition): (pp. 203-219) Elsevier, Washington DC pp. 517-528.

Humphreys, W.F., Watts, C.H.S., Cooper, S.J.B., and Leijs, R. (2009) Groundwater estuaries of salt lakes: buried pools of endemic biodiversity on the western plateau, Australia. *Hydrobiologia* 626: 79-95.

Humphreys, W.F. (2006) Aquifers: the ultimate groundwater-dependent ecosystems. *Australian Journal of Botany* 54: 115-132.

Javidkar, M., Cooper, S.J.B., King, R.A., Humphreys, W.F., Bertozzi, T., Stevens, M.I., and Austin, A.D. (2016) Molecular systematics and biodiversity of oniscidean isopods in the groundwater calcretes of central Western Australia. *Molecular Phylogenetics and Evolution* 104: 83-98.

Karanovic, T. (2004) Subterranean copepods (Crustacea: Copepoda) from arid Western Australia. *Crustaceana* Supplement 70: 1-366.

King, G.M., Judd, C., Kuske, C.R., and Smith, C. (2012a) Analysis of stomach and gut microbiomes of the eastern oyster (*Crassostrea virginica*) from coastal Louisiana, USA. *PLoS One* 7: e51475.

King, R.A., Bradford, T.M., Austin, A.D., Humphreys, W.F., and Cooper, S.J.B (2012b) Divergent molecular lineages and not-so-cryptic species: the first descriptions of stygobitic chiltoniid amphipods (Talitroidea: Chiltoniidae) from Western Australia. *Journal of Crustacean Biology* 32: 465-488.

Lee, W.J., and Brey, P.T. (2013) How microbiomes influence metazoan development: insights from history and *Drosophila* modeling of gut-microbe interactions. *Annual Review of Cell and Developmental Biology* 29: 571-592.

Leijs, R., Van Nes, E.H., Watts, C.H.S., Cooper, S.J.B, Humphreys, W.F., and Hogendoorn, K. (2012) Evolution of blind beetles in isolated aquifers: a test of alternative modes of speciation. *PLoS One* 7: e34260.

Leys, R., Watts, C.H.S., Cooper, S.J.B., and Humphreys, W.F. (2003) Evolution of subterranean diving beetles (Coleoptera: Dytiscidae: Hydroporini, Bidessini) in the arid zone of Australia. *Evolution* 57: 2819-2834.

Metchnikoff, E. (1901) L'immunité dans les maladies infectieuses. Masson. Paris

Morgan, K. (1993) Development, sedimentation and economic potential of palaeoriver systems of the Yilgarn Craton of Western Australia. *Sedimentary Geology* 85: 637-656.

Murray, B.B.R., Zeppel, M.J., Hose, G.C., and Eamus, D. (2003) Groundwater-dependent ecosystems in Australia: It's more than just water for rivers. *Ecological Management and Restoration* 4: 110-113.

Pascoe, E.L., Hauffe, H.C., Marchesi, J.R., and Perkins, S.E. (2017) Network analysis of gut microbiota literature: an overview of the research landscape in non-human animal studies. The ISME Journal 11: 2644.

Pasteur, L. (1870) Etudes sur la maladie des vers a soie: Gauthier-Villars. Paris

Pasteur, L. (1885) Methode pour prevenir la rage apres morsure. *Comptes Rendus de l'Académie des Sciences,* 101

Pasteur, L., Joubert, L., and Chamberland, C. E. (1878). La théorie des germes et ses applications à la medicine et à la chirurgie. *Comptes Rendus de l'Académie des Sciences,* 86: 1037–43. Peck, S., and Finston, T. (1993) Galapagos islands troglobites: the questions of tropical troglobites, parapatric distributions with eyed-sister-species, and their origin by parapatric speciation. *Mémoires de Biospéologie* 20: 19-37.

Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J.A. *et al*. (2009) The NIH human microbiome project. *Genome Research* 19: 2317-2323.

Ponder, W. (1986) Mound springs of the great artesian basin. (eds De Deckker, P., Williams, W.D.) *Limnology in Australia* (pp. 403-420) SpringerMelbourne VIC.

Prosdocimi, E.M., Mapelli, F., Gonella, E., Borin, S., and Crotti, E. (2015) Microbial ecology-based methods to characterize the bacterial communities of non-model insects. *Journal of Microbiological* Methods 119: 110-125.

Rouch, R., and Danielopol, D. (1987) L'origine de la faune aquatique souterraine, entre le paradigme du refuge et le modèle de la colonisation active. *Stygologia* 3: 345-372.

Sbordoni, V. (1982) Advances in speciation of cave animals. *Progress in Clinical and Biological Research* 96: 219.

Socransky, S.S., and Manganiello, S.D. (1971) The oral microbiota of man from birth to senility. *Journal of Periodontology*42: 485-496.

Tannock, G.W., and Savage, D.C. (1974) Influences of dietary and environmental stress on microbial populations in the murine gastrointestinal tract. *Infection and Immunity*9: 591-598.

Tomlinson, M. (2011) Ecological water requirements of groundwater systems: a knowledge and policy review. *Waterlines Report Series*.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., and Gordon, J.I. (2007) The human microbiome project. *Nature* 449: 804-810

Wang, Y., Gilbreath III, T.M., Kukutla, P., Yan, G., and Xu, J. (2011) Dynamic gut microbiome across life history of the malaria mosquito *Anopheles gambiae* in Kenya. *PloS One* 6: e24767.

Watts, C.H.S., and Humphreys, W.F. (1999) Three new genera and five new species of Dytiscidae (Coleoptera) from underground waters in Australia. *Records of the South Australian Museum* 32: 121-142.

Watts, C.H.S., and Humphreys, W.F. (2006) Twenty-Six New Dytiscidae (Coleoptera) of the Genera Limbodessus Guignot and Nirripirti Watts and Humphreys, from Underground Waters in Australia. *Transactions of the Royal Society of South Australia* 130: 123-185.

Statement of Authorship

Title of Paper	Diversity patterns of subterranean invertebrate fauna in calcretes of the Yilgarn Region, Western Australia
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Published in Marine and Freshwater Research

Principal Author

Name of Principal Author (Candidate)	Josephine Hyde
Contribution to the Paper	Collection of samples, analysing the data, interpretation of the data, writing the manuscript
Overall percentage (%)	65%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 15/6/18

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Steven Cooper
Contribution to the Paper	Collection of samples, editing the manuscript
Signature	Date 15/6/18

Name of Co-Author	William Humphreys
Contribution to the Paper	Collection of samples, editing the manuscript
Signature	Date 5/6/2018

Name of Co-Author	Andrew Austin

Contribution to the Paper	Editing the manuscript
Signature	Date 15/6/18

Name of Co-Author	Pablo Munguia
Contribution to the Paper	Analysing the data, interpretation of the data, writing the manuscript
Signature	Date 5 June 2018

Please cut and paste additional co-author panels here as required.

CHAPTER 2: DIVERSITY PATTERNS OF SUBTERRANEAN INVERTEBRATE FAUNA IN CALCRETES OF THE YILGARN REGION, WESTERN AUSTRALIA

Josephine Hyde[A,G,] Steven J. B. Cooper[A,B,] William F. Humphreys[C,D,] Andrew D. Austin[A] and Pablo Munguia[E,F]

[A]Australian Centre for Evolutionary Biology and Biodiversity, School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia.

[B]Evolutionary Biology Unit, South Australian Museum, North Terrace, Adelaide, SA 5000, Australia.

[C]Western Australian Museum, Welshpool DC, WA 6986, Australia.

[D]School of Animal Biology, University of Western Australia, Crawley, WA 6009, Australia.

[E]School of Biological Sciences and Environment Institute, University of Adelaide, SA 5005, Australia.

[F]RMIT Studios, RMIT University, Melbourne, Vic. 3000, Australia.

[G]Corresponding author. Email: josephine.hyde@adelaide.edu.au

Abstract

Calcrete aquifers are unique groundwater habitats containing stygobiontic species endemic to each calcrete. The evolutionary history of stygofauna suggests the calcretes in Western Australia contain multiple ancient lineages, yet populations experience episodic variation in rainfall patterns, with little-known ecological consequences. The aim of the present study was to document stygofaunal diversity patterns and determine whether they are influenced by rainfall events. The average taxon richness in boreholes peaked shortly after periods of high rainfall, and when dominant taxa were excluded, evenness decreased after periods of both high and low rainfall, indicating that dominant taxa are an important factor in driving the system. Strong abundance–distribution relationships reflected the commonality of taxon groups; common taxon groups had broad distributions and high abundance levels, whereas rare taxon groups had small distributions and low abundance. After periods of intermediate and low rainfall, taxon groups had narrower distributions and the maximum number of individuals per borehole was lower. Finally, the majority of boreholes did not show changes in evenness over the 11-year study period, suggesting a reasonably stable ecosystem with episodic fluctuations that can be attributed to rainfall events. The results of the present study indicate that diversity patterns within boreholes are driven episodically by both external and internal factors, such as rainfall and rapid borehole dominance respectively.

2.1 Introduction

Spatiotemporal patterns of diversity are highly dependent on both habitat and species traits (Chase and Leibold 2003). In general, species diversity increases with increasing habitat heterogeneity (MacArthur and MacArthur 1961) because species can occupy and exploit different niches. However, in fragmented habitats, diversity tends to decrease as dispersal and resource availability becomes limiting for new colonisation events, and competition limits population growth (Mouquet and Loreau 2003; Munguia *et al*. 2011; Munguia 2015). Patchiness in small fragmented landscapes can facilitate specialisation that may otherwise not occur and can result in the evolution of endemic species (MacArthur and MacArthur 1961). Fragmentation can also lead to some species being more abundant than others in particular environments and potentially lead to the formation of areas with only one taxon group present (MacArthur and MacArthur 1961; Davies and Margules 1998).

A unique fragmented habitat comprising groundwater calcretes in the Yilgarn region of central Western Australia was recently shown to contain a diverse ecosystem of subterranean groundwater animals known as stygofauna. These groundwater calcretes formed in the arid Yilgarn region where the mean annual rainfall is below 200 mm and potential evapotranspiration exceeds 3 m (Mann and Horwitz 1979). Such groundwater habitats are generally entirely enclosed, having been deposited from the groundwater as it approaches base level close to playas (salt lakes) in palaeodrainage systems (Humphreys 2001). The Yilgarn region contains over 200 large isolated calcretes, with several phylogeographic studies having demonstrated that species are each restricted in their distribution to a single calcrete (Cooper *et al*. 2002, 2007, 2008; Leys *et al*. 2003), an endemicity that can be attributed to the insular form of the calcretes and the nature of the intervening regolith (Anand and Paine 2002) likely preventing dispersal between calcretes (Guzik *et al*. 2011).

In subterranean habitats, the trophic structure is simplified relative to epigean habitats owing to the absence of primary producers, save in rare systems that have chemoautotrophic primary producers and parasites (Gibert and Deharveng 2002), and sparse energy resources (Huppop 2000; Culver and White 2005). Because subterranean ecosystems are continuously dark, micro-organisms (bacteria and fungi) are usually dependent on organic carbon for energy capture. The energy that does enter the system is primarily mediated by the movement of water, and the exogenic organic matter

carried in by water is, in turn, thought to be affected by both rainfall and plant growth (Deharveng and Bedos 2000; Humphreys 2012). Rainfall in the Australian arid zone is episodic both within and between years (Mann and Horwitz 1979), and amounts equivalent of mean annual rainfall may occur in a single event or, in contrast, drought can span several years. As such, carbon inputs into the calcrete aquifers are also likely to be highly variable, but the ecological effect of this variability on stygofaunal communities is unknown.

Herein we document stygofauna diversity patterns within a calcrete located at the Sturt Meadows pastoral property in central Western Australia, which contains a suite (~100) of boreholes that enable assessment of spatial and temporal ecological changes across part of the aquifer (Fig. 2.1). Because the system experiences episodic rainfall events followed by long periods of no precipitation, we focused on how rainfall may affect diversity sampled within boreholes. We calculated the number of taxon groups and faunal evenness occurring within individual boreholes and estimated changes in common and rare taxon groups driven by rainfall patterns. Given the boom-and-bust population dynamics observed in other arid zone aquatic systems (e.g. Lake Eyre aquatic invertebrates; Davis *et al*. 2013), we explored whether sampled boreholes were dominated by a single taxon group. Given the periodic changes in the water table and the heterogeneous structure of the calcrete (Guzik *et al*. 2009; Bradford *et al*. 2013), we hypothesised that specific taxon groups may become the dominant group within a borehole because of reduced connectivity during low water table periods. We discuss two potential mechanisms arising from the observed rainfall patterns that could affect the dynamics of this unique subterranean ecosystem: (1) changes in connectivity within the calcrete due to fluctuations in the water table affecting dispersal; and (2) nutrient replenishment driven by rainfall. Ultimately, these proximate, episodic mechanisms could be the drivers behind the high levels of endemicity.

2.2 Materials and methods

The study site was at a calcrete located on the Sturt Meadows pastoral station in the Yilgarn region, Western Australia (Fig. 2.1), with an area of ~43 km^2. The surface vegetation is an open Acacia woodland with lowland shrubs. In a few places, the calcrete is exposed on the surface, whereas for the most part the top of the calcrete is up to 2 m below the surface (Allford *et al*. 2008; Bradford *et al*. 2013). Bores were originally drilled for mineral exploration to a depth of 10.3 m and a diameter of 100

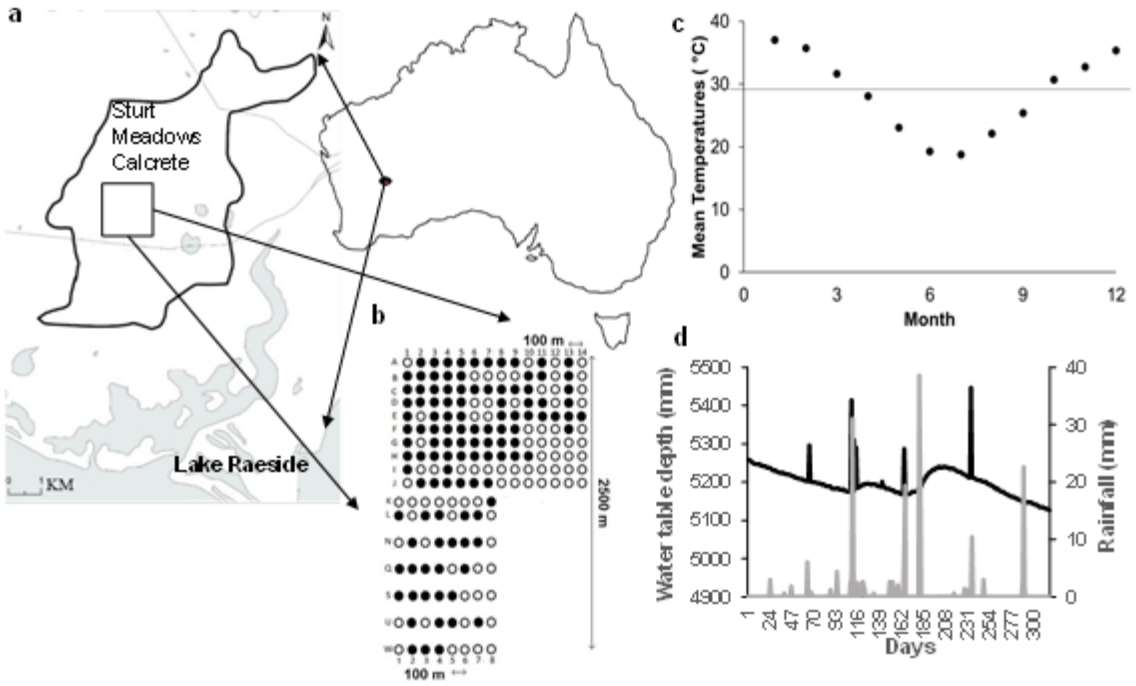

Fig. 2.1 (*a*) Location of Sturt Meadows within Australia and (*b*) the study area with sampled boreholes (black circles). (*c*) Climatological data showing mean temperature. The horizontal line indicates mean temperature for the year. (*d*) Borehole monitoring for recharge over a 12-month period at one borehole location, showing water depth (black line) and rainfall events (grey line).

mm. Two bore grids were drilled, a northern and a southern grid. The northern grid has bores spaced at 100 m in each direction and is 1.4 × 0.9 km; the southern grid has bores spaced at 100 m east–west and 200 m north–south and is 1.2 × 0.9 km (Fig. 2.1). The present study site covered both grids and an area of ~2.34 km^2. The current condition of the bores allows access to the water table from ~1.9 to 4.0 m below the surface, with water depth varying between 0.4 and 8 m among each of the 116 bores that have been accessed to date across the bore grid (Allford *et al*. 2008).

The Sturt Meadows calcrete stygofauna consists of 18 known macroinvertebrate taxa. There is a sister species triplet of dytiscid diving beetles *Paroster macrosturtensis*, *P. mesosturtensis*, and *P. microsturtensis* (Watts and Humphreys 2006), three species of chiltoniid amphipods, namely *Yilgarniella sturtensis*, *Scutachiltonia axfordi* and *Stygochiltonia bradfordae* (Bradford *et al*. 2010; King *et al*. 2012), at least four divergent lineages of oligochaete worms (Bradford *et al*. 2010) and eight copepod species (T. Karanovic, pers. comm.). There are also troglobiont species present above the water table, including two oniscidean isopod species (*Troglarmadillo sp*. and *Paraplatyarthrus occidentoniscus*; Javidkar *et al*. 2015, 2017), at least one mite species (Acari), one centipede species (Chilopoda), one palpigrade species (*Eukoenenia*

guzikae; Barranco and Harvey 2008) and one springtail species (Collembola; Bradford *et al*. 2010).

The study comprised 11 sampling events at the Sturt Meadows calcrete: 2 in 2004 (March and September), 2 in 2005 (March and April), 3 in 2006 (March, July and November), 2 in 2007 (February and May) and 1 each in 2011 and 2015 (May and April respectively). All samplings from 2004 to 2015 were undertaken using a small weighted plankton net (250 µm) hauled through the water column. In addition, sampling between 2005 and 2007 used a pump to retrieve samples. Previous research has shown that these two methods do not differ in their sampling intensity (Allford *et al*. 2008). The sampling effort from 2004 to 2015 resulted in the collection of 9118 stygobitic macroinvertebrate specimens from 512 sampling events from 116 boreholes (with 78% of boreholes sampled at least twice). Identification to family level for the amphipods, copepods and oligochaetes and species level for the beetles was performed in the field, and samples were stored in liquid nitrogen, or in 75 or 100% ethanol. Copepods and oligochaetes could not be identified to species level in the field because of a lack of taxonomic information for these groups. Morphological descriptions of three amphipod species were recently published by King *et al*. (2012), but these were not available for the early collections, where specimens had been discarded, so we were unable to identify the amphipods to species level. Therefore, our analyses clustered copepod, amphipod and oligochaete species each into distinct functional groups because these species usually share similar trophic levels (Gibert and Deharveng 2002; Bradford *et al*. 2014).

Environmental data as specified in Watts and Humphreys (2006) were collected in March 2006 and April 2015 and obtained while invertebrate samples were being collected. The nearest pertinent meteorological data were sourced from the Australian Bureau of Meteorology, namely rainfall events at Sturt Meadows from 2004 to 2015 and average temperature data from two different stations located 1.4 km apart in Leonora (~42 km from the Stuart Meadows calcrete) because neither station had complete records for the 2004–15 period. The average pan evaporation is 2400 mm $year^{-1}$, which far exceeds the average yearly rainfall of just over 200 mm.

Table 2.1 Sampling data information from 2004 to 2015 including month, period, rainfall and number of boreholes sampled

Month	Year	Period	Number of boreholes sampled	Rainfall in the 30 days before sampling (mm)
March	2004	High rainfall	14	68.4
September	2004	Low rainfall	66	3.6
March	2005	Low rainfall	26	2.6
April	2005	Low rainfall	64	1.4
March	2006	High rainfall	54	57.6
July	2006	Low rainfall	52	0
November	2006	Intermediate rainfall	52	10.4
February	2007	High rainfall	53	59
May	2007	Intermediate rainfall	54	27
May	2011	Intermediate rainfall	23	11.2
April	2015	Intermediate rainfall	53	12

2.2.1 Data Analysis

Monitoring water table levels in the aquifer from July 2006 to May 2007 indicated that recharge occurred shortly after a major rainfall event, with low rainfall events (<10 mm) having little or no affect on groundwater levels, rainfall events between 10–30 mm resulting in a moderate increase in groundwater levels and rainfall events >30 mm leading to a major recharge of the aquifer (Fig. 2.1). Therefore, we defined low-, intermediate- and high-rainfall periods based on total rainfall in the 30 days before sampling (Table 2.1). Low rainfall was defined as <10 mm rain in the previous 30 days, intermediate rainfall was defined as <30 mm rainfall in the previous 30 days and high rainfall was defined as ≥30 mm rainfall in the previous 30 days.

We first compared taxon richness and evenness among the three rainfall periods (high, intermediate and low rainfall) using mixed models (e.g. Darnell *et al*. 2015). Rainfall period was used as a fixed factor and sampling year was used as a random effect to account for year-to-year variability. Borehole occupancy, including unoccupied sites, was compared among rainfall periods for each taxon group using a contingency test to contrast periods and a Cochran–Mantel–Haenszel test to assess among-year differences in the number of sites occupied.

To determine the effects that the different rainfall events had on the various taxon groups, site occupancy was calculated for all taxon groups. The average number

of individuals' present was calculated for each rainfall period. The taxon groups were ordered in the graph by their rank site occupancy in the high-rainfall period.

Taxon group abundance–distribution patterns were compared for each of the three rainfall periods. An analysis of covariance (ANCOVA) tested differences in the relationship between maximum abundance within boreholes against the number of boreholes occupied by each taxon group. A positive relationship between borehole abundance and distribution was expected because it is a common pattern across systems and taxa (Magurran 1988; Munguia 2014). Therefore, we followed the ANCOVA with individual regressions to test whether the slope of abundance to distribution was different from zero in each of the three rainfall periods. Data were log transformed to meet parametric assumptions.

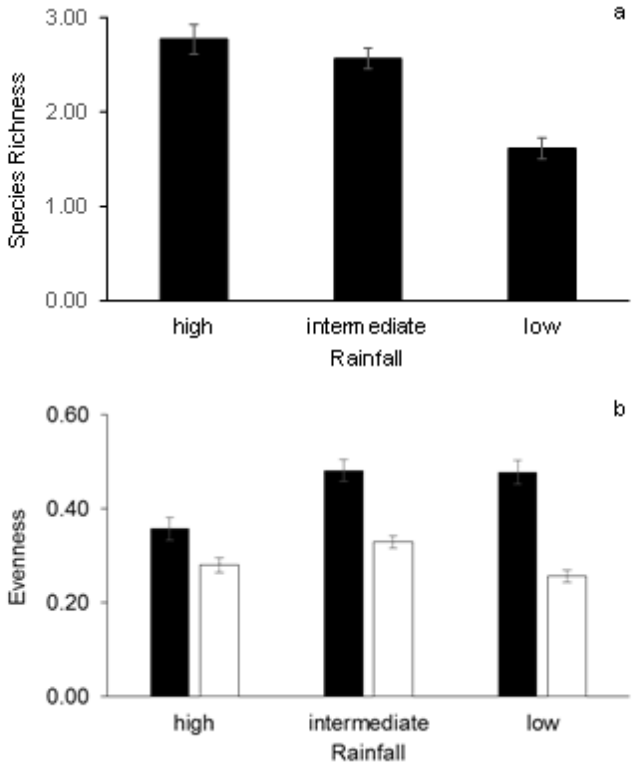

Fig. 2.2 (*a*) Species richness and (*b*) evenness for the three different episodic phases in Sturt Meadows calcrete. In (*b*), black columns represent total evenness, whereas open columns represent evenness with dominant taxa removed. Data are the mean ± s.e.m. Low rainfall <10 mm rain in the previous 30 days, intermediate rainfall <30 mm rainfall in the previous 30 days and high rainfall ≥30 mm rainfall in the previous 30 days.

Given the hypothesis that dominant taxa were driving the system, we further investigated the observed sites where a single taxon group was exclusively found (henceforth referred to as 'dominant taxon sites'). To determine whether all taxon groups were equally likely to become dominant, the proportion of sites with dominant

taxon groups from the total number of sites in a given sampling period averaged across each rainfall period was calculated and a Chi-Square test was used to compare their frequency across the three rainfall periods.

To test changes in long-term diversity patterns, the 18 boreholes that were repeatedly sampled most often over the 11 sampling events were selected and evenness was compared in both 2006 and 2015 using a paired *t*-test. Finally, a principal component analysis (PCA) was used to associate environmental parameters of each individual borehole (depth, O_2, temperature, pH and salinity) from each of the 2006 and 2015 surveys. Principal components were rotated using varimax and Factor 1 was used to represent the environmental gradient present in boreholes. Next, a mixed model tested the effect of environment (as PCA Factor 1) on richness, total abundance and evenness using year as a random effect. All analyses were performed in JMP (SAS Institute, Cary, NC, USA).

2.3 Results

Stygofaunal diversity in boreholes differed among the three rainfall periods. Diversity was lowest during the low-rainfall period (Fig. 2.2a) and greatest during the high-rainfall period ($F = 8.057$, $P < 0.001$), with year-to-year variation accounting for 13% of the variance. Evenness also differed among rainfall periods, with boreholes showing lowest evenness during the high-rainfall period (Fig. 2.2b) and highest evenness during the intermediate- and low-rainfall periods ($F = 7.316$, $P = 0.001$), with year-to-year variation accounting for 7.6% of the variance. When removing sites with a single dominant taxon group, both the high- and low-rainfall periods had the lowest evenness (Fig. 2.2b) and the intermediate-rainfall period maintained the highest evenness ($F = 10.029$, $P < 0.001$), with year-to-year variation accounting for 6.7% of the variance.

The effect of rainfall on the distribution of different taxon groups varied (Fig. 2.3). For example, copepods occupied a large proportion of sites during high- and intermediate-rainfall periods (57 and 51% respectively), but few sites (16%) during low-rainfall periods. In contrast, the dytiscid beetle *P. microsturtensis* did not significantly change proportional occupation of sites with changes in rainfall period (0, 2 and 1.5% for high, intermediate and low respectively; Table 2.2).

The more boreholes a taxon group occupied, the greater its abundance (ANCOVA, $F = 4.46$, $P = 0.01$; Fig. 2.4), yet this pattern varied within each rainfall

period. A positive relationship between rainfall events and maximum abundance was observed in each rainfall period (Fig. 2.4). During high-rainfall events, a species' maximum abundance was independent of its distribution among boreholes (Fig. 2.4a; $F = 3.04$, d.f. = 1, $P = 0.11$, $R^2 = 0.43$). In the intermediate-rainfall season, the relationship between maximum abundance and distribution was marginally significant (Fig. 2.4b; $F = 6.22$, d.f. = 1, $P = 0.05$, $R^2 = 0.55$), whereas during the period of low rainfall a significant relationship was found between maximum abundance and distribution (Fig. 2.4c; $F = 11.21$, d.f. = 1, $P = 0.02$, $R^2 = 0.69$). Copepods had the greatest abundance levels during high- and intermediate-rainfall periods, yet they did not affect the overall results of the regressions.

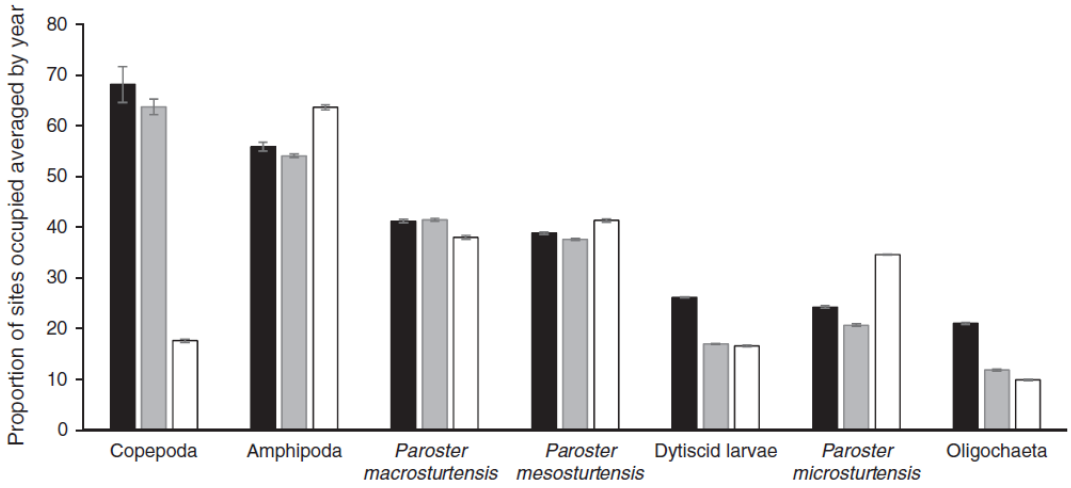

Fig. 2.3 Proportion of occupied sites averaged by year by high rainfall (black columns), intermediate rainfall (grey columns) and low rainfall (white columns) by the seven different species groups in Sturt Meadows calcrete. Data are the mean ± s.e.m.

The proportion of sites that had dominant taxon groups varied among rainfall periods ($\chi^2 = 30$, $P < 0.05$: 10.7% during high rainfall, 37.6% during intermediate rainfall and 30% during low rainfall; Fig. 2.5). For sites that had dominant taxon groups, these groups were mainly either amphipods or copepods. However, dominance shifted; during intermediate- and high-rainfall periods, copepods dominated; during low-rainfall periods amphipods dominated.

Borehole diversity remained stable between 2006 and 2015 (Fig. 2.6), with evenness not differing between these two years (paired t-test, $t = 0.77$, $P = 0.45$), and remaining reasonably low (J' (evenness) = 0.32 and 0.39 in 2006 and 2015

respectively). There were four sites that showed increases in evenness, and this pattern was driven by the dominant taxon groups in 2015.

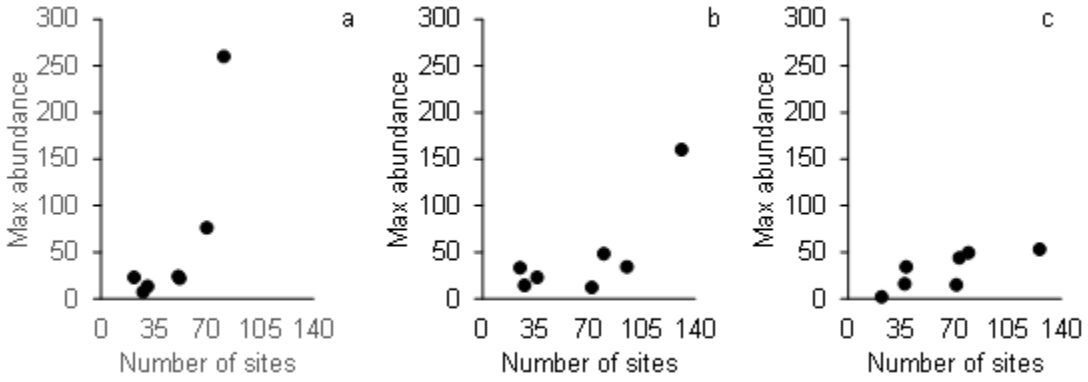

Fig. 2.4 Maximum abundance for the total number of individuals over the entire sampling period by borehole, for periods with (a) high, (ab) intermediate (b) and (c) low (c) rainfall periods.

Table 2.2 Chi-Square distribution tables calculated for the relative abundances of each taxon group from each rainfall period. A Cochran–Mantel–Haenszel (CMH) test was used to assess among-year differences in the number of sites occupied (Fig. 2.3)

Taxon group				CMH test		
	d.f.	χ^2	*P*-value	χ^2	d.f.	*P*-value
Copepoda	2	137.697	<0.0001	48.152	2	<0.0001
Amphipoda	2	4.804	0.091	4.295	2	0.117
Paroster mesosturtensis	2	4.605	0.100	3.393	2	0.183
Paroster microsturtensis	2	1.962	0.374	7.387	2	0.024
Dytiscid	2	6.259	0.0437	12.413	2	0.002
Paroster macrosturtensis	2	8.675	0.0131	0.279	2	0.869
Oligochaeta	2	19.213	<0.0001	10.733	2	0.005

Boreholes showed consistency in environmental gradients between the two years of 2006 and 2015 (Table 2.3). In both years, an environmental gradient was produced in PC Factor 1 (PC1) from high temperature and pH to high salinity and greater depth (Table 2.3). Oxygen did not contribute greatly to this gradient; instead, it was influential on an orthogonal gradient (PC2; Table 2.3). Taxon richness increased with the environmental gradient ($F = 4.11$, $P = 0.05$) where year explained only 0.8%, of the total variance (Fig. 2.7a, b). Similarly, evenness dropped as the scores on PC1 increased

(F = 5.65, P = 0.02) and year did not explain any of the variance in the model (Fig. 2.7c, d). However, these relationships were weak because environment only explained a small proportion of the variation in either taxonomic richness or evenness (Fig. 2.7).

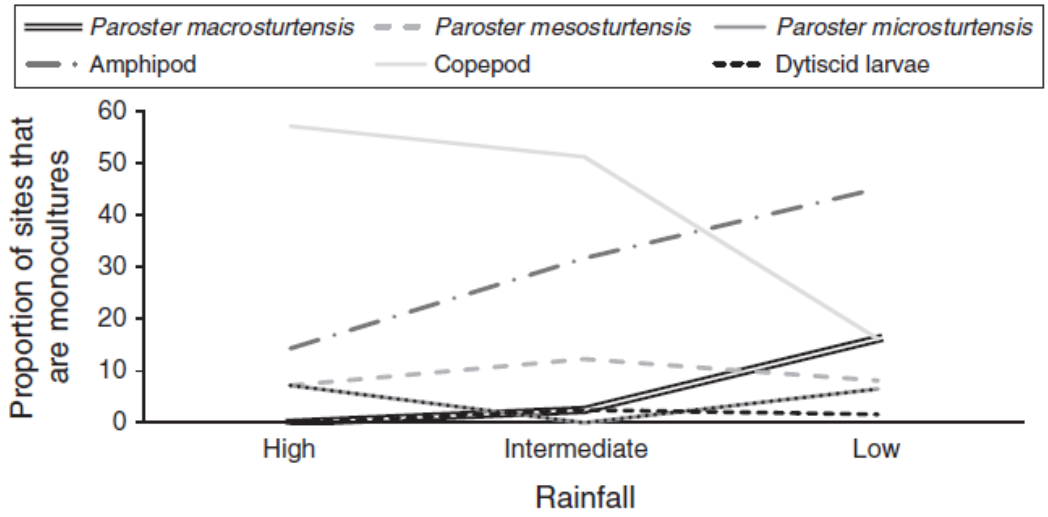

Fig. 2.5 Proportion of sites by taxon group that were present as dominant taxa (only one taxon group present in an individual borehole).

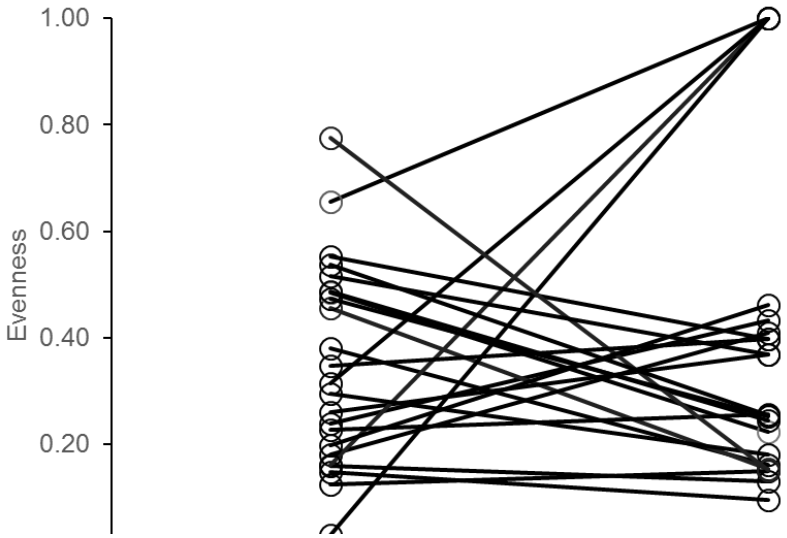

Fig. 2.6 Species evenness (J') of the 20 most sampled boreholes of the total sampling period in 2006 and 2015.

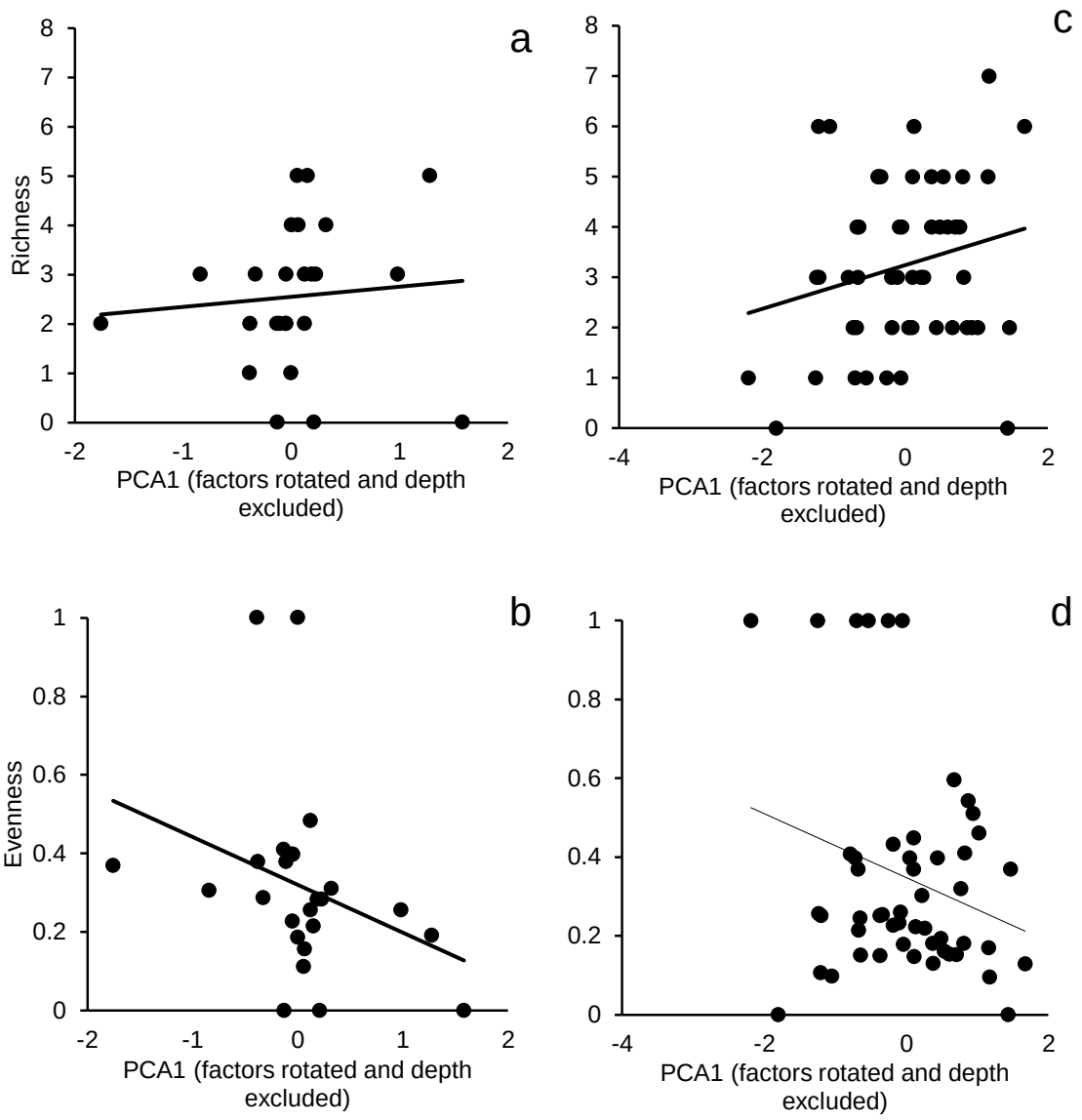

Fig. 2.7 (a, b) Species richness in 2006 (a) and 2015 (b) by the environmental factors calculated into principal component (PC) 1 and then rotated and depth excluded. (c, d) Species evenness in 2006 (c) and 2015 (d) by the environmental factors calculated into PC1 and then rotated and depth excluded. However, the R^2 values for the models were low: (a) $R^2 = 0.01$, (b) $R^2 = 0.05$, (c) $R^2 = 0.10$ and (d) $R^2 = 0.06$.

Table 2.3 Principal component analysis loadings of the environmental parameters for Factors 1 and 2 in 2006 and 2015. Numbers in parentheses are the percentage variance explained by each factor. Salinity was measured on the practical salinity scale (PSS).

	2006		2015	
	Factor 1 (37.77%)	Factor 2 (29.95%)	Factor 1 (47.16%)	Factor 2 (25.34%)
Temperature (°C)	–0.562	–0.741	0.927	–0.026
pH	–0.502	0.384	0.663	0.329
Salinity (PSS)	0.753	–0.055	–0.465	–0.629
O_2 (mg L^{-1})	–0.202	0.890	–0.101	0.872
Depth (m)	0.845	0.079	–0.912	–0.015

2.5 Discussion

At Sturt Meadows, the subterranean invertebrate community sampled within boreholes exhibited shifts in composition with an increase in taxon richness during periods of high and intermediate rainfall. It also showed periods of high taxon evenness during times of intermediate and low rainfall after sites with a single dominant taxon group were excluded from the analyses. Variation in community composition appears to be driven to a large extent by the dominant taxon groups and, in particular, copepods. The proportion of sites that had dominant taxon groups decreased during high-rainfall periods, whereas the composition of dominant taxon sites also changed with rainfall period. This system has dynamic variation that is determined by rainfall. However, evidence from long-term temporal patterns suggests that over long periods of time the system is also stable.

Significant precipitation events and subsequent aquifer recharge increased borehole occupancy. After moderate to large precipitation events, the water table in the calcrete increased (Fig. 2.1), possibly leading to easier movement between different areas of the calcrete. After these rainfall events, the proportion of sites that contained dominant taxon groups decreased, suggesting that precipitation events affect taxon group composition of individual boreholes. Because the timing and amount of rainfall vary, we suggest that the connectivity of the calcrete changes and stochastic dispersal events are likely to play a role in the frequency and abundance of taxon groups within individual boreholes. When the water table decreases, taxa may become isolated at specific locations and concentrated into a smaller volume of water. Habitat heterogeneity, possibly resulting from porosity and water chemistry changes, is also likely to affect species composition in these subterranean systems.

Water chemistry can be important in structuring invertebrates' communities in fresh water, with studies detailing a decrease in taxon richness where water chemistry is extreme (Heino 2000), for example when the pH is below 5 or salinity is >1.5 (calculated from Reeves *et al*. 2007). Conversely, except for salinity, water chemistry seems to have remarkably little effect on the distribution of ostracods in groundwater systems of the Pilbara Region of Western Australia, including in calcretes (Reeves *et al*. 2007), suggesting that individual taxonomic groups may show idiosyncratic responses to water chemistry. High pH and low oxygen concentrations create thresholds (Adlassnig *et al*. 2012) that can cause local extinctions by changing prey abundance or nutrients in the water column. Although the distribution of species in calcretes is affected by rainfall, it is also possible that other factors, such as the spatial heterogeneity of the calcrete due to a fluctuating water table, provide niche space availability and the potential for refugia in deep calcrete deposits, as well as extinction of populations in shallow calcrete areas (Bradford *et al*. 2013). Potentially, this could result in temporary or long-term isolation of species, thus affecting their distribution and abundance.

The Western Australian calcretes have provided habitat for stygofauna for millions of years. The isolation of these calcretes following post-Miocene aridification of the Australian continent has been inferred from the distribution of multiple obligate stygobiotic lineages and molecular phylogeographic studies (Cooper *et al*. 2002; Leijs *et al*. 2012). Owing to the long-term stability and isolation of the calcretes, and the long evolutionary history of their resident communities, we would expect localised borehole extinctions and year-to-year variation. However, rainfall events cause rapid and extreme changes to the environment and a platform for environmental variation.

The chemocline of the water table within calcretes is expected to support a complex microbiological community, based on studies of similarly complex anchialine systems (Humphreys *et al*. 2009). Salinity in groundwater calcretes increases towards groundwater base level, typically as it moves towards a salt lake. At the Sturt Meadows study site, salinity increases across the calcrete towards Lake Raeside (Humphreys *et al*. 2009). Infiltration of rainfall carrying particulate organic carbon and dissolved organic carbon into the calcrete would affect the microbiological community, which is likely to be the lowest trophic level in the food chain. Changes to the composition and abundance

of the microbial community would potentially cascade up the system, affecting the invertebrates and their location and abundance.

The present study has revealed a highly dynamic and episodically rainfall-dependent subterranean system that has ancient lineages of taxa inhabiting groundwater calcretes. Although one of the most significant factors in driving diversity patterns within the boreholes is periodic recharge from rainfall, this does not explain all the changes within the system. It is possible that changes in the water table, the introduction of nutrients, or a combination of factors are also affecting changes in species composition and abundance, and further research will be necessary to determine whether this is the case.

Acknowledgements

The authors thank Adam Allford for the sampling data he provided to this study, and Michelle Guzik and Tessa Bradford for assistance with collections. The authors also thank Flora, Peter and Paul Axford of the Sturt Meadows Station for providing both accommodation and access to their property. In addition, the authors thank Ron Smernik for comments that greatly improved the manuscript. This research was funded by Australian Research Council linkage grants (LP0348753, LP100200494 and LP140100555) to Steven J. B. Cooper, William F. Humphreys, Andrew D. Austin and Pablo Munguia, with industry partners Newmont Australia, Placer Dome Asia Pacific, Minara Resources Limited, the Western Australian Museum, the South Australian Museum, Biota Environmental Sciences, Bennelongia Environmental Consultants and the Department of Parks and Wildlife (WA).

References

Adlassnig, W., Koller-Peroutka, M., Bauer, S., Koshkin, E., Lendl, T., and Lichtscheidl, I. K. (2012). Endocytotic uptake of nutrients in carnivorous plants. *The Plant Journal* 71, 303–313. doi:10.1111/j.1365-313X.2012.04997.x

Allford, A., Cooper, S. J. B., Humphreys, W. F., and Austin, A. D. (2008). Diversity and distribution of groundwater fauna in a calcrete aquifer: does sampling method influence the story? *Invertebrate Systematics* 22, 127–138. doi:10.1071/IS07058

Anand, R. R., and Paine, M. (2002). Regolith geology of the Yilgarn Craton, Western Australia: implications for exploration. *Australian Journal of Earth Sciences* 49, 3–162. doi:10.1046/j.1440-0952.2002.00912.x

Barranco, P., and Harvey, M. S. (2008). The first indigenous palpigrade from Australia: a new species of Eukoenenia (Palpigradi: Eukoeneniidae). *Invertebrate Systematics* 22, 227–233. doi:10.1071/IS07031

Bradford, T.M., Adams, M., Humphreys, W. F., Austin, A. D., and Cooper, S. J. B. (2010). DNA barcoding of stygofauna uncovers cryptic amphipod diversity in a calcrete aquifer in Western Australia's arid zone. *Molecular Ecology Resources* 10, 41–50. doi:10.1111/j.1755-0998.2009.02706.x

Bradford, T.M., Adams, M., Guzik, M. T., Humphreys, W. F., Austin, A. D., and Cooper, S. J. B. (2013). Patterns of population genetic variation in sympatric chiltoniid amphipods within a calcrete aquifer reveal a dynamic subterranean environment. *Heredity* 111, 77–85. doi:10.1038/hdy.2013.22

Bradford, T.M., Humphreys, W. F., Austin, A. D., and Cooper, S. J. B. (2014). Identification of trophic niches of subterranean diving beetles in a calcrete aquifer by DNA and stable isotope analyses. *Marine and Freshwater Research* 65, 95–104. doi:10.1071/MF12356

Chase, J. M., and Leibold, M. A. (2003). 'Ecological Niches: Linking Classical and Contemporary Approaches.' (University of Chicago Press: Chicago, IL, USA.)

Cooper, S. J. B., Hinze, S., Leys, R., Watts, C. H. S., and Humphreys, W. F. (2002). Islands under the desert: molecular systematics and evolutionary origins of stygobitic water beetles (Coleoptera: Dytiscidae) from central Western Australia. *Invertebrate Systematics* 16, 589–590.

Cooper, S. J. B., Bradbury, J. H., Saint, K. M., Leys, R., Austin, A. D., and Humphreys, W. F. (2007). Subterranean archipelago in the Australian arid zone: mitochondrial DNA phylogeography of amphipods from central Western Australia. *Molecular Ecology* 16, 1533–1544. doi:10.1111/j.1365-294X.2007.03261.x

Cooper, S. J. B., Saint, K. M., Taiti, S., Austin, A. D., and Humphreys, W. F. (2008). Subterranean archipelago: mitochondrial DNA phylogeography of stygobitic isopods (Oniscidea: Haloniscus) from the Yilgarn region of Western Australia. *Invertebrate Systematics* 22, 195–203. doi:10.1071/IS07039

Culver, D. C., and White, W. B. (2005). 'Encyclopedia of Caves.' (Elsevier: Amsterdam, Netherlands.)

Darnell, M. Z., Nicholson, H. S., and Munguia, P. (2015). Thermal ecology of the fiddler crab *Uca panacea*: thermal constraints and organismal responses. *Journal of Thermal Biology* 52, 157–165.

Davies, K. F., and Margules, C. R. (1998). Effects of habitat fragmentation on carabid beetles: experimental evidence. *Journal of Animal Ecology* 67, 460–471. doi:10.1046/j.1365-2656.1998.00210.x

Davis, J., Pavlova, A., Thompson, R., and Sunnucks, P. (2013). Evolutionary refugia and ecological refuges: key concepts for conserving Australian arid zone freshwater biodiversity under climate change. *Global Change Biology* 19, 1970–1984. doi:10.1111/gcb.12203

Deharveng, L., and Bedos, A. (2000). The cave fauna of Southeast Asia. Origin, evolution andecology. In *'Subterranean Ecosystems'*. (Eds H. Wilkins, D. C, Culver, W. F., Humphreys,) pp. 603–632. (Elsevier: Amsterdam and New York.)

Gibert, J., and Deharveng, L. (2002). Subterranean ecosystems: a truncated functional biodiversity. *Bioscience* 52, 473–481. doi:10.1641/0006-3568(2002)052[0473:SEATFB]2.0.CO;2

Guzik, M.T., Cooper, S. J. B., Humphreys, W. F., and Austin, A. D. (2009). Fine-scale comparative phylogeography of a sympatric sister species triplet of subterranean diving beetles from a single calcrete aquifer in Western Australia. *Molecular Ecology* 18, 3683–3698. doi:10.1111/j.1365-294X.2009.04296.x

Guzik, M. T., Austin, A. D., Cooper, S. J. B., Harvey, M. S., Humphreys, W. F., Bradford, T.M., Eberhard, S. M., King, R. A., Leys, R., and Muirhead, K. A.

(2011). Is the Australian subterranean fauna uniquely diverse? *Invertebrate Systematics* 24, 407–418. doi:10.1071/IS10038

Heino, J. (2000). Lentic macroinvertebrate assemblage structure along gradients in spatial heterogeneity, habitat size and water chemistry. *Hydrobiologia* 418, 229–242. doi:10.1023/A:1003969217686

Humphreys, W. F. (2001). Groundwater calcrete aquifers in the Australian arid zone: the context to an unfolding plethora of stygal biodiversity. *Records of the Western Australian Museum* 64(Suppl.), 63–83. doi:10.18195/issn.0313-122x.64.2001.063-083

Humphreys, W. F. (2012). Diversity patterns in Australia. In '*Encyclopedia of Caves*', 2nd edn. (Eds W. B. White and D. C. Culver.) pp. 203–219. (Elsevier: Amsterdam and New York)

Humphreys, W. F., Watts, C. H. S., Cooper, S. J. B., and Leijs, R. (2009). Groundwater estuaries of salt lakes: buried pools of endemic biodiversity on the western plateau, Australia. *Hydrobiologia* 626, 79–95. doi:10.1007/s10750-009-9738-4

Huppop, K. (2000). How do cave animals cope with the food scarcity in caves? In *'Subterranean Ecosystems'*. (Eds H. Wilkins, D. C, Culver, W. F., Humphreys.) pp. 159–188. (Elsevier: Amsterdam and New York)

Javidkar, M., Cooper, S. J. B., King, R. A., Humphreys, W. F., and Austin, A. D. (2015). Molecular phylogenetic analyses reveal a new southern hemisphere oniscidean family (Crustacea: Isopoda) with a unique water transport system. *Invertebrate Systematics* 29, 554–577. doi:10.1071/IS15010

Javidkar, M., King, R. A., Cooper, S. J. B., Humphreys, W. F., and Austin, A. D. (2017). Taxonomy of *Paraplatyarthrus* Javidkar and King (Isopoda: Oniscidea: Paraplatyarthridae) with description of five new species from Western Australia, and comments on Australian *Trichorhina* Budde-Lunde, 1908 (Platyarthridae). *Zootaxa* 4243, 401–431. doi:10.11646/zootaxa.4243.3.1

King, R. A., Bradford, T.M., Austin, A. D., Humphreys, W. F., and Cooper, S. J. B. (2012). Divergent molecular lineages and not-so-cryptic species: the first descriptions of stygobitic chiltoniid amphipods (Talitroidea: Chiltoniidae) from Western Australia. *Journal of Crustacean Biology* 32, 465–488. doi:10.1163/193724012X626566

Leijs, R., van Nes, E. H., Watts, C. H. S., Cooper, S. J. B., Humphreys, W. F., and Hogendoorn, K. (2012). Evolution of blind beetles in isolated aquifers: a test of

alternative modes of speciation. *PLoS One* 7(3), e34260. doi:10.1371/journal.pone.0034260

Leys, R., Watts, C. H. S., Cooper, S. J. B., and Humphreys, W. F. (2003). Evolution of subterranean diving beetles (Coleoptera: Dytiscidae Hydroporini, Bidessini) in the arid zone of Australia. *Evolution* 57, 2819–2834.

MacArthur, R. H., and MacArthur, J. W. (1961). On bird species diversity. *Ecology* 42, 594–598. doi:10.2307/1932254

Magurran, A. E. (1988). Why diversity? In 'Ecological Diversity and its Measurement'. (Eds A. E., Magurran.) pp. 1–5. (Springer: Amsterdam, Netherlands.)

Mann, A., and Horwitz, R. (1979). Groundwater calcrete deposits in Australia some observations from Western Australia. *Journal of the Geological Society of Australia* 26, 293–303. doi:10.1080/00167617908729092

Mouquet, N., and Loreau, M. (2003). Community patterns in source–sink metacommunities. *American Naturalist* 162, 544–557. doi:10.1086/378857

Munguia, P. (2014). Life history affects how species experience succession in pen shell metacommunities. *Oecologia* 174, 1335–1344. doi:10.1007/s00442-013-2849-7

Munguia, P. (2015). Role of sources and temporal sinks in a marine amphipod. *Biology Letters* 11, 20140864. doi:10.1098/rsbl.2014.0864

Munguia, P., Osman, R. W., Hamilton, J., Whitlatch, R., and Zajac, R. (2011). Changes in habitat heterogeneity alter marine sessile benthic communities. *Ecological Applications* 21, 925–935. doi:10.1890/09-2398.1

Reeves, J. M., De Deckker, P., and Halse, S. A. (2007). Groundwater Ostracods from the arid Pilbara region of northwestern Australia: distribution and water chemistry. *Hydrobiologia* 585, 99–118. doi:10.1007/s10750-007-0632-7

Watts, C. H. S., and Humphreys, W. F. (2006). Twenty-six new dytiscidae (Coleoptera) of the genera *Limbodessus* Guignot and *Nirripirti* Watts and Humphreys, from underground waters in Australia. *Transactions of the Royal Society of South Australia* 130, 123–185. doi:10.1080/3721426.2006.10887055

Statement of Authorship

Title of Paper	Diversity and geographic differences in viromes of the gut of Australian subterranean groundwater diving beetles
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Written in the style of the journal Environmental Microbiology

Principal Author

Name of Principal Author (Candidate)	Josephine Hyde
Contribution to the Paper	Collection of samples, molecular work, analysing the data, interpretation of the data, writing the manuscript
Overall percentage (%)	65 %
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 15/6/18

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Lisa Dann
Contribution to the Paper	Analysing data, interpretation of data, editing of manuscript
Signature	Date 20/06/2018

Name of Co-Author	William Humphreys
Contribution to the Paper	Collection of samples, editing of manuscript
Signature	Date per 19/6/2018

Name of Co-Author	Steven Cooper
Contribution to the Paper	Supervision of project, Collection of samples, editing of manuscript
Signature	Date 15/6/2018

Name of Co-Author	Andrew Austin
Contribution to the Paper	Supervision of project, Editing of the manuscript
Signature	Date 15/6/18.

Name of Co-Author	James Mitchell
Contribution to the Paper	Supervision of project, Discussion of concepts, editing of manuscript
Signature	Date 25/6/18

Please cut and paste additional co-author panels

CHAPTER 3: DIVERSITY AND GEOGRAPHIC DIFFERENCES IN VIROMES OF THE GUT OF AUSTRALIAN SUBTERRANEAN GROUNDWATER DIVING BEETLES

Josephine Hyde[1,6], Lisa M. Dann[2], William F. Humphreys[3,4], Steven J. B. Cooper[1,5], Andrew D. Austin[1], James G. Mitchell[2]

Australian Centre for Evolutionary Biology and Biodiversity, School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia

College of Science and Engineering at Flinders University, Adelaide, South Australia

Western Australian Museum, Welshpool DC, Western Australia 6986, Australia

School of Animal Biology, University of Western Australia, Crawley, Western Australia 6009, Australia

Evolutionary Biology Unit, South Australian Museum, North Tce, Adelaide, SA 5000, Australia

Corresponding author. Email: josephine.hyde@adelaide.edu.au

Abstract

Gut viromes are important for investigating the physiology and health of hosts as they stimulate immunity and help determine fitness. Despite this, there is little work on insect gut viromes and their external influences. Here we present stygobiontic (i.e. subterranean and aquatic) beetles as models to investigate virome-host-location interactions. Using random shotgun sequencing of six beetle species, including adult and larval specimens, we found that their gut viromes are host and life-stage specific; adult beetle viromes were dominated by vertebrate viruses, while larval viromes were dominated by invertebrate viruses. Location was important for the taxonomic makeup of viromes, with a Herpesviridae dominant at one site and a Polydnaviridae-like group dominant at a second location. The results at both sites suggest the discovery of previously unclassified viruses, while the high abundance of human-related viruses may indicate anthropogenic effects on the microbial community. This research will further our knowledge about natural insect viromes and contribute to our knowledge on how these species interact with their subterranean environment.

3.1 Introduction

The suggestion that disruption in commensal gut microbial flora can lead to a disease state and that microbes are essential to digestion and nutrition, remained nearly dormant for a century (Pasteur, 1870, 1885; Metchnikoff, 1901). However, high-throughput DNA sequencing has allowed investigation of the complex relationships between the gut microbiome (bacteria, archaea, fungi and viruses (Columpsi *et al*., 2016) and their host, and have revolutionised microbiology. Consequently, microbiome research has re-emerged as one of the most important areas relevant to health and animal physiology (Lee and Brey, 2013).

Microbial communities play key roles in the host's fitness including the host's metabolism, fecundity, immunity, and longevity (Dillon and Dillon, 2004; Wang *et al*., 2011). The gut microbiome is complex and comprises a community of bacteria, fungi, archaea and viruses, with viruses being the numerically dominant component (Sekirov *et al*., 2010; Columpsi *et al*., 2016). Despite the abundance of viral communities in the gut, termed the 'virome', most research has focused on the bacterial microbiome to the exclusion of the virome (Sekirov *et al*., 2010; Yun *et al*., 2014). Analyses of gut viromes have suggested that the most abundant viral type in animal gut viromes are bacteriophages, i.e. viruses that infect bacteria, with viral composition proposed to reflect the evolutionary history of the bacterial microbiota (Minot *et al*., 2011; Pride *et al*., 2012; Columpsi *et al*., 2016).

The composition of the gut virome is dependent on selective pressures, such as the host's environmental history and diet. The host's environment can impose selective pressures that lead the host and their microbiome to coevolve and adapt to these pressures. As the gut structure evolved, so has the complexity and diversity of its microbiome (Kostic *et al*., 2013). The host's diet can influence directly and indirectly the gut virome, with the diet acting as a reservoir for viruses and imposing a selection pressure on the microbiome influencing virus colonisation (Minot *et al*., 2011; Columpsi *et al*., 2016).

To investigate the gut virome, host model organisms are required. Previously, invertebrate and vertebrate models have facilitated the investigation of the host-virome relationship at varying levels of microbiome complexity due to varying levels of experimental control (Lee and Brey, 2013; Grover and Kashyap, 2014; Ma *et al*., 2015). By using these model host organisms, various important factors, such as those that

enable symbiosis have been discovered and have enabled improved understanding, particularly of the human microbiome (Kostic *et al*., 2013). Key model systems for gut microbiomes of increasing complexity include the Hawaiian bobtail squid (*Euprymna scolopes*), the fruit fly (*Drosophila melanogaster*), the Zebrafish (*Danio rerio*), and the mouse (*Mus musculus*) (Kostic *et al*., 2013). *Drosophila melanogaster* has been used to reveal new host-microbe relationships, which are conserved in humans. Additionally, *D. melanogaster* can also reveal associations that apply to other insects which are common vectors of infectious diseases and important in agriculture. Most model organisms currently are laboratory animals. However, analyses suggest that the microbiomes of laboratory-raised organisms do not necessarily reflect those in wild-caught individuals (Chandler *et al*., 2011; Rosshart *et al*., 2017). Here we overcome this limitation by using non-lab animals from a remote field location to assess the composition of a natural virome.

Understanding wild insect biology is important for disease and agriculture. The use of a wild insect species as a model organism is advantageous as gut virome mutualism is often conserved among higher-order taxa. Such a model can function as an ecological 'test-tube' and may be instrumental in dissecting host-virome and viral-bacterial relationships in insects. In this study, we use wild populations of subterranean aquatic diving beetles (Dytiscidae) that are found in isolated calcrete aquifers in the arid region of central Western Australia. These insect species have extremely circumscribed distributions and phenologies (Watts and Humphreys, 2009), and have been isolated over significant geological time (Leys *et al*., 2003), thus allowing evolutionary processes underlying colonisation and maintenance of the virome to be investigated thoroughly.

These subterranean beetles are already models for regressive evolution e.g. loss of eyes in subterranean animals; (Leys *et al*., 2005; Tierney *et al*., 2015), and are from the genera *Paroster*, *Limbodessus*, *Neobidessodes* and *Copelatus*, comprising in total about 100 species (Watts and Humphreys, 2004, 2009). These stygofaunal (i.e. subterranean and aquatic) species are characterised as extreme short-range endemics, with most species described to date being restricted in their distribution to a single aquifer (Cooper *et al*., 2002; Leys *et al*., 2003). Additionally, phylogenetic analyses suggest that colonisation and isolation of species within aquifers occurred during the aridification of the Australian interior 3-8 million years ago. Most species (>80%) have

evolved independently from a small number of surface ancestors and the remaining species are likely to have speciated within individual aquifers (Leys *et al*., 2003; Leijs *et al*., 2012). This biodiversity hotspot of subterranean beetles represents an ideal model system for development of a natural baseline for wild insect viromes and represents a natural repeated experiment to investigate evolutionary processes associated with the establishment and maintenance of the gut virome.

We present a metagenomics analysis of uncultured viral communities using random shotgun sequencing from six subterranean diving beetle species, specifically *Limbodessus palmulaoides, L. lapostaae, L. windarraensis, Paroster macrosturtensis, P. mesosturtensis,* and *P. microsturtensis* (Watts and Humphreys, 2006). These species are from two different calcretes that are separated by ca 130 km, which each contained larval and adult stages of the beetle species. Our goal was to answer several key ecological questions about these communities: (i) what taxa comprise the viral communities in the different beetle species? (ii) does each beetle species have a unique viral community? (iii) do the differences between communities depend on the evolutionary history of the beetle species? and (iv) do the viral communities differ between the developmental stages of the host (i.e. larvae vs adults)? By answering these questions, the complex interactions that occur naturally within the wild between hosts and their gut virome can be determined. This understanding will further our knowledge about natural insect viromes and contribute to our knowledge on how these species interact with their subterranean environment.

3.2 Results

3.2.1 Viral Metagenomics overview

Raw unjoined sequence reads in FASTQ format underwent quality control where low quality sequences and non-viral sequences were removed (Table S3.1). A total of 0.96% of all sequences were removed from the samples from Sturt Meadows, while 0.62% of all sequences from Laverton were removed. According to the taxonomic results the top 10 most abundant viral families over all the beetle samples were Herpesviridae, Baculoviridae, a Polydnaviridae-like group, Myoviridae, Siphoviridae, Phycodnaviridae, Microviridae, Mimiviridae, Poxviridae, and Pandoraviridae. Habitat affiliations of the most abundant viruses were also explored, and were grouped into seven categories, microalgae, amoeba, bacteria, invertebrates, plants, vertebrates and other. The ‘other’ category included viruses that did not have any known host

affiliations, while one virus had a bicosoecid flagellate as a known host. Individuals from Laverton and Sturt Meadows had viruses with known host affiliations in all seven categories (Fig. 3.1a).

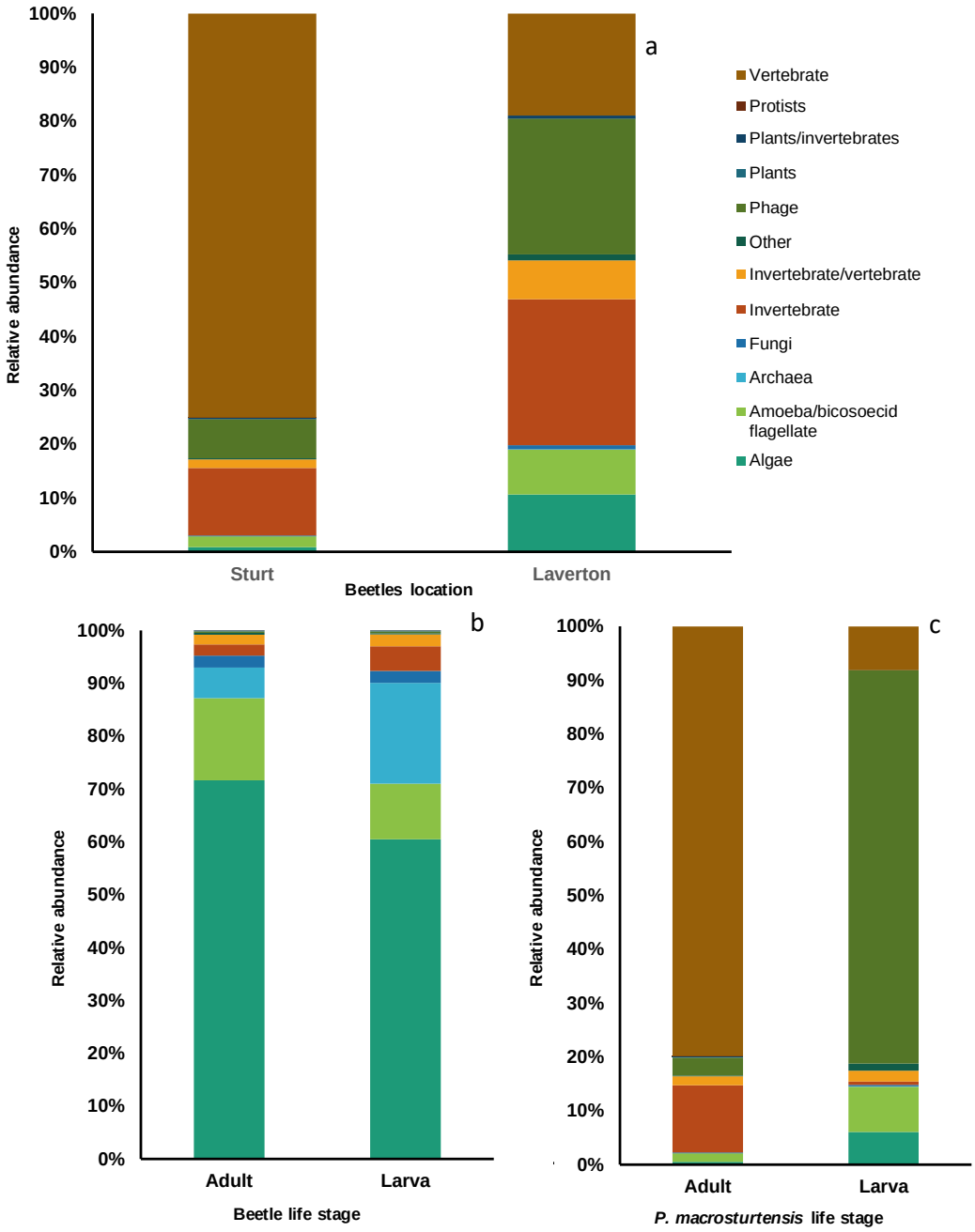

Fig. 3.1 Viral relative abundance by host type (a) comparing Sturt Meadows individuals to Laverton individuals (b) comparing all adult to larval beetles regardless of location (c) comparing all adult *Paroster macrosturtensis* individuals to larvae

3.2.2 Viral results by habitat affiliation

For the *Limbodessus* species from Laverton, 53 viral families were identified (Table S3.5). The seven most abundant families made up 80.5% of the average viral abundance (Table 3.1). For *L. palmulaoides* and *L. windarraensis* the most abundant viral family was the Polydnaviridae-like group (23.2% and 55.1% respectively), and for *L. lapostaae* the most abundant viral family was Myoviridae (38.5%) (Table 3.1). For the *Paroster* species from Sturt Meadows, 56 viral families were identified (Table S3.2). When considering the most abundant species, the top seven families made up 94.8% of the total average viral abundances (Table 3.2), For *P. macrosturtensis, P. mesosturtensis* and *P. microsturtensis*, the most abundant viral family, was Herpesviridae (70.2%, 80.0%, and 81.1% respectively) (Table 3.2).

Table 3.1 Average viral abundances for *Limbodessus* species (Laverton calcrete) by family

Known Host	Family	Average Abundance (%)
Invertebrate	Polydnaviridae	22.33
Vertebrate	Herpesviridae	15.62
Phage	Myoviridae	14.72
Algae	Phycodnaviridae	10.63
Phage	Siphoviridae	7.05
Amoeba/Bicosoecid Flagellate	Mimiviridae	4.74
Invertebrate/Vertebrate	Poxviridae	4.67

Table 3.2 Average viral abundances for *Paroster* species (Sturt Meadows calcrete) by family

Known Host	Family	Average Abundance (%)
Vertebrate	Herpesviridae	74.82
Invertebrate	Baculoviridae	9.86
Phage	Myoviridae	3.13
Invertebrate	Polydnaviridae	2.27
Phage	Siphoviridae	1.83
Phage	Microviridae	1.72
Amoeba/Bicosoecid Flagellate	Mimiviridae	1.03

The Pilou evenness value (J) was constrained between 0-1. J was 0.13 at Laverton, while at Sturt Meadows it was 0.07 indicating a bias among virome communities. The beetle species at Laverton had a collective Shannon-Wiener index (H') value of 4.18, while the Sturt Meadows beetles had an H' of 1.68, indicating a substantial difference in diversity and evenness between the sites. Comparisons between the viromes of beetles from different sites and developmental stages employing multi-dimensional scaling showed that the beetle gut viromes cluster by location (Fig. 3.2a).

SIMPER analysis was used to calculate the dissimilarity between species (Table S3.3), with the average dissimilarity between the species at Laverton ranging from 60.01% to 62.07%, whereas the average dissimilarity among species at Sturt Meadows ranged from 25.14% to 64.18% (Table S3.3). The average group dissimilarity among beetle species from the two sites ranged from 57.04% to 79.00%. The top viral species that drove the dissimilarity among the beetle species from Laverton were in the families Myoviridae, Asfarviridae, and the Polydnaviridae-like group. The top viral species that drove the dissimilarity among the beetle species from Sturt Meadows were in the families Herpviridae, the Polydnaviridae-like group and Pandoraviridae. The top viral species that drove the dissimilarity among the beetle species between the two sites were Herpviridae, Polydnaviridae and Myoviridae.

3.2.3 Viral results by host species

Comparisons among the six beetle species suggest that gut virome diversity was associated with the host species (Fig. 3.2c). SIMPER analysis was also performed to assess within species similarity for species where enough individuals had been sampled (Table S3.4). Average similarity within *L. palmulaoides* was 42.9%, with similarity contributions above 1% accounting for 33.9% of the total contributions (Table S3.4). Average similarity within *P. macrosturtensis* was 37.9% with similarity contributions greater than 1% accounting for 21.6% of the total contributions (Table S3.4). Finally, *P. microsturtensis* had a within taxon average similarity of 72.2% with similarity contributions above 1% accounting for 54.1% of total contributions (Table S3.4). The top viral species that were driving the similarity within beetle species came from the Polydnaviridae-like group and Baculoviridae for *L. palmulaoides*, Herpesviridae and

Mimiviridae for *P. macrosturtesis*, and Herpesviridae and Baculoviridae for *P. microsturtensis.*

3.2.4 Viral results by developmental stage

Analysis of the viral species assigned to the gut viromes showed that three families were unique to the larval stage, while eight families were unique to the adult stage and 50 families were shared between both developmental stages (Table S3.2). When combining the developmental stages from both sites for larvae, the most abundant host association was vertebrate (60.5%), followed by phage (19.1%), and then invertebrate (10.6%) (Fig. 3.1b). For the adult stage, the most abundant host association was again vertebrate (71.6%), however, the next highest abundance was invertebrate (15.5%), followed by phage (5.8%) (Fig. 3.1b). Four families containing giant viruses were identified, additionally, a family of viruses that typically co-infects organisms with giant viruses, was also found. Pandoraviridae, Mimiviridae, Marseilleviridae and Pithoviridae were found in all individuals sequenced, while virophages were only found in the adult stage of *P. macrosturtensis*. Additionally, all five families including the virophages were found in the single amphipod sequenced. The giant virus sequences made up more of the top 100 viral sequences for the larval stage (3.1%) than in adults (0.9%). Additionally, a comparison of adults and larvae of *P. macrosturtensis* showed a large change of viral species, with the larva being dominated by phage (73.1%), followed by viruses associated with amoebas (8.4%), while the adults were dominated by vertebrate associated viruses (80.0%) followed by invertebrate viruses (12.5%) (Fig. 3.1c). Combining the site data, larval beetles had an H’ value of 2.59 while the adults had an H’ value of 1.9, indicating a substantial difference in virome diversity between life-history stages. However, the evenness was not considerably different between the stages; larval beetles had an evenness value of 0.09, while the adults had an evenness value of 0.08.

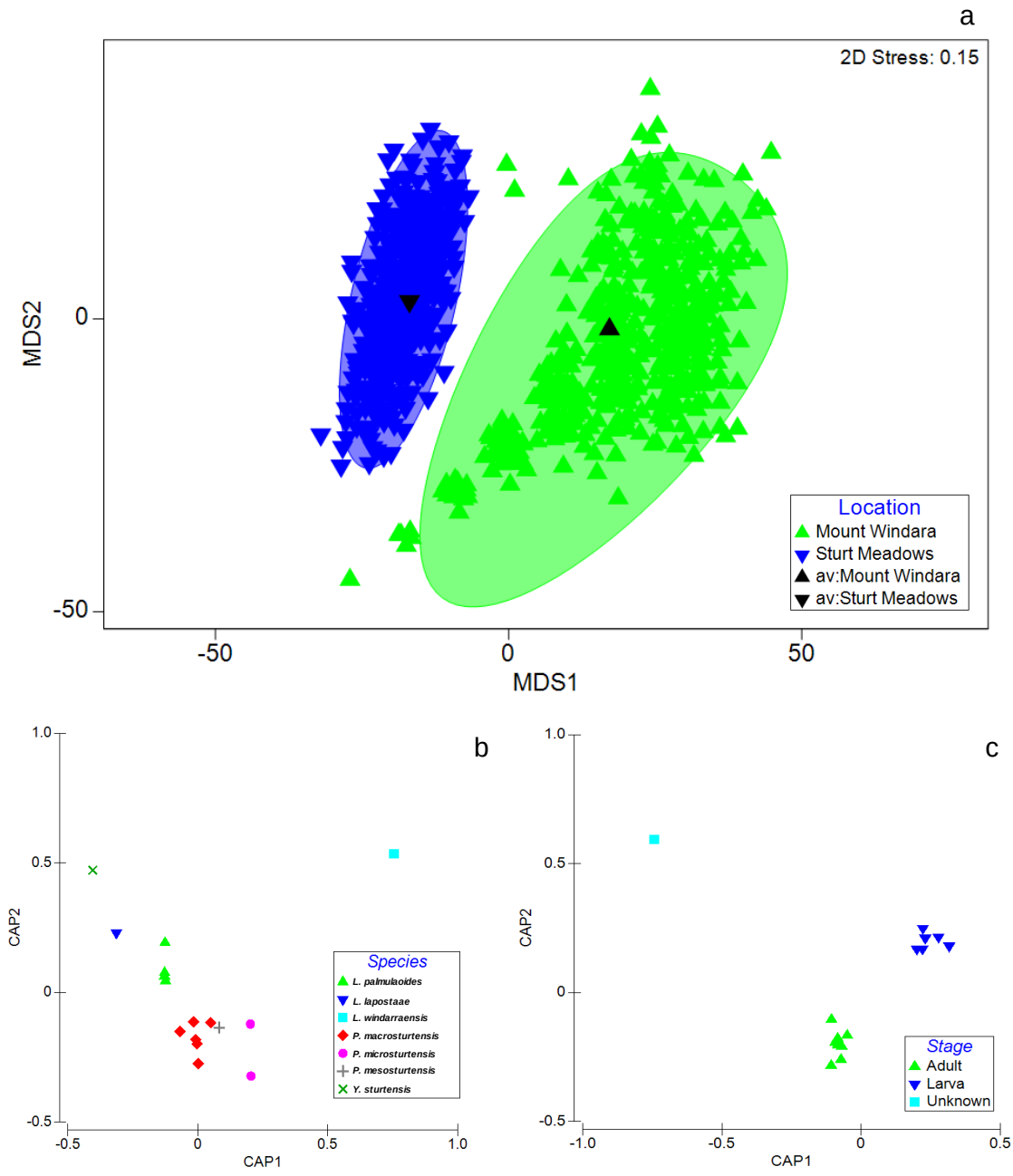

Fig. 3.2 (a) Multidimensional scaling analysis of viral species by beetle location Confidence intervals for the ellipses were 95% bootstrap regions. Symbols are the bootstrap samples and the black Av symbols are the average for each location. (b) Canonical Analysis of Principal coordinates (CAP) analysis of beetles by species, with amphipod as an outgroup; beetles are in the genera *Limbodessus* and *Paroster*, the amphipod comes from the genus *Yilgarniella* (c) CAP analysis of beetle species by life stage. Beetles are classified as either adult or unknown, the amphipod is classified as unknown.

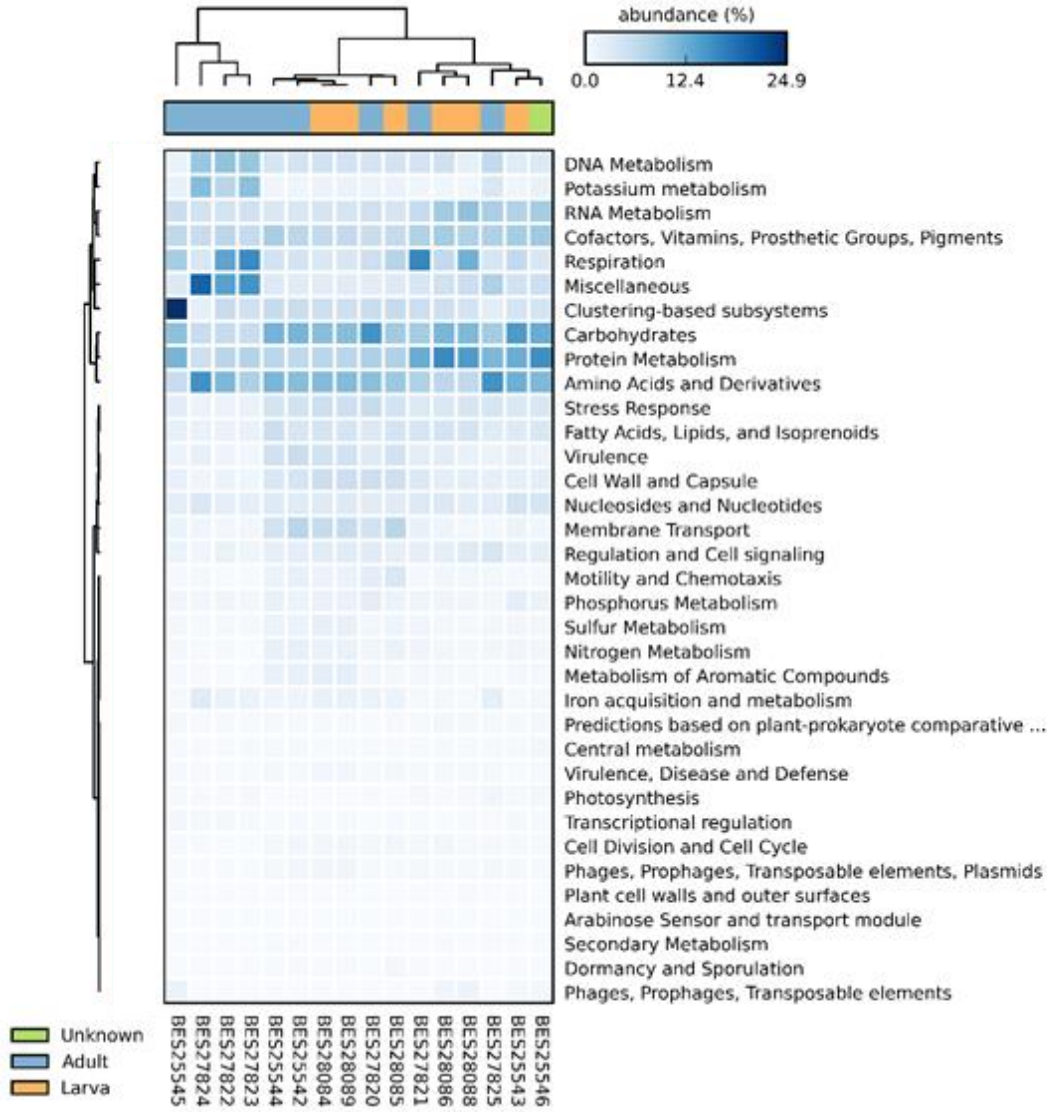

Fig. 3.3 Heatmap of functional genes at subsystem level 1; ANOVA and Welch's t-test (uncorrected) were performed on the data and the data were filtered by p-value of > 0.05 and an effect size < 0.8. Dendrograms show hierarchical clustering of samples and functional genes. The colours across the top indicate sample location with blue indicating beetles from Laverton, Orange indicating beetles from Sturt Meadows and Green indicating the amphipod from Sturt Meadows

3.2.5 FUNCTIONAL ANALYSES

Functional gene analysis was performed on the viral sequences to two different hierarchical levels of SEED classification, which organises genes by functional classification into a five level hierarchy (Overbeek *et al*., 2013): the subsystem level 1 and the subsystem level 2 (Table S3.5). At subsystem level 1, the analyses returned 36 different subsystems, 17 of which made up the top 10 most abundant functions returned over the 16 viral metagenomes (Fig. 3.3). Carbohydrates, amino acids and derivatives, protein metabolism, respiration, miscellaneous and clustering-based subsystems were

the most abundant functions in at least one metagenome. At subsystem level 2, STAMP analyses returned four important functions when comparing between beetle genera: inorganic and organic sulfur assimilation, co-enzyme A and protein secretion system type III (Fig. 3.4a). When comparing developmental stage, 20 important functions at subsystem level 2 were returned (Fig. 3.4b). These functions were represented by eight different parent categories at subsystem level 1: clustering based subsystems; membrane transport; amino acids and derivatives; DNA metabolism; cofactors, vitamins, prosthetic groups and pigments; acids, lipids, and isoprenoids; metabolism of aromatic compounds; and phages, prophages, transposable elements.

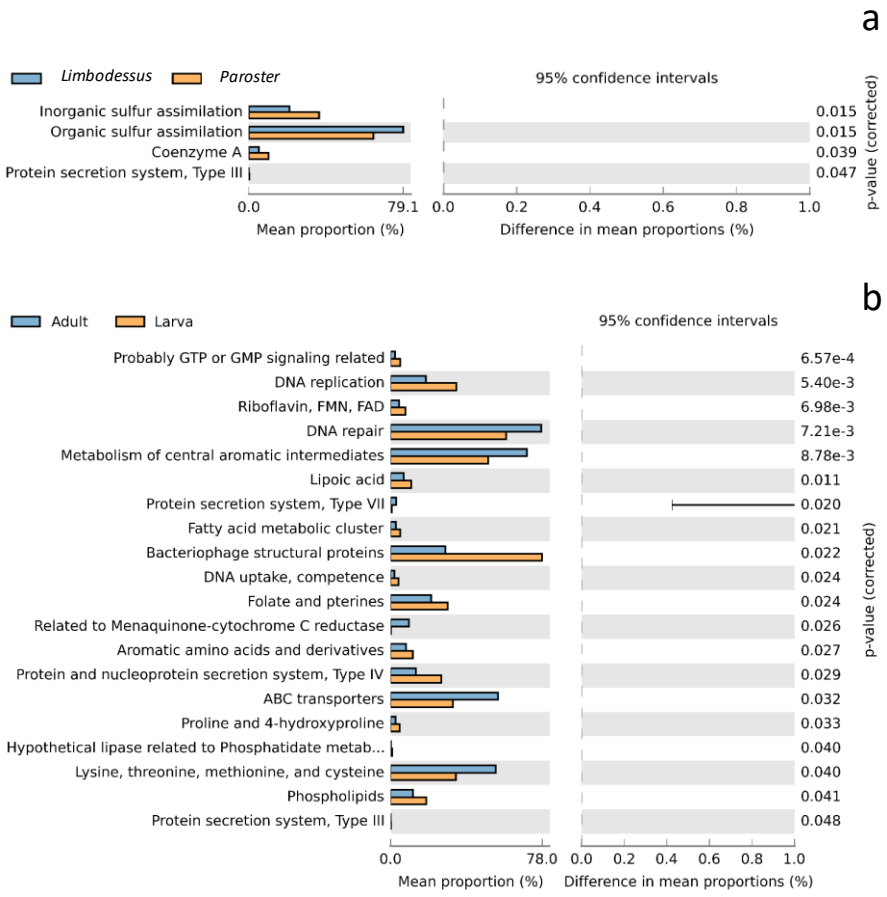

Fig. 3.4 Pairwise comparison bar plots of functional genes using welch's t-testto determine significant differences in genes between the two groups, by (a) life stage and (b) genus. Data were filtered by p-value >0.05.

3.3 Discussion

Using metagenomic analysis we have characterised six stygofaunal beetle species, from two different sites and two developmental stages, with an amphipod crustacean for comparison, for a total of 16 metagenomes. In this first study of the viral communities of stygofaunal beetles we show that the virome is influenced by location (Fig. 3.2a), and by the developmental stage of the beetles (Figs. 3.1b-c, 3.2c). Additionally, there is some evidences that the virome is also influenced by beetle species (Fig. 3.2b). The differences between beetle viromes within each calcrete suggests that the different beetle species and life stages are interacting within the same environmental conditions in significantly different ways and, potentially, their viral communities are being altered in response to genetic and physiological differences between species. However due to the uneven sampling between the different species, it is difficult to make any strong conclusions about the role of species in forming the virome beyond correlation at this stage. Herpesviridae is the first and second most common viral family found at Sturt Meadows and Laverton respectively (Tables 3.1, 3.2). While Herpesviridae viruses are not commonly found in water, this viral family has been found in several aquatic (e.g. *Hydra*) and terrestrial animal gut viromes (Grasis *et al*., 2014; Fawaz *et al*., 2016). Grasis *et al*. (2014) proposed that *Hydra* are actively selecting for the virus, potentially because they confer some benefit to them (Grasis *et al*., 2014). The presence of vertebrates on the surface, particularly cattle at the Sturt Meadows pastoral property, may result in a higher than normal *Herpes* viral load in the groundwater which may increase the viral load in the beetles (Nandi *et al*., 2009). The large number of vertebrate viruses found in these beetles has not been found in either of the model organisms, *D. melanogaster* or *C. elegans* (Félix *et al*., 2011; Unckless, 2011), however, their presence in other species in the wild, such as *Hydra*, suggests that the presence of vertebrate viruses might be widespread in wild invertebrate populations (Grasis *et al*., 2014). Another possibility is that these sequences are misallocated to Herpesviridae, and other vertebrate viruses, as these are the closest genomes available on the database. Thus, they may represent novel viral groups not present in any databases.

Polydnaviridae are insect viruses that are known to be exclusively symbiotic with wasps that are endoparasitic on lepidopteran larval hosts (Strand and Drezen, 2012). As these beetles moved underground between 5-10 million years ago (Cooper *et*

al., 2002) and the environment that the beetles live in is aquatic and below the surface under carbonate rock, the possibility that they are parasitised by wasps is highly unlikely. The most likely explanation is that this is a new virus species and the closest genomes already sequenced are from the family Polydnaviridae.

The family Phycodnaviridae was the third most abundant viral family in the Laverton calcrete beetles and the sixth most abundant family in the Sturt Meadows beetles (Fig. 3.1a). This family is known to infect marine and freshwater algae and plays a role in regulating the growth of their algal host by causing the death and lysis of their host, which can release nutrients into the water (Sigee, 2005). As subterranean environments are known to be nutrient poor, with the absence of photosynthetic primary producers (Gibert and Deharveng, 2002), the abundance of Phycodnaviridae could be evidence of algae as a nutrient source. These viruses are likely to be present in algae and the surrounding water following lysis (Wilson *et al*., 2009). Therefore, it would allow their accumulation within the beetles due to general exposure and consumption of algae by organisms lower in the food chain. Additionally, the viruses could enter the beetles by viral attachment directly on the surface of the beetles' prey. Aquifer recharge, and hence nutrient supplementation from algae, occurs through rainfall events that are rare in this arid landscape of Western Australia. However, flooding of the surface area by episodic downpours (Humphreys 2001) may permit algal blooms in the surface water prior to infiltration or flooding from the salt lake (Humphreys *et al*., 2009).

Giant viruses, such as in the families Mimiviridae and Pandoraviridae were abundant in adult and larval beetles (Fig. 3.1b). Giant viruses may be a normal part of human gut viromes, but their presence in insects may be underestimated due to the size selective filters that are used before sequencing (Popgeorgiev *et al*., 2013). Despite a lack of light in the groundwater aquifers, plant roots often occur in abundance. It is thought that the plant roots in the calcretes are supplying a signification amount of primary energy into the system (Jasinska *et al*., 1996) and therefore plant viruses were expected to be present in the beetle viromes. However, the beetles at both Laverton and Sturt Meadows have a low abundance of plant viruses, in their viromes (Fig. 3.1a). These viruses came from two different families: Potyviridae, Tymoviridae. However, of the plant viruses the beetles carried there was a higher proportion of plant viruses that are known to have invertebrates as vectors (Caulimoviridae, Tospoviridae) than any other type of plant virus. This low viral load of plant only viruses could be due to the

fact that beetles are the top predators in the system and there are too many intermediate steps between the plants and the beetles for the viruses to survive in such trophically distant hosts (Mortensen 1993). Another possibility could be due to sequence bias as this study sequenced DNA viruses while most plant and algal viruses are RNA viruses.

Analysis showed that viral communities differed significantly between the two sites (Figs. 3.1a, 3.2a). Coupled with the results of the Shannon-Weiner index analysis, these results suggest that the viruses in beetles from Laverton are significantly more diverse than those from Sturt Meadows. The lower viral diversity (H') of the beetles from Sturt Meadows is not due to them being extreme short range endemics as this is the case for the beetles at both sites. Therefore, most of the lower viral diversity must be explained by other factors. This said, there is evidence that all the beetles at Sturt Meadows evolved from a single stygobiont ancestor, and speciated within the calcrete (Leijs *et al*., 2012; Langille pers. Comm.) further reducing their exposure to different viruses, which may explain some of the lower diversity at this site. Beetle viromes also clustered strongly by developmental stage (Figs. 3.1b, 3.2b), suggesting that the viral communities are highly influenced by life stage. These beetles are stygobiontic at every life history stage, moult numerous times during development, shedding the lining of their foregut and hindgut each time (Engel and Moran, 2013). Additionally, with each stage of development, the gut is remodelled significantly, therefore providing an unstable habitat for the gut microbiota (Engel and Moran, 2013).

Previous studies on the gut microbiome of insects have suggested that diet can play a large role in bacterial composition and diversity (Yun *et al*., 2014). Yun *et al*. (2014) demonstrated that the Shannon diversity index changed depending on the diet of the insect, with omnivores having a higher diversity index than herbivores or carnivores. Therefore, the larval stages might have a different diet to the adult beetles, which could explain the different diversity indexes. The different feeding methods; adults masticate their prey whereas larvae inject digestive enzymes and suck out their prey, which also indicates that the larva have a different diet to the adult beetles, could explain the different viral compositions between beetle developmental stages. This virome difference is most obviously seen in *P. macros*turtensis (Fig. 3.1c) larvae where their feeding mechanism could render them more likely to be exposed to the viruses of the micro-invertebrates they feed on, which could account for the high numbers of bacterial, amoebal and algal viruses present in their gut viromes.

Analyses showed that the viral functional genes were stable between sites and individuals (Fig. 3.3), with the top 10 viral functional genes at both sites found to be the same, but in a slightly different order of expression. The most important subsystem functional differences between the two sites were involved with organic and inorganic sulphur assimilation, which may be due more to the geology of the two sites than the viral species concerned (Fig. 3.4b). While some of the most important functional differences between the two developmental stages involve bacteriophage structural proteins (Fig. 3.4a), this is most likely due to the larger number of bacterial phage in larvae compared to the adults.

The methods used in this study are based on the currently available sequenced genomes in public databases. As many taxa are not currently represented in the genome databases, it is possible that highly abundant, but un-sequenced organisms, have been excluded from the list of abundant taxa. Additionally, although taxa have been matched to what is available on sequence databases it is highly likely, given that both the groundwater environment and the stygobiontic beetles have not been previously studied, that numerous new viral species have been encountered. As such, viral species have been matched to what is the closest match available from the database, which may not necessarily be the same species. However, all of the metagenomes used in this study have the same issues in this regards and therefore should not present an issue in comparisons among the sites, beetle species and life stages.

3.4 Conclusions

Stygobitic diving beetles host complex viral communities, which vary greatly by location. The gut virome composition, while unique to each individual, varies with host developmental stage, and location. The viral families Herpesviridae and Polydnaviridae largely drove the differences found between the two locations. The results suggest that while these beetles are occupying the same environmental locations at each calcrete, life style factors, such as diet, are likely driving the virome differences between different developmental stages. These results provided further insights into wild insect viromes and their composition within a subterranean environment.

3.5 Experimental Procedures

The study sites were two groundwater calcrete deposits in the Yilgarn region of central Western Australia; one calcrete is located on the Sturt Meadows pastoral station (28.7155° S, 120.8931° E) and the second calcrete is located on Laverton station

(28.3983° S, 122.2038° E). At both sites, the surface vegetation is an open *Acacia* woodland with lowland shrubs. In a few places, the calcrete is exposed on the surface, whereas for the most part the top of the calcrete is up to 2 m below the surface (Allford *et al*., 2008; Bradford *et al*., 2013). The Sturt Meadows calcrete stygofauna consists of 18 known macroinvertebrate taxa, including a sister species triplet of dytiscid diving beetles; *P. macrosturtensis*, *P. mesosturtensis*, and *P. microsturtensis* (Watts and Humphreys, 2006). The two calcretes both drain towards salt lakes, which represent the base level of the regional groundwater. The Laverton calcrete stygofauna is more diverse than that at Sturt Meadows and has yet to be fully characterised, but it includes three dytiscid diving beetle species *L. palmulaoides*, *L. lapostaae,* and *L. windarraensis* (Watts and Humphreys, 2006). The Laverton calcrete has four main sampling sites, referred to as Windarra, Shady Well, Quandong and Erlistoun. For this study all sampling occurred at Windarra.

3.5.1 Specimen collection

Fifteen individuals comprising six beetle species from two genera were collected from Sturt Meadows and Laverton (Table 3.3). *Limbodessus palmulaoides*, *P. macrosturtensis*, *P. microsturtensis* adults and larvae and a *P. mesosturtensis, L. lapostaae,* and *L. windarraensis* adults were collected. Additionally, an amphipod (*Yilgarniella sturtensis*) from Sturt Meadows was also collected for comparison. Specimens were identified and all but two, were preserved in liquid Nitrogen and stored at -80 °C at the South Australian Regional Facility for Molecular Ecology and Evolution (University of Adelaide, South Australia). An additional specimen (BES25545) was preserved in 100% ethanol and then stored at -20 °C. Another specimen of *P. macrosturtensis* (BES25542) was brought back alive from the field and maintained at 25 °C in the original groundwater collected from Sturt Meadows. It was fed an amphipod, just prior to it being euthenased with 100% ethanol and its DNA extracted. The collection localities, dates and collectors are listed in Table 3.3.

Table 3.3 Collection information for the beetles and amphipod sequenced in this study

Genus	Species	Life Stage	Location	Collection date	Collector	Sample Number
Limbodessus	*palmulaoides*	adult	Laverton, Western Australia	Sep-15	K.K. Jones; S.J.B. Cooper; B. Langille	25542
Limbodessus	*palmulaoides*	adult	Laverton, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper;	27821
Limbodessus	*palmulaoides*	larva	Laverton, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper;	28086
Limbodessus	*palmulaoides*	larva	Laverton, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper;	28088
Limbodessus	*lapostaae*	adult	Laverton, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper;	27825
Limbodessus	*windarraensis*	adult	Laverton, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper;	27820
Paroster	*macrosturtensis*	adult	Sturt Meadows, Western Australia	Sep-15	K.K. Jones; S.J.B. Cooper; B. Langille	25544
Paroster	*macrosturtensis*	adult	Sturt Meadows, Western Australia	Nov-06	W.F. Humphreys; S.J.B. Cooper; A. Allford	25545
Paroster	*macrosturtensis*	adult	Sturt Meadows, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper; J. Hyde	27822
Paroster	*macrosturtensis*	larva	Sturt Meadows, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper; J. Hyde	28084
Paroster	*macrosturtensis*	larva	Sturt Meadows, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper; J. Hyde	28085
Paroster	*macrosturtensis*	larva	Sturt Meadows, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper; J. Hyde	28089
Paroster	*mesosturtensis*	adult	Sturt Meadows, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper; J. Hyde	27823
Paroster	*microsturtensis*	larva	Sturt Meadows, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper; J. Hyde	25543
Paroster	*microsturtensis*	adult	Sturt Meadows, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper; J. Hyde	27824
Yilgarniella	*sturtensis*	unknown	Sturt Meadows, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper; J. Hyde	25546

3.5.2 DNA extraction and sequencing

Genomic DNA was extracted from whole specimens using a modified Gentra Pure-Gene DNA purification kit protocol (Gentra systems Minneapolis MN, USA). All extractions were performed inside a UV hood and the individual beetles were placed under UV for 30 s in an effort to reduce the amount of surface microbial contamination. The genomic DNA was sent for library construction and sequencing at the Australian Genome Research Facility (AGRF) and libraries were prepared using a Nextera DNA library prep kit (Caruccio, 2011). Each library contained only one specimen and between five and six libraries were run in a lane. The DNA was sequenced using an Illumina Miseq, and yielded 150 or 300 bp paired end reads.

3.5.3 Taxonomic analyses

Sequenced DNA in FASTQ format was quality filtered and trimmed. Additionally, adapters, unknown terminal bases, poly-A tails, and low-quality 3′ read regions were removed via FqTrim (Pertea, 2015). Paired-ends were joined and combined with the unjoined forward reads. RiboPicker was used to remove 16s, 18s, 28s, and 5.8s ribosomal RNA to increase the quality of sequences and relevance of the results (Schmieder, *et al*., 2012). Bowtie 2 was used to remove human sequences using the *H. sapiens* UCSC hg18 Bowtie 2 index (Deng *et al*., 2015; Langmead and Salzberg, 2012). Sequences were dereplicated for 100% sequence similarity using USEARCH (Edgar, 2010). Dereplicated sequences were assembled using Velvet, with a k-mer length of 49 (Zerbino and Birney 2008). These contigs were then analyzed via tBLASTx with the NCBI viral RefSeq database using an e-value of 10^{-7} (Deng *et al*., 2015). Viral taxonomic representation was determined via Galaxy (Blankenberg *et al*., 2010; Giardine *et al*., 2005; Goecks, Nekrutenko, and Taylor, 2010). Velvet was employed for assembly due to its success with previous Illumina viral metagenomic datasets, showing highly reliable contig construction using short read sequencing, as well as its compatibility with the file formats obtained from postprocessing (Vázquez-Castellanos *et al.*, 2014). VelvetOptimiser was employed to determine optimum assembly parameters. Hash values from 20 to 399 were explored with a k-mer length of 49 chosen to allow a balance between specificity and sensitivity. The short paired fasta file option was used with a minimum contig length of 100 and a coverage cut-off value of 15×. Velvet was selected over MetaVelvet as no significant difference was found between them (Vázquez-Castellanos *et al*., 2014).

3.5.4 Statistical analysis

Data were analysed using PRIMER (v7) following square root transformation and for high abundance species down weighted using the default setting. Metric multidimensional scaling (MDS) analysis was performed to determine the similarity of the samples by location. The data were then presence/absence overall transformed and Canonical Analysis of Principal coordinates (CAP) was performed by life stage to determine similarity. Fifteen individuals were classified correctly (93.75%) and one was misclassified with an error rate of 6.25%. CAP was also performed by species to determine similarity. The total classified correctly was 12/16 (75%) with a misclassification error of 25%. Four of the six species were always classified correctly.

SIMPER analysis was performed to determine the relative similarity of species contribution between samples. A one-way analysis was performed with a cut off for low contributions of 90%; the factor group used was beetle species. Species diversity was calculated using the Shannon-Wiener index (H') and evenness was calculated using Pilou evenness (J). Evenness is a measure of how homogenous a community or ecosystem is; a community where all species are equally common has a high degree of evenness. The Shannon-Wiener index values of H' range from 0-5; they typically are 1.5 to 3.5 and are rarely above 4 (Shannon, 1948). The H' value increases as both the richness and evenness of the community increases. The Shannon-Wiener index (H') was calculated to estimate the diversity in each of the two calcretes, it uses abundance and evenness to calculate the diversity of the ecosystem. Both were calculated in R using the Vegan package. The functional data were further analysed in STAMP 2.1.3 (Parks *et al*., 2014) in which Welch's t-test was used to compare the data by genus and by developmental stage using Clustering-based subsystem level two (Fig. 3.8). The data were filtered removing all features with $p > 0.05$.

Acknowledgments

The authors would like to thank Flora, Peter and Paul Axford of the Sturt Meadows Station for providing both accommodation and access to their property. We would also like to thank Dr Robert Edwards and Dr Elizabeth Dinsdale for their advice during this project. This research was funded by Australian Research Council linkage grants (LP0348753, LP 100200494 and LP140100555) to SJBC, WFH, ADA and PM, with industry partners Newmont Australia, Placer Dome Asia Pacific, Minara Resources Limited, the Western Australian Museum, the South Australian Museum, Rio Tinto, Biota Environmental Sciences, Bennelongia, Department of Biodiversity, Conservation and Attractions (WA).

References

Allford, A., Cooper, S.J.B., Humphreys, W.F., and Austin, A.D. (2008) Diversity and distribution of groundwater fauna in a calcrete aquifer: does sampling method influence the story? *Invertebrate Systematics* 22: 127-138.

Bradford, T.M., Adams, M., Guzik, M.T., Humphreys, W.F., Austin, A.D., and Cooper, S.J.B. (2013) Patterns of population genetic variation in sympatric chiltoniid amphipods within a calcrete aquifer reveal a dynamic subterranean environment. *Heredity* 111: 77.

Caruccio, N. (2011) Preparation of next-generation sequencing libraries using Nextera™ technology: simultaneous DNA fragmentation and adaptor tagging by in vitro transposition. High-Throughput Next Generation Sequencing: *Methods and Applications*: 241-255.

Chandler, J.A., Lang, J.M., Bhatnagar, S., Eisen, J.A., and Kopp, A. (2011) Bacterial communities of diverse Drosophila species: ecological context of a host–microbe model system. *PLoS Genetics* 7: e1002272.

Columpsi, P., Sacchi, P., Zuccaro, V., Cima, S., Sarda, C., Mariani, M. *et al*. (2016) Beyond the gut bacterial microbiota: The gut virome. *Journal of Medical Virology* 88: 1467-1472.

Cooper, S.J.B, Hinze, S., Leys, R., Watts, C.H.S., and Humphreys, W.F. (2002) Islands under the desert: molecular systematics and evolutionary origins of stygobitic water beetles (Coleoptera: Dytiscidae) from central Western Australia. *Invertebrate Systematics* 16: 589-590.

Dillon, R.J., and Dillon, V.M. (2004) The gut bacteria of insects: nonpathogenic interactions. *Annual Reviews in Entomology* 49: 71-92.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, 2460–2461. doi:10.1093/bioinformatics/btq461

Engel, P., and Moran, N.A. (2013) The gut microbiota of insects – diversity in structure and function. *FEMS Microbiology Reviews* 37: 699-735.

Fawaz, M., Vijayakumar, P., Mishra, A., Gandhale, P.N., Dutta, R., Kamble, N.M. *et al*. (2016) Duck gut viral metagenome analysis captures snapshot of viral diversity. *Gut Pathogens* 8: 30.

Félix, M.-A., Ashe, A., Piffaretti, J., Wu, G., Nuez, I., Bélicard, T. *et al*. (2011) Natural and experimental infection of *Caenorhabditis* nematodes by novel viruses related to nodaviruses. *PLoS Biology* 9: e1000586.

Gibert, J., and Deharveng, L. (2002) Subterranean Ecosystems: A Truncated Functional Biodiversity *BioScience* 52: 473-481.

Grasis, J.A., Lachnit, T., Anton-Erxleben, F., Lim, Y.W., Schmieder, R., Fraune, S. *et al*. (2014) Species-specific viromes in the ancestral holobiont Hydra. *PLoS One* 9: e109952.

Grover, M., and Kashyap, P.C. (2014) Germ-free mice as a model to study effect of gut microbiota on host physiology. *Neurogastroenterology and Motility* 26: 745-748.

Humphreys, W.F. (2001). Groundwater calcrete aquifers in the Australian arid zone: the context to an unfolding plethora of stygal biodiversity. Pp 63-83 in Subterranean Biology in Australia 2000, W.F. Humphreys and M.S. Harvey (eds). *Records of the Western Australian Museum,* Supplement No. 64.

Humphreys, W.F., Watts, C.H.S., Cooper, S.J.B., Leijs, R. (2009) Groundwater estuaries of salt lakes: buried pools of endemic biodiversity on the western plateau, Australia. *Hydrobiologia* 626: 79–95.

Jasinska, E. J., Knott, B., McComb, A. J., (1996). Root Mats in Ground Water: A Fauna-Rich Cave Habitat. *Journal of the North American Benthological Society* 15: 508-519.

Kostic, A.D., Howitt, M.R., and Garrett, W.S. (2013) Exploring host–microbiota interactions in animal models and humans. *Genes and Development* 27: 701-718.

Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357–359.

Lee, W.-J., and Brey, P.T. (2013) How microbiomes influence metazoan development: insights from history and *Drosophila* modeling of gut-microbe interactions. *Annual Review of Cell and Developmental Biology* 29: 571-592.

Leijs, R., Van Nes, E.H., Watts, C.H.S., Cooper, S.J.B., Humphreys, W.F., and Hogendoorn, K. (2012) Evolution of blind beetles in isolated aquifers: a test of alternative modes of speciation. *PLoS One* 7: e34260.

Leys, R., Watts, C.H.S., Cooper, S.J.B, and Humphreys, W.F. (2003) Evolution of subterranean diving beetles (Coleoptera: Dytiscidae Hydroporini, Bidessini) in the arid zone of Australia. *Evolution* 57: 2819-2834.

Leys, R., Cooper, S.J.B, Strecker, U., and Wilkens, H. (2005) Regressive evolution of an eye pigment gene in independently evolved eyeless subterranean diving beetles. *Biology Letters* 1: 496-499.

Ma, D., Storelli, G., Mitchell, M., and Leulier, F. (2015) Studying host-microbiota mutualism in Drosophila: harnessing the power of gnotobiotic flies. *Journal of Biomedical Science*38: 285-293.

Metchnikoff, E. (1901) L'immunité dans les maladies infectieuses: Masson.

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D. *et al*. (2011) The human gut virome: inter-individual variation and dynamic response to diet. *Genome Research* 21: 1616-1625.

Nandi, S., Kumar, M., Manohar, M., and Chauhan, R. (2009) Bovine herpes virus infections in cattle. *Animal Health Research Reviews* 10: 85-98.

Overbeek, R., R. Olson, G. D. Pusch, G. J. Olsen, J. J. Davis, T. Disz, R. A. Edwards, S. Gerdes, B. Parrello and M. Shukla (2013). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research* 42: 206-214.

Parks, D.H., Tyson, G.W., Hugenholtz, P., and Beiko, R.G. (2014) STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30: 3123-3124.

Pasteur, L. (1870) Etudes sur la maladie des vers a soie: Gauthier-Villars. Paris

Pasteur, L. (1885) Methode pour prevenir la rage apres morsure. *Comptes Rendus de l'Académie des Sciences*, 101Pertea, G. (2015). Fqtrim: v0.9.4 release. *Zenodo*. 10.5281/zenodo.20552 Retrieved from http://dx.doi.org/10.5281/zenodo.20552

Popgeorgiev, N., Temmam, S., Raoult, D., and Desnues, C. (2013) Describing the silent human virome with an emphasis on giant viruses. *Intervirology* 56: 395-412.

Pride, D.T., Salzman, J., Haynes, M., Rohwer, F., Davis-Long, C., White III, R.A. *et al*. (2012) Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. *The ISME Journal* 6: 915.

Rosales SM, Vega Thurber R (2015) Correction: Brain Meta-Transcriptomics from Harbor Seals to Infer the Role of the Microbiome and Virome in a Stranding Event. *PLoS One* 10(12): e0146208.https://doi.org/10.1371/journal.pone.0146208

Rosshart, S.P., Vassallo, B.G., Angeletti, D., Hutchinson, D.S., Morgan, A.P., Takeda, K. *et al*. (2017) Wild mouse gut microbiota promotes host fitness and improves disease resistance. *Cell* 171: 1015-1028. e1013.

Schmieder, R., Lim, Y. W., and Edwards, R. (2012). Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics*, 2012, 433–435. [PMID: 22155869]

Sekirov, I., Russell, S.L., Antunes, L.C.M., and Finlay, B.B. (2010) Gut microbiota in health and disease. *Physiological Reviews* 90: 859-904.

Shannon, C.E. (1948) A mathematical theory of communication. *Bell System Technical Journal*27: 623-656.

Sigee, D.C. (2005) Viruses: Major Parasites in the Freshwater Environment. In *Freshwater Microbiology: Biodiversity and Dynamic Interactions of Microorganisms in the Aquatic Environment*. New Delhi, India: Wiley, pp. 339-369.

Strand, M., and Drezen, J. (2012) Family polydnaviridae. *Virus taxonomy: ninth report of the international committee on taxonomy of viruses* Amsterdam: Elsevier: 237-248.

Tierney, S.M., Cooper, S.J.B., Saint, K.M., Bertozzi, T., Hyde, J., Humphreys, W.F., and Austin, A.D. (2015) Opsin transcripts of predatory diving beetles: a comparison of surface and subterranean photic niches. *Royal Society Open Science* 2: 140386.

Unckless, R.L. (2011) A DNA virus of Drosophila. *PLoS One* 6: e26564.

Wang, Y., Gilbreath III, T.M., Kukutla, P., Yan, G., and Xu, J. (2011) Dynamic gut microbiome across life history of the malaria mosquito *Anopheles gambiae* in Kenya. *PloS One* 6: e24767.

Vázquez-Castellanos, J. F., García-López, R., Pérez-Brocal, V., Pignatelli, M., and Moya, A. (2014). Comparison of different assembly and annotation tools on analysis of simulated viral metagenomic communities in the gut. *BMC Genomics*, 15, 1–20.

Watts, C.H.S., and Humphreys, W.F. (2004) Thirteen new Dytiscidae (Coleoptera) of the genera *Boongurrus* Larson, *Tjirtudessus* Watts and Humphreys and *Nirripirti* Watts and Humphreys, from underground waters in Australia. *Transactions of the Royal Society of South Australia* 128: 99-129.

Watts, C.H.S., and Humphreys, W.F. (2006) Twenty-Six New Dytiscidae (Coleoptera) of the Genera *Limbodessus* Guignot and *Nirripirti* Watts and Humphreys, from Underground Waters in Australia. *Transactions of the Royal Society of South Australia* 130: 123-185.

Watts, C.H.S., and Humphreys, W.F. (2009) Fourteen new Dytiscidae (Coleoptera) of the genera *Limbodessus* Guignot, *Paroster* Sharp, and *Exocelina* Broun from underground waters in Australia. *Transactions of the Royal Society of South Australia* 133: 62-107.

Wilson, W., Van Etten, J.L., and Allen, M. (2009) The Phycodnaviridae: the story of how tiny giants rule the world. (eds Van Etten, J) *Lesser Known Large dsDNA Viruses*: Lincon NE, Springer, pp. 1-42.

Yun, J.-H., Roh, S.W., Whon, T.W., Jung, M.-J., Kim, M.-S., Park, D.-S. *et al*. (2014) Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Applied and Environmental Microbiology* 80: 5254-5264.

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18, 821–829.

Supplementary Tables

Table S3.1 Information on sequence numbers and data percentages kept after data went through the quality control pipeline

Sample	Genus	Total sequences (raw)	FqTrim (%)	Ribopicker (%)	Bowtie2 (%)	USEARCH (%)
BES27820	*Limbodessus*	1,082,334	0.27	0.02	0.02	0.53
BES27821	*Limbodessus*	1,320,946	0.01	0.02	0.01	0.87
BES27825	*Limbodessus*	1,476,630	0.46	0.02	0.01	0.66
BES28086	*Limbodessus*	3,598,466	0.01	0.07	0.05	0.47
BES28088	*Limbodessus*	2,897,727	0.01	0.10	0.01	0.38
BES25542	*Limbodessus*	4,417,482	0.02	0.21	0.13	0.06
		14,793,585	0.07	0.11	0.06	0.38
BES27823	*Paroster*	958,312	0.47	0.07	0.02	0.21
BES27824	*Paroster*	1,614,519	0.29	0.05	0.03	0.52
BES27822	*Paroster*	1,716,381	0.40	0.01	0.02	0.40
BES28085	*Paroster*	2,524,576	0.00	0.09	0.00	2.89
BES28084	*Paroster*	2,536,524	0.00	0.60	0.00	0.69
BES28089	*Paroster*	2,784,735	0.00	0.65	0.00	0.19
BES25543	*Paroster*	3,056,701	0.13	0.06	0.01	0.20
BES25544	*Paroster*	3,732,270	0.09	0.10	0.00	0.12
BES25545	*Paroster*	1,302,320	0.07	0.04	0.02	0.33
		20,226,338	0.12	0.21	0.01	0.62
BES25546	*Yilgarniella*	4,226,806	0.06	0.13	0.01	0.15

Table S3.2 Complete list of viral abundances for all viral families identified for all beetle individuals sequenced. Host associations were determined using the Virus-Host DB **(https://www.genome.jp/virushostdb/)**

	L. palmulaoides				*L. lapostaae*	*L. windarraensis*	*P. macrosturtensis*						*P. microsturtensis*		*P. mesosturtensis*	*Y. sturtensis*	
Family	25542	27821	28088	28086	27825	27820	25544	25545	27822	28085	28089	28084	27824	25543	27823	25546	Host Associat
Phycodnaviridae	20184	504	458	1271	448	1576	1980	1038	875	2145	2415	3420	538	2570	372	4406	Algae
Pandoraviridae	4816	147	347	649	196	34	3790	245	1636	425	1573	741	1376	4243	882	487	Amoeba
Pithoviridae	588	153	29	67	13	22	121	42	84	223	147	285	19	127	10	340	Amoeba
Marseilleviridae	450	437	95	106	97	68	151	110	570	174	96	123	162	170	283	59	Amoeba
unclassified ssDNA viruses	2	0	0	1	0	0	2	0	0	0	2	0	0	0	0	11	Amoeba
Mimiviridae	3655	1087	3075	1923	1093	58	3265	2376	1796	1547	2578	3178	385	2943	374	1531	Amoeba/Bico
Rudiviridae	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	8	Archaea
Lipothrixviridae	9	4	0	3	0	0	1	0	3	5	1	1	1	0	3	4	Archaea
Pleolipoviridae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	Archaea
unclassified archaeal viruses	34	0	0	0	0	0	15	0	0	27	13	27	0	0	0	1	Archaea
Bicaudaviridae	11	2	0	2	0	0	4	0	0	2	13	9	0	1	1	1	Archaea
Turriviridae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	Archaea
unclassified dsDNA viruses	444	69	51	81	823	332	146	169	433	131	82	152	332	197	198	1229	Fungi
Totiviridae	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	Fungi
Baculoviridae	999	1780	2383	2280	1169	155	18681	6866	58237	5	52	48	30887	34878	26101	1647	Invertebrate
Polydnaviridae	2342	29196	2909	3737	8807	4351	5271	3596	11373	0	17	1	7094	8267	4843	1490	Invertebrate
Nudiviridae	254	205	31	79	157	9	546	123	2506	45	91	165	1507	783	1039	803	Invertebrate

Nimaviridae	560	19	216	13	0	6	1	0	0	0	1	20	0	0	0	34	Invertebrate
Ascoviridae	293	39	55	63	16	11	135	9	15	141	115	142	24	45	5	29	Invertebrate
unassigned Mononegavirales	0	1	0	4	0	0	0	2	0	0	0	0	1	1	1	16	Invertebrate
unclassified dsRNA viruses	0	1	1	3	1	0	1	2	2	0	0	0	2	2	2	3	Invertebrate
Parvoviridae	35	46	54	45	3	9	3	3	5	0	0	0	1	1	2	0	Invertebrate
Phasmaviridae	3	1	3	4	1	0	0	0	5	0	0	0	0	4	0	0	Invertebrate
Iridoviridae	1222	525	462	463	3146	115	808	2179	2468	49	59	112	2835	1635	941	2532	Invertebrate/V
Poxviridae	4166	1360	1283	2299	1188	444	3224	2166	3412	209	1882	349	1756	3351	1420	1243	Invertebrate/V
Flaviviridae	0	0	0	1	0	0	0	2	0	0	0	0	1	0	1	1	Invertebrate/V
Phenuiviridae	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	Invertebrate/V
Peribunyaviridae	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	Invertebrate/V
Sphaerolipoviridae	106	0	0	0	0	0	39	0	0	0	20	24	0	2	6	0	Other
Rhabdoviridae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	Other
unclassified RNA viruses	10	21	10	9	2	4	0	0	0	0	3	0	0	0	2	9	Other
unclassified dsDNA viruses	637	70	48	68	820	329	196	161	386	342	268	374	326	140	196	1155	Other
Endornaviridae	4	0	0	0	0	0	4	0	0	0	1	0	0	0	0	0	Other
unclassified virophages	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	2	Other
unclassified viruses	253	23	11	54	4	4	108	16	19	497	126	119	7	31	4	34	Other
Siphoviridae	15181	275	119	401	178	48	6475	1536	1272	3471	9506	8154	814	1034	310	905	Phage
Microviridae	670	0	1096	1219	0	6	148	1290	6	12891	760	15229	0	257	0	566	Phage
Myoviridae	11047	807	161	527	21109	185	3793	247	13010	7341	15415	13823	880	419	844	543	Phage

Podoviridae	2301	223	9	127	631	14	537	105	38	842	2066	2376	43	46	21	301	Phage
unclassified bacterial viruses	837	42	12	50	8	8	185	4	11	1174	719	1261	5	23	5	14	Phage
unclassified dsDNA phages	292	5	9	14	2	3	94	0	1	291	207	267	0	0	0	11	Phage
Inoviridae	110	4	0	14	0	3	28	0	0	86	310	161	1	8	0	3	Phage
unclassified Caudovirales	306	3	1	0	1	0	101	0	0	89	116	180	1	0	0	1	Phage
Tectiviridae	4	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	Phage
Cystoviridae	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Phage
Tymoviridae	16	0	0	0	0	0	5	0	0	0	0	3	0	0	0	0	Plants
Potyviridae	11	3	1	4	0	1	2	0	1	3	3	11	0	1	0	0	Plants
Caulimoviridae	150	184	154	187	365	156	317	838	1309	0	0	0	1217	729	537	99	Plants And In
Tospoviridae	0	2	0	0	2	1	5	0	38	0	0	0	12	10	9	13	Plants And In
Lavidaviridae	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	1	Protists
Herpesviridae	22003	2383	245	1848	8700	732	191663	64852	426184	213	9613	874	175169	310933	153919	32398	Vertebrate
Asfarviridae	4	9	2	3	6409	3	53	6	679	10	2	0	329	86	50	868	Vertebrate
Alloherpesviridae	56	0	1	2	2	51	11	1	7	12	10	31	0	46	11	333	Vertebrate
Retroviridae	112	131	128	193	223	103	137	316	660	0	0	0	672	284	263	155	Vertebrate
Coronaviridae	0	77	0	3	0	7	0	0	0	1	0	6	0	0	1	134	Vertebrate
Adenoviridae	96	0	0	0	1	1	37	0	5	2	26	7	0	0	0	17	Vertebrate
Papillomaviridae	0	0	0	0	0	0	0	0	13	2	0	0	5	2	9	3	Vertebrate
Orthomyxoviridae	2	1	0	2	0	0	0	0	0	0	0	0	0	1	0	2	Vertebrate
Circoviridae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Vertebrate

Astroviridae	7	0	0	0	0	0	16	0	0	0	1	0	0	0	0		0	Vertebrate
Reoviridae	0	0	1	0	0	5	0	0	0	0	0	0	0	0	0		0	Vertebrate
Polyomaviridae	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0		0	Vertebrate
Hepadnaviridae	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	Vertebrate
Picornaviridae	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	Vertebrate

Table S3.3 SIMPER results from dissimilarity between species groups, within and between calcretes.

Dissimilarity between groups			
Calcrete	Groups	Species	% Contribution (≥1)
L	*L. lapostaae - L. windarraensis*	Shigella phage SHFML-11	11.10
L	*L. lapostaae - L. windarraensis*	African swine fever virus	6.02
L	*L. lapostaae - L. windarraensis*	Ovine herpesvirus 2	5.11
L	*L. lapostaae - L. windarraensis*	Human herpesvirus 6A	3.30
L	*L. lapostaae - L. windarraensis*	Paramecium bursaria Chlorella virus AR158	2.75
L	*L. lapostaae - L. windarraensis*	European catfish virus	2.66
L	*L. lapostaae - L. windarraensis*	Invertebrate iridescent virus 30	2.61
L	*L. lapostaae - L. windarraensis*	Cotesia congregata bracovirus	2.13
L	*L. lapostaae - L. windarraensis*	Orgyia pseudotsugata multiple nucleopolyhedrovirus	1.97
L	*L. lapostaae - L. windarraensis*	Megavirus chiliensis	1.92
L	*L. lapostaae - L. windarraensis*	Macaca nemestrina herpesvirus 7	1.92
L	*L. lapostaae - L. windarraensis*	Enterobacteria phage ST104	1.86
L	*L. lapostaae - L. windarraensis*	Equid herpesvirus 8	1.59
L	*L. lapostaae - L. windarraensis*	Suid herpesvirus 1	1.46
L	*L. lapostaae - L. windarraensis*	Rabbit fibroma virus	1.44
L	*L. lapostaae - L. windarraensis*	Macacine herpesvirus 4	1.10
L	*L. lapostaae - L. windarraensis*	Paramecium bursaria Chlorella virus FR483	1.02
L	*L. palmulaoides - L. lapostaae*	Shigella phage SHFML-11	7.06
L	*L. palmulaoides - L. lapostaae*	African swine fever virus	3.92
L	*L. palmulaoides - L. lapostaae*	Ovine herpesvirus 2	3.25
L	*L. palmulaoides - L. lapostaae*	Cotesia congregata bracovirus	2.52
L	*L. palmulaoides - L. lapostaae*	European catfish virus	1.82
L	*L. palmulaoides - L. lapostaae*	Human herpesvirus 6A	1.37
L	*L. palmulaoides - L. lapostaae*	Mimivirus terra2	1.34
L	*L. palmulaoides - L. lapostaae*	Invertebrate iridescent virus 30	1.25
L	*L. palmulaoides - L. lapostaae*	Enterobacteria phage ST104	1.10
L	*L. palmulaoides - L. lapostaae*	Suid herpesvirus 1	1.08
L	*L. palmulaoides - L. lapostaae*	Apis mellifera filamentous virus	1.04
L	*L. palmulaoides - L. windarraensis*	Cotesia congregata bracovirus	2.53
L	*L. palmulaoides - L. windarraensis*	Orgyia pseudotsugata multiple nucleopolyhedrovirus	2.17

L	*L. palmulaoides - L. windarraensis*	Paramecium bursaria Chlorella virus AR158	1.91
L	*L. palmulaoides - L. windarraensis*	Mimivirus terra2	1.86
L	*L. palmulaoides - L. windarraensis*	Macaca nemestrina herpesvirus 7	1.64
L	*L. palmulaoides - L. windarraensis*	Megavirus chiliensis	1.61
L	*L. palmulaoides - L. windarraensis*	Human herpesvirus 6A	1.20
L	*L. palmulaoides - L. windarraensis*	Penguinpox virus	1.07
SM	*P. macrosturtensis - P. mesosturtensis*	Human herpesvirus 6A	8.46
SM	*P. macrosturtensis - P. mesosturtensis*	Orgyia pseudotsugata multiple nucleopolyhedrovirus	3.60
SM	*P. macrosturtensis - P. mesosturtensis*	Glypta fumiferanae ichnovirus	1.21
SM	*P. macrosturtensis - P. mesosturtensis*	Papiine herpesvirus 2	1.19
SM	*P. macrosturtensis - P. mesosturtensis*	Shigella phage SHFML-11	1.05
SM	*P. macrosturtensis - P. microsturtensis*	Human herpesvirus 6A	9.65
SM	*P. macrosturtensis - P. microsturtensis*	Orgyia pseudotsugata multiple nucleopolyhedrovirus	3.77
SM	*P. macrosturtensis - P. microsturtensis*	Glypta fumiferanae ichnovirus	1.27
SM	*P. macrosturtensis - P. microsturtensis*	Papiine herpesvirus 2	1.12
SM	*P. macrosturtensis - P. microsturtensis*	Cotesia congregata bracovirus	1.03
SM	*P. microsturtensis - P. mesosturtensis*	Human herpesvirus 6A	8.99
SM	*P. microsturtensis - P. mesosturtensis*	*Pandoravirus inopinatum*	2.18
SM	*P. microsturtensis - P. mesosturtensis*	Invertebrate iridescent virus 30	2.09
SM	*P. microsturtensis - P. mesosturtensis*	Cotesia congregata bracovirus	1.95
SM	*P. microsturtensis - P. mesosturtensis*	Orgyia pseudotsugata multiple nucleopolyhedrovirus	1.77
SM	*P. microsturtensis - P. mesosturtensis*	Corynebacterium phage P1201	1.41
SM	*P. microsturtensis - P. mesosturtensis*	Penguinpox virus	1.38
SM	*P. microsturtensis - P. mesosturtensis*	Aureococcus anophagefferens virus	1.28
SM	*P. microsturtensis - P. mesosturtensis*	Shigella phage SHFML-11	1.23
SM	*P. microsturtensis - P. mesosturtensis*	Mimivirus terra2	1.19
SM	*P. microsturtensis - P. mesosturtensis*	Megavirus chiliensis	1.05
SM - L	*L. lapostaae - P. macrosturtensis*	Human herpesvirus 6A	6.48
SM - L	*L. lapostaae - P. macrosturtensis*	Shigella phage SHFML-11	3.78
SM - L	*L. lapostaae - P. macrosturtensis*	Orgyia pseudotsugata multiple nucleopolyhedrovirus	2.21
SM - L	*L. lapostaae - P. macrosturtensis*	African swine fever virus	2.19
SM - L	*L. lapostaae - P. macrosturtensis*	Ovine herpesvirus 2	1.95

SM - L	*L. lapostaae - P. macrosturtensis*	Cotesia congregata bracovirus	1.88
SM - L	*L. lapostaae - P. macrosturtensis*	Papiine herpesvirus 2	1.31
SM - L	*L. lapostaae - P. macrosturtensis*	European catfish virus	1.06
SM - L	*L. lapostaae - P. mesosturtensis*	Human herpesvirus 6A	19.52
SM - L	*L. lapostaae - P. mesosturtensis*	Orgyia pseudotsugata multiple nucleopolyhedrovirus	7.45
SM - L	*L. lapostaae - P. mesosturtensis*	Shigella phage SHFML-11	6.78
SM - L	*L. lapostaae - P. mesosturtensis*	African swine fever virus	4.15
SM - L	*L. lapostaae - P. mesosturtensis*	Ovine herpesvirus 2	3.78
SM - L	*L. lapostaae - P. mesosturtensis*	Cotesia congregata bracovirus	3.00
SM - L	*L. lapostaae - P. mesosturtensis*	Glypta fumiferanae ichnovirus	2.77
SM - L	*L. lapostaae - P. mesosturtensis*	European catfish virus	2.20
SM - L	*L. lapostaae - P. mesosturtensis*	Invertebrate iridescent virus 30	1.40
SM - L	*L. lapostaae - P. mesosturtensis*	Enterobacteria phage ST104	1.38
SM - L	*L. lapostaae - P. mesosturtensis*	Megavirus chiliensis	1.26
SM - L	*L. lapostaae - P. mesosturtensis*	Equid herpesvirus 8	1.12
SM - L	*L. lapostaae - P. mesosturtensis*	Oryctes rhinoceros virus	1.01
SM - L	*L. lapostaae - P. microsturtensis*	Human herpesvirus 6A	20.29
SM - L	*L. lapostaae - P. microsturtensis*	Orgyia pseudotsugata multiple nucleopolyhedrovirus	6.96
SM - L	*L. lapostaae - P. microsturtensis*	Shigella phage SHFML-11	6.09
SM - L	*L. lapostaae - P. microsturtensis*	Ovine herpesvirus 2	3.09
SM - L	*L. lapostaae - P. microsturtensis*	African swine fever virus	3.09
SM - L	*L. lapostaae - P. microsturtensis*	Glypta fumiferanae ichnovirus	2.57
SM - L	*L. lapostaae - P. microsturtensis*	Cotesia congregata bracovirus	1.56
SM - L	*L. lapostaae - P. microsturtensis*	European catfish virus	1.38
SM - L	*L. lapostaae - P. microsturtensis*	*Pandoravirus salinus*	1.18
SM - L	*L. lapostaae - P. microsturtensis*	*Pandoravirus inopinatum*	1.13
SM - L	*L. lapostaae - P. microsturtensis*	Enterobacteria phage ST104	1.08
SM - L	*L. lapostaae - P. microsturtensis*	Mimivirus terra2	1.02
SM - L	*L. palmulaoides - P. macrosturtensis*	Human herpesvirus 6A	6.71
SM - L	*L. palmulaoides - P. macrosturtensis*	Orgyia pseudotsugata multiple nucleopolyhedrovirus	2.23
SM - L	*L. palmulaoides - P. macrosturtensis*	Cotesia congregata bracovirus	1.83
SM - L	*L. palmulaoides - P. macrosturtensis*	Papiine herpesvirus 2	1.25
SM - L	*L. palmulaoides - P. mesosturtensis*	Human herpesvirus 6A	16.26

SM - L	*L. palmulaoides - P. mesosturtensis*	Orgyia pseudotsugata multiple nucleopolyhedrovirus	5.50
SM - L	*L. palmulaoides - P. mesosturtensis*	Glypta fumiferanae ichnovirus	2.24
SM - L	*L. palmulaoides - P. mesosturtensis*	Cotesia congregata bracovirus	2.08
SM - L	*L. palmulaoides - P. mesosturtensis*	Oryctes rhinoceros virus	1.12
SM - L	*L. palmulaoides - P. microsturtensis*	Human herpesvirus 6A	17.93
SM - L	*L. palmulaoides - P. microsturtensis*	Orgyia pseudotsugata multiple nucleopolyhedrovirus	5.58
SM - L	*L. palmulaoides - P. microsturtensis*	Glypta fumiferanae ichnovirus	2.22
SM - L	*L. palmulaoides - P. microsturtensis*	Cotesia congregata bracovirus	1.52
SM - L	*L. windarraensis - P. mesosturtensis*	Human herpesvirus 6A	24.66
SM - L	*L. windarraensis - P. mesosturtensis*	Orgyia pseudotsugata multiple nucleopolyhedrovirus	10.00
SM - L	*L. windarraensis - P. mesosturtensis*	Glypta fumiferanae ichnovirus	3.34
SM - L	*L. windarraensis - P. mesosturtensis*	Paramecium bursaria Chlorella virus AR158	2.26
SM - L	*L. windarraensis - P. mesosturtensis*	Bovine herpesvirus 1	1.74
SM - L	*L. windarraensis - P. mesosturtensis*	Oryctes rhinoceros virus	1.72
SM - L	*L. windarraensis - P. mesosturtensis*	Cotesia congregata bracovirus	1.60
SM - L	*L. windarraensis - P. mesosturtensis*	Shigella phage SHFML-11	1.60
SM - L	*L. windarraensis - P. mesosturtensis*	Macaca nemestrina herpesvirus 7	1.51
SM - L	*L. windarraensis - P. mesosturtensis*	*Pandoravirus salinus*	1.43
SM - L	*L. windarraensis - P. mesosturtensis*	Equid herpesvirus 2	1.24
SM - L	*L. windarraensis - P. mesosturtensis*	Equid herpesvirus 1	1.02
SM - L	*L. windarraensis - P. microsturtensis*	Human herpesvirus 6A	23.72
SM - L	*L. windarraensis - P. microsturtensis*	Orgyia pseudotsugata multiple nucleopolyhedrovirus	8.70
SM - L	*L. windarraensis - P. microsturtensis*	Glypta fumiferanae ichnovirus	2.92
SM - L	*L. windarraensis - P. microsturtensis*	*Pandoravirus salinus*	1.55
SM - L	*L. windarraensis - P. microsturtensis*	Paramecium bursaria Chlorella virus AR158	1.53
SM - L	*L. windarraensis - P. microsturtensis*	Invertebrate iridescent virus 30	1.52
SM - L	*L. windarraensis - P. microsturtensis*	Corynebacterium phage P1201	1.27
SM - L	*L. windarraensis - P. microsturtensis*	*Pandoravirus inopinatum*	1.23
SM - L	*L. windarraensis - P. microsturtensis*	Oryctes rhinoceros virus	1.17
SM - L	*L. windarraensis - P. microsturtensis*	Mimivirus terra2	1.14
SM - L	*L. windarraensis - P. microsturtensis*	Equid herpesvirus 2	1.10
SM - L	*L. windarraensis - P. microsturtensis*	Gryllus bimaculatus nudivirus	1.03

SM - L	*L. windarraensis - P. microsturtensis*	Macaca nemestrina herpesvirus 7	1.00
SM - L	*L. windarraensis - P.macrosturtensis*	Human herpesvirus 6A	7.60
SM - L	*L. windarraensis - P.macrosturtensis*	Orgyia pseudotsugata multiple nucleopolyhedrovirus	2.61
SM - L	*L. windarraensis - P.macrosturtensis*	Papiine herpesvirus 2	1.28
SM - L	*L. windarraensis - P.macrosturtensis*	Cotesia congregata bracovirus	1.19
SM - L	*L. windarraensis - P.macrosturtensis*	Mimivirus terra2	1.10
SM - L	*L. windarraensis - P.macrosturtensis*	Paramecium bursaria Chlorella virus AR158	1.03

Table S3.4 SIMPER results from similarity within species for *P. macrosturtensis*, *P. microsturtensis* and *L. palmulaoides*.

Similarity within groups		
Group	Species	% Contribution (≥1)
L. palmulaoides	Cotesia congregata bracovirus	6.20
L. palmulaoides	Orgyia pseudotsugata multiple nucleopolyhedrovirus	3.77
L. palmulaoides	Mimivirus terra2	3.17
L. palmulaoides	Megavirus chiliensis	2.86
L. palmulaoides	Melanoplus sanguinipes entomopoxvirus	2.43
L. palmulaoides	Neodiprion abietis NPV	1.81
L. palmulaoides	Enterobacteria phage phiX174 sensu lato	1.54
L. palmulaoides	Penguinpox virus	1.53
L. palmulaoides	Cafeteria roenbergensis virus BV-PW1	1.34
L. palmulaoides	Chrysochromulina ericina virus	1.34
L. palmulaoides	*Pandoravirus inopinatum*	1.34
L. palmulaoides	Human herpesvirus 6A	1.17
L. palmulaoides	Tokyovirus A1	1.17
L. palmulaoides	Emiliania huxleyi virus 86	1.11
L. palmulaoides	Lymphocystis disease virus - isolate China	1.11
L. palmulaoides	Choristoneura biennis entomopoxvirus 'L'	1.01
L. palmulaoides	Invertebrate iridescent virus 30	1.00
P. macrosturtensis	Human herpesvirus 6A	5.57
P. macrosturtensis	Mimivirus terra2	2.45
P. macrosturtensis	Enterobacteria phage phiX174 sensu lato	1.84
P. macrosturtensis	Orgyia pseudotsugata multiple nucleopolyhedrovirus	1.83
P. macrosturtensis	Megavirus chiliensis	1.55
P. macrosturtensis	Papiine herpesvirus 2	1.39
P. macrosturtensis	Cafeteria roenbergensis virus BV-PW1	1.32
P. macrosturtensis	*Pandoravirus salinus*	1.22
P. macrosturtensis	Aureococcus anophagefferens virus	1.19
P. macrosturtensis	Chrysochromulina ericina virus	1.16
P. macrosturtensis	Suid herpesvirus 1	1.09
P. macrosturtensis	*Pandoravirus dulcis*	1.00
P. microsturtensis	Human herpesvirus 6A	25.35
P. microsturtensis	Orgyia pseudotsugata multiple nucleopolyhedrovirus	10.6

P. microsturtensis	Glypta fumiferanae ichnovirus	3.57
P. microsturtensis	Cotesia congregata bracovirus	3.20
P. microsturtensis	Melanoplus sanguinipes entomopoxvirus	2.25
P. microsturtensis	*Pandoravirus salinus*	2.02
P. microsturtensis	Invertebrate iridescent virus 30	1.72
P. microsturtensis	Corynebacterium phage P1201	1.62
P. microsturtensis	Equid herpesvirus 2	1.41
P. microsturtensis	Oryctes rhinoceros virus	1.36
P. microsturtensis	Mimivirus terra2	1.01

Table 3.5 Viral species functional gene information from SEED subsystems level 1 and level 2

Subsystem Level 1	Subsystem Level 2	25542	25543	25544	25545	25546	27820	27821	27822	27823	27824	27825	28084	28085	28086	28088	28089
Amino Acids and Derivatives	Alanine, serine, and glycine	1473.81	96.79	575.14	20.77	69.82	46.17	53.00	46.11	25.44	43.86	14.87	1735.40	597.31	45.01	22.02	1787.07
Amino Acids and Derivatives	Arginine; urea cycle, polyamines	3426.73	53.34	1182.67	9.00	26.07	119.77	46.71	13.08	7.82	11.70	19.13	3657.72	1585.86	42.25	9.75	3852.27
Amino Acids and Derivatives	Aromatic amino acids and derivatives	2505.32	65.83	805.47	15.00	37.45	65.20	47.36	35.50	23.50	37.17	31.50	2859.33	759.66	47.75	21.00	2781.13
Amino Acids and Derivatives	Branched-chain amino acids	4323.86	120.60	1689.74	14.12	86.55	99.15	89.43	54.53	21.86	52.30	40.18	3065.15	1306.92	59.24	30.92	2907.40
Amino Acids and Derivatives	Glutamine, glutamate, aspartate, asparagine; ammonia assimilation	1651.44	71.15	664.00	9.40	36.68	59.85	36.51	16.84	5.58	20.20	14.06	1686.25	542.94	22.31	9.79	1643.56
Amino Acids and Derivatives	Histidine Metabolism	1390.18	28.58	421.60	6.50	16.00	34.01	31.16	14.47	5.10	11.88	8.50	1509.80	382.64	15.78	4.50	1548.01
Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	7120.18	450.98	2930.33	58.23	167.84	238.49	137.68	743.35	229.02	1236.66	430.48	6550.96	1978.02	112.32	62.53	6409.64
Amino Acids and Derivatives	Proline and 4-hydroxyproline	1260.68	30.12	429.52	0.75	8.97	27.60	21.66	3.48	2.10	5.70	5.63	1259.93	483.85	12.39	6.00	1248.06
Arabinose Sensor and transport module	Arabinose	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.50

Carbohydrates	Aceton metabolism	2.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Carbohydrates	Aminosugars	826.88	142.87	258.03	90.83	30.67	103.83	34.98	91.50	36.50	111.67	28.14	661.91	548.43	64.75	25.00	650.49
Carbohydrates	Central carbohydrate metabolism	9034.40	639.53	3384.31	79.92	238.33	340.23	264.62	264.27	103.32	249.90	176.94	7438.17	2609.89	303.34	156.32	7505.61
Carbohydrates	CO2 fixation	851.52	51.69	377.76	8.05	25.57	82.94	23.28	12.38	11.03	16.07	14.33	641.19	224.30	26.93	10.86	655.86
Carbohydrates	Di- and oligosaccharides	2043.22	36.30	820.28	7.87	35.90	99.43	30.28	18.08	13.62	26.40	15.83	1750.54	449.72	42.25	16.53	1877.49
Carbohydrates	Fermentation	1570.25	32.62	615.17	4.84	18.96	65.66	28.78	15.62	6.29	10.33	17.17	1959.85	709.65	30.05	11.37	1913.89
Carbohydrates	Monosaccharides	4819.34	76.90	1379.69	20.80	74.10	126.55	63.50	37.27	25.80	54.22	32.05	3808.18	672.75	83.88	28.58	3887.95
Carbohydrates	Nucleotide sugars	3.60	2.67	4.50	0.60	2.90	0.00	0.40	0.90	0.90	0.00	0.00	0.00	0.00	1.40	2.80	0.00
Carbohydrates	One-carbon Metabolism	1768.00	62.18	712.66	4.48	32.90	47.96	39.45	10.03	7.05	18.65	10.50	1334.21	636.48	23.25	10.59	1221.07
Carbohydrates	Organic acids	1448.57	17.86	363.97	3.00	16.88	38.83	16.67	11.14	3.12	3.03	3.37	1899.62	868.69	7.00	4.17	1834.47
Carbohydrates	Polysaccharides	744.72	22.10	320.17	2.40	8.57	13.24	6.73	7.40	3.57	7.43	8.00	123.34	178.04	6.70	6.13	119.98
Carbohydrates	Sugar alcohols	2067.46	22.62	697.46	6.33	18.83	88.17	23.17	14.17	5.25	17.23	8.50	1698.51	252.31	20.67	10.67	1645.51
Cell Division and Cell Cycle	Bacterial checkpoint control	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Cell Wall and Capsule	Capsular and extracellular polysacchrides	2456.21	52.18	963.40	6.67	28.32	90.40	45.15	9.83	12.65	11.77	11.12	1902.95	944.82	23.45	12.17	1835.83
Cell Wall and Capsule	Cell wall of Mycobacteria	39.83	0.00	79.00	0.00	0.00	0.00	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Cell Wall and Capsule	Gram-Negative cell wall components	2957.69	64.95	789.67	21.37	34.50	141.02	81.08	70.34	15.90	49.87	45.87	4766.27	1863.33	58.23	29.58	4699.37
Cell Wall and Capsule	Gram-Positive cell wall components	188.68	2.00	52.96	0.00	6.00	6.17	2.53	0.20	0.00	2.67	0.00	141.73	194.41	0.00	0.00	171.87
Central metabolism	TCA	2.33	0.00	5.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Clustering-based subsystems	A bicyclomycin resistance protein, a helicase, and a pseudouridine synthase	2.80	0.00	1.00	0.00	0.00	2.00	0.00	0.00	0.00	0.00	0.00	61.75	12.50	0.00	0.00	51.45
Clustering-based subsystems	alpha-proteobacterial cluster of hypotheticals	111.08	3.00	50.99	0.00	0.00	0.00	2.00	0.50	0.00	0.50	0.00	8.33	3.83	1.00	0.50	4.83
Clustering-based subsystems	Biosynthesis of galactoglycans and related lipopolysacharides	515.72	0.50	168.53	0.00	2.33	0.00	4.00	3.00	0.00	6.00	1.00	105.17	9.69	0.33	1.00	104.97
Clustering-based subsystems	Carotenoid biosynthesis	20.28	0.00	11.37	0.00	0.00	1.50	0.00	0.00	0.00	0.10	0.00	33.17	7.64	0.00	0.00	38.36
Clustering-based subsystems	Catabolism of an unknown compound	10.50	1.00	22.00	0.00	1.00	0.00	0.00	0.50	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
Clustering-based subsystems	Cell Division	1120.51	14.76	286.28	2.83	10.83	26.01	31.80	8.83	2.33	5.12	7.84	1285.42	449.99	40.02	8.03	1342.38
Clustering-based subsystems	Chemotaxis, response regulators	36.82	0.00	13.00	0.00	0.00	7.83	0.00	0.00	0.00	1.50	0.00	12.54	61.50	0.00	0.00	19.58

Clustering-based subsystems	Choline bitartrate degradation, putative	1.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	51.50	3.00	0.00	0.00	37.31
Clustering-based subsystems	Chromosome Replication	17.00	0.00	12.00	0.00	1.00	3.00	0.00	0.00	1.00	0.00	0.00	68.00	82.00	0.00	0.00	44.00
Clustering-based subsystems	clustering of 2 heat shock proteins, phosphoenolpyruvate carboxykinase and a putative hydrolase	26.35	1.33	5.99	0.00	0.00	3.73	1.25	0.00	0.00	1.00	0.00	206.05	45.58	0.33	0.00	198.93
Clustering-based subsystems	contains Thr-tRNA-syn, pyridoxine biosyn, lipid A biosyn, 3 hypos	22.76	0.00	55.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00
Clustering-based subsystems	CRISPRs and associated hypotheticals	13.00	0.00	4.00	0.00	0.00	0.50	0.00	0.00	0.00	1.00	0.00	6.00	1.00	0.00	0.00	11.00
Clustering-based subsystems	Cytochrome biogenesis	533.82	4.75	223.97	1.00	4.25	7.25	9.58	3.50	2.00	3.00	1.29	227.81	138.70	3.00	1.00	220.42
Clustering-based subsystems	DNA metabolism	6.44	0.00	6.28	0.00	0.00	1.00	0.00	0.50	0.33	0.00	0.00	60.25	7.33	0.50	0.00	51.09
Clustering-based subsystems	DNA polymerase III epsilon cluster	241.37	0.53	96.28	0.00	3.50	12.50	3.32	1.83	0.00	1.83	0.33	357.99	153.02	2.94	3.70	376.81
Clustering-based subsystems	Degradation of Polyphenols (?)	74.49	0.00	0.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.08	0.00	0.00	0.00	0.55

Clustering-based subsystems	D-tyrosyl-tR0(Tyr) deacylase (EC 3.1.-.-) cluster	41.87	0.00	7.14	0.00	0.00	0.75	0.00	0.00	0.00	0.00	0.00	61.52	12.67	0.00	0.00	54.99
Clustering-based subsystems	Fage-related, replication	236.21	0.00	69.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	135.50	0.00	0.00	0.00	127.00
Clustering-based subsystems	Fatty acid metabolic cluster	410.21	6.31	167.77	0.00	5.75	4.76	2.69	1.09	1.33	2.15	0.93	358.93	85.35	7.26	1.19	322.78
Clustering-based subsystems	Flagella protein?	38.53	1.00	17.80	0.00	0.33	0.33	0.00	0.00	0.00	0.00	0.00	15.67	0.67	0.33	0.33	12.35
Clustering-based subsystems	heat shock, cell division, proteases, and a methyltransferase	76.47	0.00	14.11	0.00	1.00	0.33	1.68	0.00	0.00	0.00	1.00	59.60	8.63	0.20	0.00	60.34
Clustering-based subsystems	Hypothetical associated with RecF	4.12	0.00	5.00	0.00	0.00	0.67	0.00	0.00	0.00	0.00	0.00	5.53	20.79	0.00	0.00	8.91
Clustering-based subsystems	Hypothetical in Lysine biosynthetic cluster	57.27	0.50	13.99	0.00	0.00	7.00	0.14	0.00	0.00	0.45	0.00	86.45	99.15	0.00	0.00	107.12
Clustering-based subsystems	Hypothetical lipase related to Phosphatidate metabolism	23.70	0.00	11.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	90.77	31.71	0.00	0.33	95.11
Clustering-based subsystems	Hypothetical protein possible functionally linked with Alanyl-tRNA synthetase	31.13	0.75	16.42	0.00	0.00	1.63	1.25	0.00	0.00	0.00	0.00	75.21	10.98	0.50	0.00	66.31

Clustering-based subsystems	Hypothetical Related to Dihydroorate Dehydrogenase	23.17	0.00	4.83	0.00	0.50	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Clustering-based subsystems	Isoprenoid/cell wall biosynthesis: PREDICTED UNDECAPRENYL DIPHOSPHATE PHOSPHATASE	143.35	2.20	52.12	2.00	0.90	1.95	5.68	0.28	0.58	0.60	0.80	172.26	169.95	2.20	1.00	185.12
Clustering-based subsystems	Lysine Biosynthesis	47.28	0.33	12.17	0.00	0.00	1.00	3.67	0.00	0.00	0.00	0.00	39.41	22.34	0.33	0.25	44.86
Clustering-based subsystems	Methylamine utilization	111.00	1.00	41.50	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	45.00	9.00	0.00	0.00	43.00
Clustering-based subsystems	Molybdopterin oxidoreductase	158.00	0.00	56.00	0.00	4.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Clustering-based subsystems	Nucleotidyl-phosphate metabolic cluster	1.00	0.00	0.50	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	6.33	15.83	0.00	0.00	2.00
Clustering-based subsystems	Pigment biosynthesis	5.50	2.00	5.50	0.00	0.50	0.00	0.00	0.50	0.00	0.50	0.00	0.00	0.00	1.00	0.00	0.00
Clustering-based subsystems	Probably GTP or GMP sig0ling related	235.27	4.79	65.74	0.00	0.00	9.28	5.25	0.50	0.83	0.70	1.50	355.94	132.86	6.17	1.50	326.27
Clustering-based subsystems	Probably organic hydroperoxide	90.67	0.00	26.50	0.33	0.50	1.33	0.83	0.00	0.00	0.83	0.00	106.17	25.00	0.00	0.50	93.00

	resistance related hypothetical protein																
Clustering-based subsystems	Probably Ybbk-related hypothetical membrane proteins	82.00	2.00	33.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	1.00	70.00	118.00	1.00	0.00	70.50
Clustering-based subsystems	Protein export?	87.43	1.00	68.33	0.00	0.00	3.00	2.00	0.00	0.00	1.00	1.00	23.33	60.00	1.50	0.00	27.42
Clustering-based subsystems	proteosome related	128.70	11.00	98.17	1.00	6.00	1.50	4.50	5.00	6.00	3.00	3.00	61.08	20.55	7.00	1.50	54.04
Clustering-based subsystems	Putative asociate of RNA polymerase sigma-54 factor rpoN	558.21	6.00	139.33	0.00	4.50	2.50	3.00	5.00	0.00	1.00	2.00	201.48	57.00	6.50	2.00	171.64
Clustering-based subsystems	Putative GGDEF domain protein related to agglutinin secretion	41.67	0.00	17.00	0.00	0.00	28.00	3.00	1.00	0.00	2.00	2.00	38.00	264.17	0.00	0.00	42.00
Clustering-based subsystems	Putative Isoquinoline 1-oxidoreductase subunit	187.48	1.33	52.57	0.33	0.50	2.40	0.00	1.33	0.33	0.00	0.33	56.21	32.33	0.50	0.00	53.95
Clustering-based subsystems	Putrescine/GABA utilization cluster-temporal,to add to SSs	165.66	1.00	43.70	0.00	0.33	11.67	3.00	1.00	0.00	1.50	0.50	132.82	44.90	0.25	0.00	153.89
Clustering-based subsystems	Pyruvate kinase associated cluster	21.34	0.00	5.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00

Clustering-based subsystems	Recombination related cluster	42.00	0.00	15.50	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	3.50	10.70	0.00	0.00	4.00
Clustering-based subsystems	recX and regulatory cluster	1.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	7.17	1.00	0.00	0.50
Clustering-based subsystems	Related to Menaquinone-cytochrome C reductase	0.00	2.00	5.00	1.00	2.00	1.00	0.00	17.00	9.00	17.00	6.00	0.00	5.33	0.00	0.00	2.00
Clustering-based subsystems	Ribosomal Protein L28P relates to a set of uncharacterized proteins	24.81	12.00	23.63	4.00	7.00	0.00	1.00	16.50	7.00	5.00	1.00	4.65	16.49	2.00	2.50	2.75
Clustering-based subsystems	Ribosome-related cluster	82.31	0.33	21.17	0.00	0.00	5.67	1.33	0.50	0.20	0.50	0.00	182.99	75.92	0.50	0.50	165.49
Clustering-based subsystems	Shiga toxin cluster	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.00	0.00	0.00	0.00	14.00
Clustering-based subsystems	Sulfatases and sulfatase modifying factor 1 (and a hypothetical)	262.52	3.00	110.21	0.00	3.50	1.00	1.00	0.00	0.00	1.00	7.00	82.79	5.33	3.25	2.00	55.09
Clustering-based subsystems	Tartronate-semialdehyde related area (links to pyridoxine and aldorate metabolism)	14.23	0.00	3.00	0.00	0.00	2.00	0.00	0.00	0.00	0.00	0.00	27.59	0.25	0.00	0.00	26.11

Clustering-based subsystems	Three hypotheticals linked to lipoprotein biosynthesis	57.33	1.50	24.33	0.00	0.50	0.00	5.50	0.00	0.50	0.00	0.50	3.00	0.00	1.00	0.00	2.00
Clustering-based subsystems	TldD cluster	326.50	4.83	67.08	1.00	2.17	12.23	4.83	9.90	3.00	21.00	0.50	398.55	98.16	3.75	2.50	396.08
Clustering-based subsystems	tRNA sulfuration	66.84	12.24	42.14	0.00	5.58	3.67	3.25	2.50	1.00	3.50	1.33	81.18	27.40	2.83	1.33	86.70
Clustering-based subsystems	Translation	389.20	3.75	106.00	1.00	1.00	21.07	4.88	0.00	1.00	4.50	0.00	446.08	283.95	5.50	0.50	449.71
Clustering-based subsystems	Tricarboxylate transporter	1049.83	10.00	367.86	1.00	1.50	65.50	1.00	1.00	0.00	1.00	0.50	198.24	82.00	1.00	1.00	168.24
Clustering-based subsystems	Two related proteases	5.50	0.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	11.18	3.00	0.00	6.15
Clustering-based subsystems	Type III secretion system, extended	8.00	0.50	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.50	0.00	0.00	0.00	4.50
Clustering-based subsystems	Urate degradation	115.00	1.00	41.00	0.00	1.00	5.00	0.00	0.00	0.00	0.00	0.00	124.00	10.00	0.00	0.00	130.50
Cofactors, Vitamins, Prosthetic Groups, Pigments	NAD and NADP	979.33	143.58	459.97	19.50	38.73	36.50	40.76	216.50	105.00	282.42	100.33	1101.13	377.72	42.08	14.50	1055.40
Cofactors, Vitamins, Prosthetic Groups, Pigments	Biotin	2884.88	48.01	1906.59	14.79	34.68	49.49	48.91	25.78	7.36	29.53	31.00	1359.65	443.12	59.67	28.58	1358.54
Cofactors, Vitamins, Prosthetic Groups, Pigments	Coenzyme A	747.94	37.08	355.70	24.50	25.42	23.75	20.58	114.67	25.00	30.50	11.00	862.81	240.09	21.67	7.00	848.73

Cofactors, Vitamins, Prosthetic Groups, Pigments	Coenzyme F420	28.67	4.00	55.42	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Cofactors, Vitamins, Prosthetic Groups, Pigments	Coenzyme M	8.00	0.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cofactors, Vitamins, Prosthetic Groups, Pigments	Fe-S clusters	251.87	6.82	91.00	1.70	4.78	4.77	7.29	1.83	3.00	3.35	1.84	150.33	75.88	10.85	6.79	165.39
Cofactors, Vitamins, Prosthetic Groups, Pigments	Folate and pterines	2626.98	174.16	871.06	31.36	112.15	78.97	119.02	85.56	48.52	84.67	57.01	2451.52	1472.99	146.17	63.81	2375.70
Cofactors, Vitamins, Prosthetic Groups, Pigments	Lipoic acid	1416.12	46.94	416.71	8.60	38.62	24.75	39.14	24.95	8.90	20.59	16.68	1500.82	347.79	47.10	17.40	1457.23
Cofactors, Vitamins, Prosthetic Groups, Pigments	Pyridoxine	764.12	28.85	292.68	2.56	10.11	24.14	36.38	20.67	4.50	7.50	12.64	669.17	269.68	19.42	7.40	717.17
Cofactors, Vitamins, Prosthetic Groups, Pigments	Quinone cofactors	1312.38	34.17	558.03	12.17	23.92	51.20	51.02	25.13	5.88	7.50	13.36	1829.83	613.00	24.83	10.00	1753.20
Cofactors, Vitamins, Prosthetic Groups, Pigments	Riboflavin, FMN, FAD	766.45	26.33	254.61	6.50	15.40	22.64	27.37	10.58	4.00	15.50	11.80	910.55	471.41	29.73	14.20	920.37

Cofactors, Vitamins, Prosthetic Groups, Pigments	Tetrapyrroles	1312.81	27.83	457.52	13.00	8.50	16.50	19.83	8.67	2.00	15.20	10.88	533.77	190.22	13.50	6.00	530.29
Cofactors, Vitamins, Prosthetic Groups, Pigments	Thiamin	46.99	1.00	12.57	0.61	0.61	1.07	0.28	0.00	0.00	0.27	0.08	21.02	17.10	0.84	0.00	28.36
DNA Metabolism	CRISPs	175.00	1.00	72.00	2.00	1.00	0.50	1.00	0.00	0.00	1.00	0.00	31.00	1.00	8.00	0.00	34.00
DNA Metabolism	DNA recombination	213.62	5.17	85.39	0.00	0.96	2.09	6.66	4.44	0.00	0.50	2.03	184.47	29.74	5.67	0.50	174.46
DNA Metabolism	DNA repair	5474.47	111.77	1892.57	25.50	68.50	152.94	142.20	769.40	342.67	806.10	203.50	5902.64	1977.57	153.77	25.07	5709.09
DNA Metabolism	DNA replication	2676.39	70.49	767.38	7.00	46.45	67.22	73.33	36.39	30.00	35.33	21.67	2391.43	1233.79	79.14	22.90	2477.01
DNA Metabolism	DNA uptake, competence	227.46	4.30	63.58	0.50	1.00	16.50	4.50	2.00	1.50	2.50	2.25	350.54	189.91	9.00	3.00	329.09
Dormancy and Sporulation	Spore DNA protection	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	17.00	0.00	0.00	0.00
Fatty Acids, Lipids, and Isoprenoids	Fatty acids	4496.04	159.25	2607.34	38.67	86.15	113.34	170.14	90.50	56.00	112.76	78.00	4320.48	2112.74	131.66	75.17	4314.38
Fatty Acids, Lipids, and Isoprenoids	Isoprenoids	1078.73	26.44	490.88	4.37	27.93	32.45	31.68	24.09	2.26	20.28	14.03	1275.20	254.28	42.39	16.25	1211.49
Fatty Acids, Lipids, and Isoprenoids	Phospholipids	1510.68	30.41	538.21	2.33	25.12	45.70	25.96	8.37	3.58	8.45	10.70	1822.91	543.36	40.25	12.58	1848.99
Fatty Acids, Lipids, and Isoprenoids	Triacylglycerols	181.50	0.00	78.83	0.00	1.00	4.00	1.00	1.00	0.00	0.00	0.00	445.37	98.09	0.00	1.00	441.97
Iron acquisition and metabolism	Siderophores	738.67	0.50	233.17	0.00	0.00	14.33	1.00	1.00	1.00	2.33	0.00	1856.48	288.57	2.00	0.00	1768.56

Membrane Transport	ABC transporters	4685.81	47.83	1087.92	24.50	6.50	90.72	23.00	26.00	11.00	37.00	7.00	1736.40	611.74	12.50	4.00	1707.51
Membrane Transport	Protein and nucleoprotein secretion system, Type IV	766.87	8.17	215.42	1.50	4.00	38.66	25.32	4.23	2.50	2.50	2.58	1761.25	846.28	17.33	4.00	1707.15
Membrane Transport	Protein secretion system, Chaperone-Usher pathway (CU)	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	254.33	4.00	0.00	0.00	246.83
Membrane Transport	Protein secretion system, Type II	795.50	6.00	214.50	0.00	0.00	39.00	3.00	4.00	2.00	5.00	0.00	437.83	569.00	4.00	1.00	398.42
Membrane Transport	Protein secretion system, Type III	7.00	0.50	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	29.50	13.00	0.00	0.00	24.50
Membrane Transport	Protein secretion system, Type VI	613.58	1.00	149.32	0.00	0.33	70.17	11.24	0.00	2.33	4.67	1.10	2586.29	721.12	6.07	0.33	2543.20
Membrane Transport	Protein secretion system, Type VII	59.00	0.00	116.00	0.00	0.00	9.60	4.00	2.00	1.00	1.00	0.00	10.00	38.00	1.00	0.00	17.00
Membrane Transport	Protein secretion system, Type VIII (Extracellular nucleation/precipitation pathway, ENP)	22.00	0.00	30.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	114.00	0.00	0.00	0.00
Membrane Transport	Protein translocation across cytoplasmic membrane	149.20	4.00	65.67	0.00	1.00	7.00	1.50	0.00	0.00	0.00	1.50	239.14	94.39	2.50	1.00	261.29

Membrane Transport	Sugar Phosphotransferase Systems, PTS	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.14	0.00	0.00	0.00
Membrane Transport	TRAP transporters	99.00	0.00	13.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	17.00	2.00	0.00	0.00	15.00
Membrane Transport	Uni- Sym- and Antiporters	537.84	7.00	219.50	0.00	1.00	7.00	12.00	13.00	2.00	6.00	0.00	645.40	348.67	3.00	0.00	693.04
Metabolism of Aromatic Compounds	Anaerobic degradation of aromatic compounds	143.76	7.00	34.89	0.00	4.20	1.00	1.17	4.00	0.00	1.00	0.00	102.73	28.01	2.00	3.00	75.65
Metabolism of Aromatic Compounds	Metabolism of central aromatic intermediates	2093.71	13.18	645.07	4.83	7.13	17.83	12.56	6.33	3.43	10.50	4.50	2045.89	229.69	11.09	4.33	1988.58
Metabolism of Aromatic Compounds	Peripheral pathways for catabolism of aromatic compounds	996.96	6.24	315.51	0.00	1.32	13.16	8.84	2.48	0.15	2.61	1.73	2151.53	137.18	10.18	1.69	1951.89
Miscellaneous	Plant-Prokaryote comparative genomics	4371.81	199.32	1644.68	45.52	139.49	93.40	162.91	111.30	68.78	123.99	111.68	3654.18	1777.85	173.06	97.96	3611.81
Motility and Chemotaxis	Flagellar motility in Prokaryota	2398.90	15.74	582.88	4.67	9.25	86.10	19.53	6.28	1.48	18.57	13.17	1564.98	2047.96	20.97	12.37	1645.58
Motility and Chemotaxis	Social motility and nonflagellar swimming in bacteria	629.13	1.00	143.86	1.00	2.00	9.21	8.40	0.00	0.50	3.33	1.83	795.45	284.32	5.00	1.00	788.27

Nucleosides and Nucleotides	Purines	3247.80	238.54	1278.28	45.17	110.83	125.48	119.50	148.33	68.23	289.25	73.83	3654.32	1346.20	84.60	40.33	3580.96
Nucleosides and Nucleotides	Pyrimidines	1917.51	43.33	654.09	5.50	28.65	33.77	49.26	25.39	12.00	22.83	19.52	1760.91	560.35	55.58	19.00	1790.58
Phages, Prophages, Transposable elements	Bacteriophage integration/excision/lysogeny	10.50	0.00	9.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	46.50	1.00	1.00	0.00	49.00
Phages, Prophages, Transposable elements	Bacteriophage structural proteins	65.42	32.00	32.83	35.00	20.00	0.00	0.00	0.00	0.00	0.00	0.00	36.00	43.00	43.00	35.00	35.50
Phages, Prophages, Transposable elements	Phage family-specific subsystems	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Phages, Prophages, Transposable elements	Phage Host Interactions	2.00	0.00	4.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.00	1.00	0.00	0.00	5.00
Phages, Prophages, Transposable elements	Superinfection Exclusion	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.00	0.00	0.00	0.00	8.00
Phages, Prophages, Transposable elements, Plasmids	Pathogenicity islands	153.80	5.00	60.82	0.00	0.00	10.60	2.95	0.33	0.00	1.50	0.50	271.96	66.67	1.70	0.00	291.02
Phages, Prophages, Transposable elements, Plasmids	Phages, Prophages	399.07	15.57	148.98	4.00	8.00	2.00	16.75	11.44	1.00	5.03	12.79	1049.55	146.65	11.50	10.00	1077.73

Phages, Prophages, Transposable elements, Plasmids	Plasmid related functions	2.00	0.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Phages, Prophages, Transposable elements, Plasmids	Transposable elements	873.39	2.50	97.25	0.00	1.00	3.00	5.50	1.00	0.00	0.00	1.00	339.69	49.25	3.00	0.00	315.66
Photosynthesis	Electron transport and photophosphorylation	26.50	25.00	24.50	5.67	17.00	7.67	13.67	10.33	8.33	17.33	22.00	0.00	0.00	9.00	7.00	1.00
Photosynthesis	Light-harvesting complexes	1.50	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Protein Metabolism	Protein biosynthesis	9373.32	515.80	3363.26	77.78	324.29	303.94	472.29	265.97	132.35	226.19	224.01	8516.94	3582.84	560.26	198.44	8622.04
Protein Metabolism	Protein degradation	2723.45	202.67	1095.03	58.17	163.00	121.40	160.96	161.83	90.23	150.83	114.95	2570.12	1645.97	177.50	93.78	2586.69
Protein Metabolism	Protein folding	1203.10	144.13	408.23	25.60	76.75	52.27	67.70	63.03	28.13	59.50	36.42	1585.47	532.59	68.98	49.73	1548.50
Protein Metabolism	Protein processing and modification	780.72	57.00	349.58	93.50	47.26	35.84	35.34	74.39	36.75	27.27	23.50	861.42	402.31	49.78	20.29	758.55
Protein Metabolism	Secretion	88.33	0.00	28.00	2.00	1.00	3.00	2.00	0.00	0.00	0.00	1.00	347.50	7.00	0.00	0.00	326.00
Protein Metabolism	Selenoproteins	422.97	24.00	170.17	4.00	22.50	18.50	21.96	9.00	11.00	11.50	15.75	460.92	238.21	14.17	9.00	412.70
Regulation and Cell signaling	Programmed Cell Death and Toxin-antitoxin Systems	345.50	5.00	72.50	0.00	2.00	20.00	8.00	2.50	1.00	2.50	0.00	606.31	224.59	1.00	1.00	579.17
Regulation and Cell signaling	Proteolytic pathway	37.14	45.83	32.17	14.50	27.00	22.83	44.50	83.53	22.00	43.00	110.00	0.00	0.00	43.50	30.67	0.00

Regulation and Cell signaling	Quorum sensing and biofilm formation	180.91	0.11	22.95	0.50	0.00	30.25	0.00	0.25	0.50	0.70	0.50	342.89	35.66	1.00	0.00	325.07
Regulation and Cell signaling	Regulation of virulence	8.00	0.00	4.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	49.00	15.00	0.00	0.00	39.00
Regulation and Cell signaling	Signal transduction in Eukaryotes	85.61	63.33	60.83	12.00	46.00	10.20	23.50	27.50	13.00	26.00	15.00	119.34	34.12	36.50	35.00	118.68
Respiration	ATP synthases	833.50	101.00	289.00	7.00	34.00	28.00	37.00	20.00	9.00	30.00	24.00	704.00	373.00	46.00	21.00	708.00
Respiration	Electron accepting reactions	2730.31	136.89	1068.62	77.60	31.35	128.63	425.14	511.16	310.64	118.83	38.83	1831.56	2175.09	89.81	105.24	1905.34
Respiration	Electron donating reactions	3295.46	154.94	1324.76	75.67	43.50	88.88	307.38	329.05	191.60	127.92	47.78	2753.08	1668.65	155.26	125.62	2770.11
Respiration	General Stress Response and Statio0ry Phase Response	78.50	1.00	31.07	0.00	0.00	4.00	2.50	2.00	0.00	0.20	0.67	268.28	47.71	0.00	0.00	260.35
Respiration	Mitochondrial electron transport system in plants	85.02	44.71	48.83	29.50	12.77	36.23	198.36	235.50	121.90	35.17	18.67	0.00	0.00	22.99	44.91	0.67
Respiration	Plastidial (cyanobacterial) electron transport system	0.00	0.00	0.00	0.00	1.00	0.00	0.00	4.00	4.00	3.00	2.00	0.00	0.00	0.00	0.00	1.00
Respiration	Sodium Ion-Coupled Energetics	10.83	0.00	6.10	0.00	0.50	0.59	0.00	0.00	0.00	0.00	0.00	1.46	38.33	0.00	0.00	2.08
RNA Metabolism	RNA processing and modification	5738.87	349.44	1873.81	73.24	203.22	214.78	216.68	215.75	103.13	168.62	195.65	6771.53	2551.52	266.38	148.47	6762.13

RNA Metabolism	Transcription	2239.06	229.31	898.95	49.00	152.92	119.90	140.90	154.50	89.27	255.57	108.81	1779.99	697.96	210.74	114.80	1786.36
Secondary Metabolism	Bacterial cytostatics, differentiation factors and antibiotics	33.50	0.00	7.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	13.50	0.00	0.00	0.00	11.00
Secondary Metabolism	Biologically active compounds in metazoan cell defence and differentiation	109.97	0.50	47.07	0.00	0.50	0.33	0.33	0.33	0.00	0.33	0.00	27.18	1.00	0.25	0.00	23.03
Secondary Metabolism	Biosynthesis of phenylpropanoids	45.58	1.00	23.33	0.00	1.00	0.33	0.50	1.00	0.00	2.25	1.25	37.77	0.00	1.00	0.50	28.89
Secondary Metabolism	Lipid-derived mediators	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Secondary Metabolism	Plant Alkaloids	10.50	4.50	3.00	0.50	2.57	1.00	3.00	3.00	2.00	0.50	2.50	0.00	0.00	5.67	1.00	0.00
Secondary Metabolism	Plant Hormones	40.21	0.00	20.23	0.30	0.00	2.31	0.67	0.00	1.00	0.00	0.00	33.33	20.89	0.00	0.00	27.35
Stress Response	Acid stress	4.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Stress Response	Cold shock	108.50	2.00	42.00	0.00	3.00	3.00	1.00	0.00	0.00	2.00	2.00	79.00	28.00	4.00	1.00	72.00
Stress Response	Dessication stress	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00
Stress Response	Detoxification	809.86	15.33	300.19	1.00	13.50	40.86	15.50	5.50	1.00	8.05	9.33	1009.95	269.26	15.17	5.00	1037.62
Stress Response	Heat shock	1369.29	86.85	547.63	24.15	40.67	48.30	57.76	48.55	31.02	35.24	40.02	1189.43	450.60	58.42	38.75	1174.20
Stress Response	Osmotic stress	1875.94	48.88	710.38	16.30	33.28	97.42	34.17	31.53	14.00	31.73	26.62	2174.47	434.10	42.68	29.75	2140.78

Stress Response	Oxidative stress	3039.09	70.87	1083.52	9.10	45.22	107.83	52.68	29.24	13.87	34.53	57.77	3404.14	1343.19	48.07	18.80	3338.08
Stress Response	Periplasmic Stress	304.40	1.00	85.00	0.00	0.00	6.67	5.25	0.50	0.60	0.50	2.62	242.61	123.97	3.50	3.00	241.25
Sulfur Metabolism	Inorganic sulfur assimilation	506.66	4.50	227.54	3.17	2.33	6.80	11.37	4.77	1.50	3.33	0.33	783.79	305.57	3.83	0.50	740.78
Sulfur Metabolism	Organic sulfur assimilation	1551.90	14.53	398.34	4.20	6.20	26.56	23.84	4.70	3.44	5.72	2.72	2247.65	288.58	9.35	5.70	2036.93
Virulence	Fimbriae of the Chaperone/Usher Assembly Pathway	34.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	180.00	4.00	0.00	0.00	179.00
Virulence	Resistance to antibiotics and toxic compounds	9735.80	111.03	2708.15	18.38	37.73	167.60	82.13	81.50	24.17	159.33	33.42	8378.23	3361.25	77.77	26.43	8236.88
Virulence	Type III, Type IV, Type VI, ESAT secretion systems	1347.85	12.00	206.50	2.00	1.00	0.00	15.00	2.00	0.00	1.50	1.50	95.50	42.00	9.00	4.00	103.82
Virulence, Disease and Defense	Adhesion	146.57	2.89	22.80	2.50	4.50	9.00	3.00	2.33	1.50	6.50	5.50	280.70	18.31	4.00	2.00	286.37
Virulence, Disease and Defense	Bacteriocins, ribosomally synthesized antibacterial peptides	316.47	1.50	66.47	0.83	2.17	11.67	6.45	2.00	1.00	0.00	1.50	867.44	297.06	1.00	0.25	802.42
Virulence, Disease and Defense	Detection	13.22	2.33	6.94	0.00	2.00	2.00	0.50	2.00	2.00	2.33	1.67	20.22	33.99	3.75	3.00	17.55
Virulence, Disease and Defense	Invasion and intracellular resistance	171.00	2.00	219.33	2.00	3.00	0.00	2.00	1.00	0.00	1.00	0.00	69.00	46.56	1.00	1.00	82.00

Virulence, Disease and Defense	Toxins and superantigens	3.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	2.00	1.00	0.00	0.00	8.00
unclassifed	unclassifed	51906.09	1209.02	16786.84	765.11	628.78	1329.13	982.38	2513.31	1301.63	3425.97	644.68	44268.91	20045.24	836.15	345.74	44364.91

Statement of Authorship

Title of Paper	Characterising the microbiome of Australian subterranean diving beetles
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Written in the style of the journal FEMS

Principal Author

Name of Principal Author (Candidate)	Josephine Hyde
Contribution to the Paper	Collection of samples, molecular work, analysing the data, interpretation of the data, writing the manuscript
Overall percentage (%)	70 %
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 15/6/18

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Lisa Dann
Contribution to the Paper	Discussion of concepts, editing of manuscript
Signature	Date 20/06/2018

Name of Co-Author	William Humphreys
Contribution to the Paper	Editing of the manuscript, collection of samples
Signature	Date 19/6/2018

Name of Co-Author	Steven Cooper

Name of Co-Author	Steven Cooper
Contribution to the Paper	Supervision of project, Collection of samples, editing of manuscript
Signature	Date 15/6/2018

Name of Co-Author	Andrew Austin
Contribution to the Paper	Supervision of project, Editing of the manuscript
Signature	Date 15/6/18.

Name of Co-Author	James Mitchell
Contribution to the Paper	Supervision of project, Discussion of concepts, editing of manuscript
Signature	Date 25/6/18

Please cut and paste additional co-author panels

CHAPTER 4: Characterising the microbiome of Australian subterranean diving beetles

Josephine Hyde[1,6] Lisa M. Dann[2] William F. Humphreys[3,4] Steven J. B. Cooper[1,5] Andrew D. Austin[1] James G Mitchell[2]

1. Australian Centre for Evolutionary Biology and Biodiversity, School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia
2. College of Science and Engineering at Flinders University, Adelaide, South Australia
3. Western Australian Museum, Welshpool DC, Western Australia 6986, Australia
4. School of Animal Biology, University of Western Australia, Crawley, Western Australia 6009, Australia
5. Evolutionary Biology Unit, South Australian Museum, North Tce, Adelaide, SA 5000, Australia
6. Corresponding author. Email: josephine.hyde@adelaide.edu.au

Abstract

The microbial communities found within the gut of organisms are vital for host health. These gut microbiomes are determined by host genetics and their environment. Model organisms are used to explore the relationship between the microbiome, host and environment. However, most model organisms are laboratory-raised, which significantly affects their microbiome. We present the first gut microbial analyses, based on shot-gun sequencing (metagenomic analyses), of six species of subterranean diving beetles from two different life stages and two separate calcrete aquifers in the Yilgarn Region of Western Australia. The results showed distinct taxonomic patterns within the microbial communities, which were related particularly to the life stage of the beetle. Specifically, microbial communities in the adult beetles had more unclassified genera than larval beetles, while *Acinetobacter* dominated the microbiome of larval beetles. *Wolbachia* was found in two of the six species, which is the first time this genus have been reported in subterranean aquatic insects. *Spiroplasma* was also found in three of the beetle species, suggesting that there may be a bacterial influence on the reproduction of some of the beetle species. The majority of sequences were unclassifiable, suggesting that beetle microbiomes contain a large number of novel organisms.

4.1 Introduction

Animals are holobionts, and it has been suggested that no insect description is complete without the inclusion of their associated microbiomes (Gill *et al*., 2006; Prosdocimi *et al*., 2015). The composition of gut microbiomes reflects the natural selection of microbes and is biased towards mutual cooperation and stability within this complex ecosystem (Dillon and Dillon 2004). Gut microbiomes have a direct beneficial effect on the host. For example, gut symbionts improve the host immune system (Lee *et al*., 2013), prevent pathogenic colonisation (Ryu *et al*., 2008), and alter host development (Shin *et al*., 2011).

As our understanding of how a microbiome influences host health and disease, a complete understanding of how the gut microbiome is initially assembled and maintained is becoming increasingly important (Kinross *et al*., 2008; Blaser, 2014). While many animal microbiomes have been reported, comprehensive analyses have been limited, and the distribution and abundance of common bacterial phyla remain largely unexplored (Degli Esposti and Romero, 2017). Model host-microbe systems provide the opportunity for experimental control to dissect and understand their complexity (Kostic *et al*., 2013). Common model organisms for investigating the gut microbiome include the Hawaiian bobtail squid (*Euprymna scolopes*), the fruit fly (*Drosophila melanogaster*), the zebrafish (*Danio rerio*), and the mouse (*Mus musculus*) (Kostic *et al*., 2013). These models are helping to unravel the complex relationship that host genetics and environment have with the composition of the gut microbiome. *Drosophila*, in particular, is an important model system for a variety of reasons, not only because it is a model organism used in genetics, development and disease, but also because it is providing a better understanding of the interactions between the host genome and microbiome. Understanding the microbiome of this model species can also be extrapolated not only to other arthropods but humans as well.

The subterranean dytiscid beetles from the arid zone of central Australia include species from four different genera, specifically *Paroster*, *Limbodessus*, *Neobidessodes* and *Copelatus*, and include about 100 described species (Watts and Humphreys, 2004, 2009). These beetles are incredibly short-range endemics (Harvey, 2002), as all the described species are restricted to a single calcrete aquifer, which generally encompasses <1000 km^2 (Cooper *et al*., 2002; Leys *et al*., 2003; Leijs *et al*., 2012). Phylogenetic analyses suggest that the colonisation and subsequent isolation of the

beetles within aquifers occurred 5-10 million years ago and coincided with aridification of the Australian interior (Leys *et al*., 2003). The majority of the beetles (> 80%) evolved from a small number of surface species; the remaining species are suggested to have speciated underground from a subterranean ancestor (Leys *et al*., 2003, 2012). These subterranean beetles would help in the development of a baseline of wild insect microbiota, and would allow for investigations into the role that evolutionary processes, host genotypes and the environment have had on the establishment and maintenance of their microbiomes.

In this study, we used metagenomic analyses to investigate the gut microbial diversity of six subterranean beetle species from calcretes located within the arid zone of central Western Australia. The species investigated were from two distinct aquifers separated by ca 130 km. To determine whether location influence the gut microbiomes, two distinct sampling sites, a mining site and pastoral site, were chosen. The beetles investigated included species that had evolved from surface, and subterranean ancestors and both adults and larvae were analysed. Our goal was to start to develop a baseline for understanding microbial diversity for these beetles and to answer specific questions about the ecology of their microbial communities, i.e. (i) What is the make-up of the bacterial communities? (ii) Is there an identifiable microbiome pattern for each of the beetle species? (iii) Are any differences in the microbial communities related to the developmental stage of the host? (iv) Are microbial communities site specific?

4.2 Experimental Procedures

The first site, where beetle species were studied, was a calcrete located on the Sturt Meadows pastoral station in the Yilgarn Region, Western Australia with an area of ~43 km^2. The surface vegetation is an open Acacia woodland with lowland shrubs. In a few places, the calcrete is exposed on the surface, whereas for the most part, the top of the calcrete is up to 2 m below the surface (Allford *et al*., 2008; Bradford *et al*., 2013). The Sturt Meadows calcrete stygofauna consists of 18 known macroinvertebrate taxa, including a sister species triplet of dytiscid diving beetles *Paroster macrosturtensis* (large), *P. mesosturtensis* (medium), and *P. microsturtensis* (small) (Watts and Humphreys 2006). The second study site was a calcrete located on Laverton Downs station next to the Mt Windarra underground mine, approximately 130 km north-east of the Sturt Meadows site. The surface vegetation was also open Acacia woodland with lowland shrubs. The Laverton calcrete stygofauna is thought to be more diverse than

the Sturt Meadows calcrete, but is yet to be fully characterised; it contains three dytiscid diving beetle species *Limbodessus palmulaoides* (large), *L. lapostaae* (small), and *L. windarraensis* (medium) (Watts and Humphreys 2006). The two calcretes are on separate palaeodrainages, located upstream of salt lakes.

4.2.1 Specimen collection:

Fifteen beetles were collected including *L. palmulaoides* (four individuals), *P. macrosturtensis* (six individuals), *P. microsturtensis* (two individuals) adults and larvae, and *P. mesosturtensis, L. lapostaae*, and *L. windarraensis* (one individual each) adults (Table 4.1). The collected specimens were identified, and all, but two, were preserved in liquid Nitrogen and stored at -80 °C in the South Australian Regional Facility for Molecular Ecology and Evolution (University of Adelaide). An additional specimen (25545) was preserved in 100% ethanol and then stored at -20 °C in the Australian Centre for Evolutionary Biology and Biodiversity (University of Adelaide). The final specimen (25542) was euthanised with 100% ethanol just before DNA extraction occurred. The collection localities, dates and collectors are listed in Table 4.1.

4.2.2 DNA extraction and sequencing:

Genomic DNA was extracted from whole specimens using a modified Gentra Pure-Gene DNA purification kit protocol (Gentra systems Minneapolis MN, USA). All extractions were performed inside a UV hood to reduce the amount of surface microbial contamination. The DNA concentrations were checked before library construction using a Quantus Fluorometer. The genomic DNA was sent for library construction, and sequencing at the Australian Genome Research Facility (AGRF) in Adelaide; libraries were prepared using a Nextera DNA library prep kit (Caruccio 2011) and methods specified by the manufacturer. Each library contained only one individual specimen and between five and six libraries were run per lane of an Illumina MiSeq platform (one run with 150-bp paired-end reads, and two runs with 300-bp paired-end reads).

4.2.3 Taxonomic and Functional Analyses

The resulting raw DNA sequences were sorted into specimen groups based on their index barcodes, which were then removed, and sequences with read lengths less than 30 bp, containing Ns or quality scores < 30 were removed using TRIMMOMATIC (Bolger *et al*., 2014) and PRINSEQ (Schmieder and Edwards, 2011). Host-associated

Table 4.1 Collection information for the dytiscid beetles species collected from the Laverton and Sturt Meadows calcretes.

Genus	Species	Life Stage	Location	Collection date	Collector	Sample Number	Collection
Limbodessus	*lapostaae*	adult	Laverton, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper;	27825	Liquid Nitrogen
Limbodessus	*palmulaoides*	adult	Laverton, Western Australia	Sep-15	K.K. Jones; S.J.B. Cooper; B. Langille	25542	Ethanol
Limbodessus	*palmulaoides*	adult	Laverton, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper;	27821	Liquid Nitrogen
Limbodessus	*palmulaoides*	larva	Laverton, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper;	28086	Liquid Nitrogen
Limbodessus	*palmulaoides*	larva	Laverton, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper;	28088	Liquid Nitrogen
Limbodessus	*windarraensis*	adult	Laverton, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper;	27820	Liquid Nitrogen
Paroster	*macrosturtensis*	adult	Sturt Meadows, Western Australia	Sep-15	K.K. Jones; S.J.B. Cooper; B. Langille	25544	Liquid Nitrogen
Paroster	*macrosturtensis*	adult	Sturt Meadows, Western Australia	Nov-06	W.F. Humphreys; S.J.B. Cooper; A. Allford	25545	Ethanol
Paroster	*macrosturtensis*	adult	Sturt Meadows, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper; J. Hyde	27822	Liquid Nitrogen
Paroster	*macrosturtensis*	larva	Sturt Meadows, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper; J. Hyde	28084	Liquid Nitrogen
Paroster	*macrosturtensis*	larva	Sturt Meadows, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper; J. Hyde	28085	Liquid Nitrogen
Paroster	*macrosturtensis*	larva	Sturt Meadows, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper; J. Hyde	28089	Liquid Nitrogen
Paroster	*mesosturtensis*	adult	Sturt Meadows, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper; J. Hyde	27823	Liquid Nitrogen
Paroster	*microsturtensis*	adult	Sturt Meadows, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper; J. Hyde	27824	Liquid Nitrogen
Paroster	*microsturtensis*	larva	Sturt Meadows, Western Australia	Apr-15	W.F. Humphreys; S.J.B. Cooper; J. Hyde	25543	Liquid Nitrogen

sequences were removed using Bowtie2 (Langmead and Salzberg, 2012) (Table S4.1). Additionally, possible prey-associated sequences were also removed (Table S4.1). Sequences were then assembled into contigs using SPAdes, selecting for k-mer lengths 21, 33, 55 and 77, to balance between sensitivity and specificity, and using the –only-assembler function (Bankevich *et al*., 2012) and binned using MetaBAT on the default setting with three minimum samples (Kang *et al*., 2015). The resulting bins were taxonomically and functionally annotated using the default settings in FOCUS (Silva *et al*., 2014) and SUPERFOCUS (Silva *et al*., 2015). The functional data were further analysed in STAMP 2.1.3 (Parks *et al*., 2014), in which Welch's t-test was used to compare the data by genus and developmental stage via clustering-based subsystem level one using a p-value ≤ 0.05. Additionally, bacterial sequences were run through the Anvi'o (Eren *et al*., 2015) metagenomics workflow where they were taxonomically annotated using Centrifuge (Kim *et al*., 2016) and functionally annotated using PROKKA (Seemann, 2014). These two annotation pipelines were then compared to each other to form a consensus annotation.

4.2.4 Statistical analysis

Assembled sequences were analysed using PRIMER (v6), where data were square-root transformed with high abundance species down-weighted using the default setting. Metric multidimensional scaling (MDS) analysis was performed to determine the similarity of the samples by location. Presence/absence data were then used for canonical analysis of principal coordinates (CAP) to determine the similarity between species. SIMPER analysis was performed to determine the relative similarity of species contributions between samples. A one-way analysis was performed with a cut-off of 90% for low contributions, and the factor group used was beetle species. One-way ANOSIM was also conducted, with 999 permutations, to determine if there was a significant difference between the sample sites (Clarke, 1993; Clarke and Warwick, 2001).

4.3 Results

Raw unjoined sequence reads in FASTQ format underwent quality control where low-quality sequences and non-bacterial sequences were removed (Table S4.2). A total of 0.38% of all sequences were removed from the beetle samples from Sturt Meadows, while 0.40% of all sequences from the Laverton beetle samples were removed. On the assumption that after running the sequences through the complete

pipeline only bacterial sequences remained, comparisons were made between the bacterial microbiomes of the beetles at different locations, and developmental stages via MDS. These analyses showed that the beetle gut microbiomes cluster by location (Fig. 4.1). CAP analyses also showed that the bacteria cluster by host size and species (Fig. 4.2).

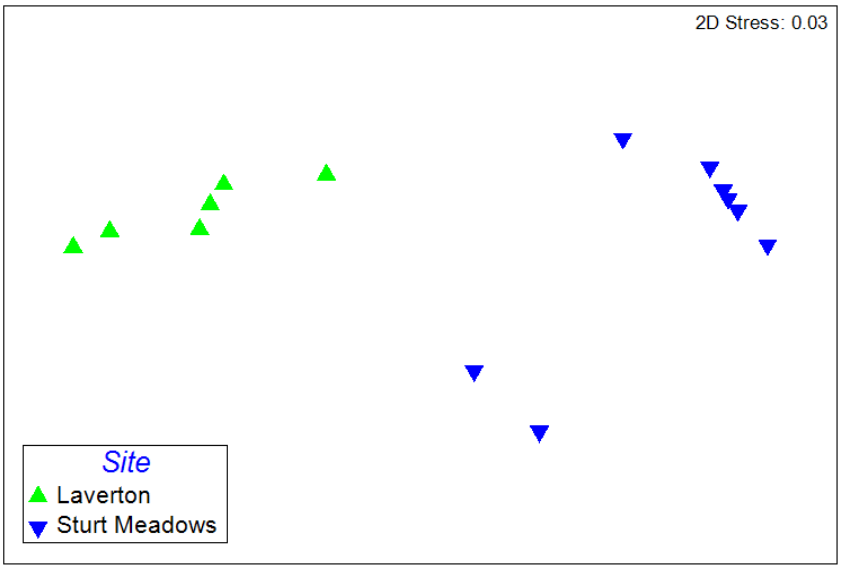

Fig. 4.1 Multidimensional scaling analysis of bacterial communities by beetle host location. Data were normalised and transformed by square root before CAP analysis was performed. Blue triangles are all beetles found at the Sturt Meadows calcrete regardless of life stage and green triangles are all beetles found at the Laverton calcrete regardless of life stage.

Table 4.2 The total number and relative abundance of the different bacterial phyla by collection location.The None claasification refers to bacterial sequences that did not match anything in the database

	Laverton		Sturt Meadows	
Phylum	No. of Bins	% of Bins	No. of Bins	% of Bins
Proteobacteria	21	58.33	24	50.00
Bacteroidetes	3	8.33	3	6.25
None	9	25.00	15	31.25
Tenericutes	1	2.78	0	0.00
Actinobacteria	2	5.56	4	8.33
Firmicutes	0	0.00	2	4.17

For *Limbodessus*, 36 bins were identified of which 27 could be classified as a known phylum (Tables 4.2, Fig. 4.3). Of these, a further 17 could be identified to genera. Of the remaining bins, five matched multiple genera, despite the contamination percentage being below the recommended contamination level, and were classified as Unknown.

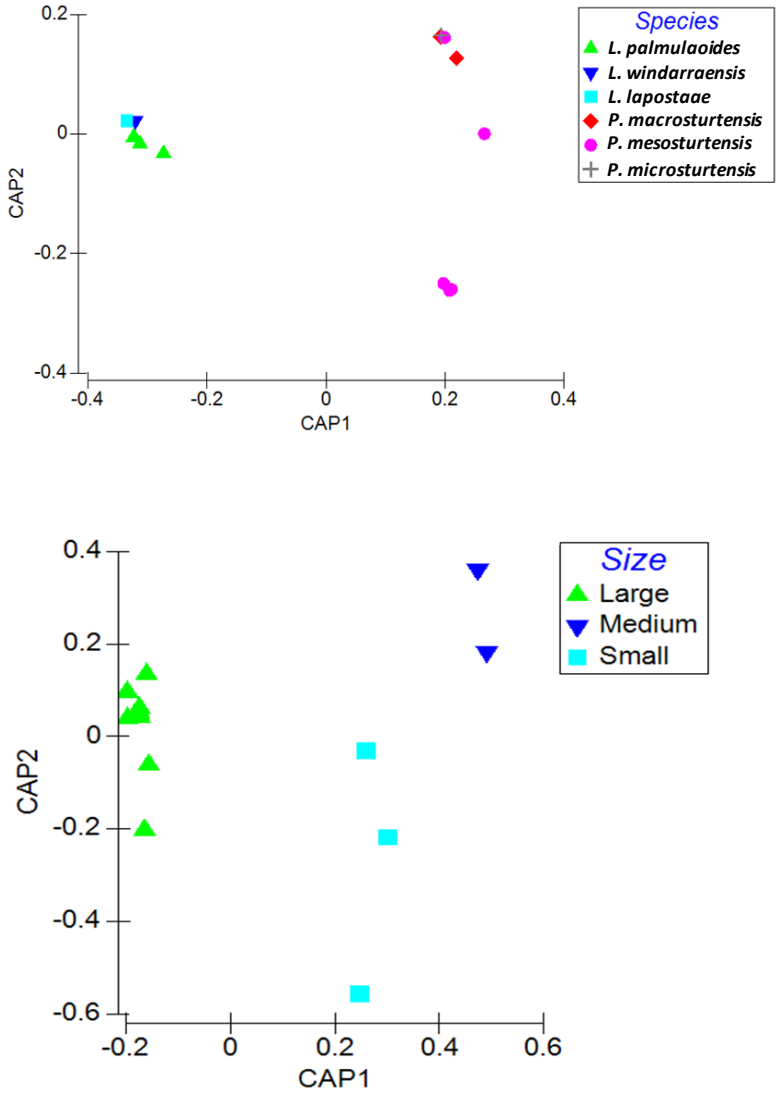

Fig. 4.2 (a) Canonical Analysis of Principal Coordinates (CAP) analysis of the bacterial communities of the beetles by beetle species (b) CAP analysis of the bacterial communities of the beetles by beetle size. Beetle species in each calcrete are classified by small, medium and large. Large beetles were *Limbodessus palmulaoides* and *Paroster macrosturtensis*, medium beetles were *L. windarraensis* and *P. mesosturtensis* and small beetles were *L. lapostaae* and *P. microsturtensis*

Table 4.3 The total number and relative abundance of the different bacterial genera by collection location..The None claasification refers to bacterial sequences that did not match anything in the database. The Unknown classification refers to bacterial sequences that were under the contamination threshold, but still matched multiple genera

	Laverton		Sturt Meadows	
Genus	No. of Bins	% of Bins	No. of Bins	% of Bins
None	14	38.89	20	41.67
Acinetobacter	2	5.56	12	25.00
Unknown	5	13.89	3	6.25
Mycobacterium	1	2.78	3	6.25
Stenotrophomonas	0	0.00	3	6.25
Bacillus	0	0.00	2	4.17
Pseudomonas	0	0.00	2	4.17
Pseudonocardia	0	0.00	1	2.08
Serratia	0	0.00	1	2.08
Shewanella	0	0.00	1	2.08
Methyloversatilis	2	5.56	0	0.00
Achromobacter	1	2.78	0	0.00
Acidovorax	1	2.78	0	0.00
Alcanivorax	1	2.78	0	0.00
Dyadobacter	1	2.78	0	0.00
Microbacterium	1	2.78	0	0.00
Muricauda	1	2.78	0	0.00
Niastella	1	2.78	0	0.00
Ochrobactrum	1	2.78	0	0.00
Pannonibacter	1	2.78	0	0.00
Sphingopyxis	1	2.78	0	0.00
Spiroplasma	1	2.78	0	0.00
Wolbachia	1	2.78	0	0.00

The rest of the bins did not match to anything on either the RefSeq or FOCUS databases and were classified as None (Table 4.2). The most abundant phylum recorded for the beetles in Laverton was Proteobacteria (58.8 %), followed by Bacteroidetes (8.8 %). The most abundant genera were *Acinetobacter* and *Methyloversatilis* (both 5.6 %) (Table S4.3).

For *Paroster* at Sturt Meadows, 48 bins were identified of which 33 could be classified to a known phylum (Table 4.3, Fig. 4.3). Of these 25 could be identified to genera. Of the remaining bins, three matched to multiple genera despite the

contamination percentage being below the recommended level (and were classified asUnknown). The rest of the bins did not match to anything in the database (and were classified as None) (Table 4.3). The most abundant phylum for beetles in Sturt meadows was Proteobacteria (52.1 %) followed by Actinobacteria (8.3 %). The most abundant genera were *Acinetobacter* (25.0 %), followed by *Mycobacterium* and *Stenotrophomonas* (both 6.3 %) (Table S4.4).

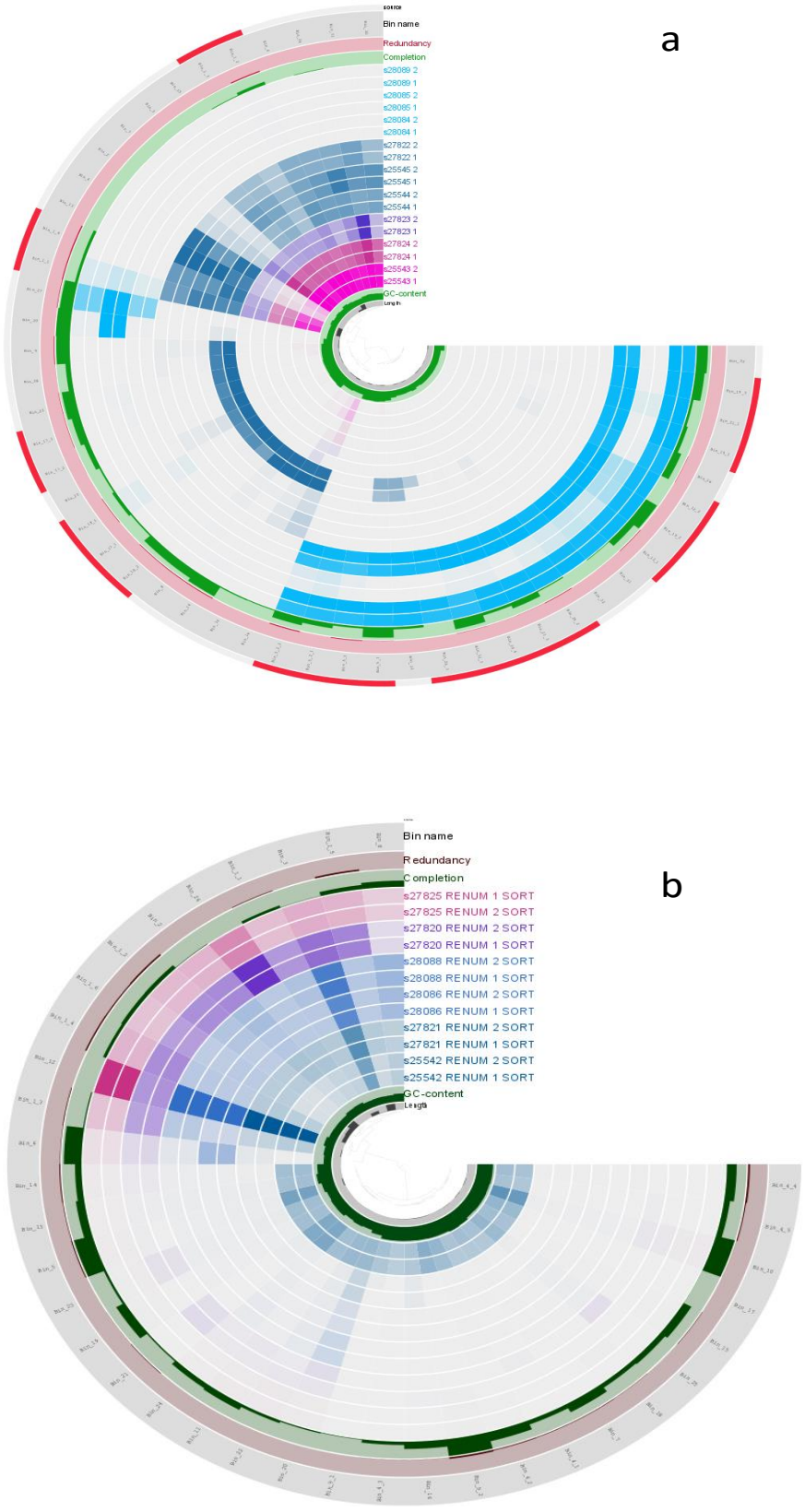

Fig. 4.3 a) Anvi'o plot of the abundance of bacterial species at Sturt Meadows by species; pink, purple and blue lines represent *P. microsturtensis*, *P. mesosturtensis*, and *P. macrosturtensis* respectively. Dark pink and blue lines represent adult individuals, and light pink and blue lines represent larvae. Bin name corresponds to SI Table 4 b) Anvi'o plot of the abundance of bacterial species at Laverton by species; the pink, purple and blue lines represent *L. lapostaae*, *L. windarraensis* and *L. palmuloaides* respectively Dark blue lines are adult individuals and light blue lines are larvae.

Functional gene analysis at SEED subsystem level 1 identified 34 different subsystems among the bacteria from Sturt Meadows, and 33 different subsystems among bacteria from Laverton (Fig. S4.1). The 'Plant cell walls and outer surfaces' subsystem category was present at Sturt Meadows but absent at Laverton (Fig. S4.1). In the Laverton calcrete, the top 10 most abundant subsystems across the six individual beetles resulted in 19 subsystems. Three subsystems ('Protein Metabolism', 'Carbohydrates', and 'Cofactors, Vitamins, Prosthetic Groups and Pigments') were shared across all bacterial genera in the calcrete (Table S4.5).

In the Sturt Meadows calcrete, the top 10 most abundant subsystems across the individual beetles resulted in 18 subsystems; two subsystems, 'Carbohydrates', and 'Amino Acids and Derivatives' were shared across all bacterial genera in the calcrete (Table S4.5). Each calcrete had two subsystems in the top 10 subsystems that were unique to each calcrete. 'Motility and Chemotaxis', and 'Phages, Prophages, Transposable elements and Plasmids' were unique to Laverton and 'Metabolism of Aromatic Compounds', and 'Phosphorus Metabolism' were unique to Sturt Meadows. The analyses returned two functional differences when comparing by location: 'Protein Metabolism', and 'Cofactors, Vitamins, Prosthetic Groups and Pigments' (Fig. S4.1).

Comparisons among the six beetle species suggest that gut microbial diversity was strongly associated with location, with only two known genera, *Acinetobacter* and *Mycobacterium*, shared across the two calcretes. SIMPER analysis showed that Laverton had an average similarity of 34.5%, while Sturt Meadows had a lower average similarity of 24.9% (Table S4.6). Comparison between the two sites revealed an average dissimilarity of 99.6%. The main bacterial groups driving the dissimilarity within/between the sites were species from within the None classification. The three beetle species with more than two individuals and therefore enough individuals to calculate within species similarities ranged in similarity from *P. microsturtensis* (64.3%) to *P. macrosturtensis* (20.6%). Only one species pair between the two calcretes did not have a species dissimilarity of 100%, which was *P. macrosturtensis* and *L. palmulaoides*, which had a dissimilarity of 99.01%. In both calcretes, small and medium-sized beetle species (i.e. *P. mesosturtensis* and *P. microsturtensis*, respectively) had much lower species dissimilarity between each other than when compared with the largest species, *P. macrosturtensis* (Table S4.6). As for the site dissimilarity

calculations the bacterial genera that appeared to be driving the similarity within species and dissimilarity between species from within the None classification.

4.4 Discussion

Here we show that the gut microbial communities within stygobiotic beetles species are heavily influenced by the location (Fig. 4.1), developmental stage and size of the beetle (Fig. 4.2). The results showed that the most abundant bacterial phylum at the two sampling sites was Proteobacteria (Table 4.2). This phylum has been found in other insect microbiomes, and all other phyla found in the beetles here have been found in other insect microbiomes (Yun *et al*., 2014). The genera that were found at Laverton and Sturt Meadows were mostly genera known to occur in other arthropod microbiomes (Andreotti *et al*., 2011; Shelomi *et al*., 2013; Montagna *et al*., 2015; Degli Esposti and Romero, 2017). However, Sturt Meadows and Laverton beetles had bacteria that are known to be endosymbiotic with insects, specifically *Spiroplasma* and *Wolbachia* at Laverton and *Pseudonocardia* at Sturt Meadows (Table 4.3). Additionally, a genus that is known to include pathogenic insect bacteria, *Serratia*, was also present at Sturt Meadows (Table 4.3).

Marine associated bacteria were present at both sites (Table 4.3). Particularly, *Alcanivorax*, a hydrocarbonoclastic bacterium, was found at the Laverton calcrete (Table 4.3). Members of this genus are n-alkane-degrading marine bacteria which have been found in crude-oil-containing seawater when nitrogen and phosphorus nutrients are abundant. These bacteria may be present as the calcrete is close to a nickel sulphide mine and an active calcrete quarry, which potentially are releasing hydrocarbons into the environment as part of the mining process. Alternatively, *Alcanivorax* may be present in the calcrete due to the presence of *Acacia* and *Eucalyptus* on the surface. Both plant groups have n-alkanes present in their leaf wax, and it has been shown that the concentration of n-alkanes increases in plants in the arid centre of Australia (Hoffmann *et al*., 2013). *Shewanella* was present at Sturt Meadows (Table 4.3). This bacterium has been found previously in the guts of marine invertebrates (Kim *et al*., 2007; Leigh *et al*., 2017), and are facultative anaerobes which can use a wide variety of metals as electron acceptors, including iron (Tiedje, 2002). While Sturt Meadows does not have any active mining, the geology of the site includes nickel, cobalt and laterite ore. The high amounts of these minerals in the environment may be enhancing the presence of *Shewanella* in the guts of the beetles present.

When comparing the gut microbial communities from the different host beetles, a significant difference was observed between adult and larval *P. macrosturtensis* (Fig. 4.3). Specifically, the adult beetles had more unclassified bacterial taxa, while the majority of larval bacterial sequences were *Acinetobacter*. Studies of the gut microbiome in other insects have shown that the presence of certain bacterial species was negatively correlated with the presence of other bacterial species. For example, in the locust *Schistocerca gregaria*, the abundance of the bacterium *Serratia marcescens*, was negatively correlated with the abundance of symbiotic gut bacterial species (Dillon and Charnley, 2002; Dillon *et al*., 2005). Therefore, the high abundance of *Acinetobacter* in *P. macrosturtensis* larvae could be decreasing the abundance of the other bacterial species present in the beetles at Sturt Meadows (Douglas, 2015). Another pattern that has been found in mosquitoes is that *Acinetobacter* prevalence correlates with gender (Minard *et al*., 2012); as the gender of the beetle larvae in this study are unknown, it cannot be ruled out that there is a gender imbalance between the larval and adult beetles sequenced, thus skewing the results.

There are consistent co-occurrences of beetle species with distinct non-overlapping sizes in multiple calcretes (Cooper *et al*., 2002; Humphreys *et al*., 2009; Leijs *et al*., 2012). Studies have hypothesised that speciation leading to multiple sympatric beetle species has occurred; because of ecological-niche partitioning within the calcrete, where the species use the calcrete resources differently in order for them to co-exist (Leys *et al*., 2003; Juan *et al*., 2010; Leijs *et al*., 2012; Bradford *et al*., 2014). Interestingly, the differentiation of the bacterial microbiome between the two life stages of *P. macrosturtensis* may be influenced by differences in their diet.

Functional analyses identified sulfur metabolism at both sites, suggesting chemoautotrophic metabolism is occurring (Table S4.5). Previous research has suggested that the Sturt Meadows calcrete may derive energy from plant sources (Bradford *et al*., 2014). However, it is likely that the energy in the calcrete comes from the bacteria utilising a combination of energy sources, with additional carbon input coming from both plant roots and recharge events (Hyde *et al*., 2018). Additionally, at both calcretes functional genes from bacteria associated with photosynthesis were found (Table S4.5). As these calcrete systems are in complete darkness, the occurrence of these photosynthesis genes may indicate the presence of cyanobacteria that, while obviously not photosynthetic in the dark, have preserved their photosynthetic pathways

(Baulina, 2012). These bacteria most likely enter the calcrete during recharge after heavy rainfall and metabolise heterotrophically within the calcrete. It is also possible that the bacteria are dormant or even dying after entering the calcrete if they are not capable of alternative methods of metabolism.

Previous research has shown that there is a significant difference in the microbiomes of laboratory-raised compared to wild caught animals (Chandler *et al*., 2011). To confirm this finding, one of the beetle species investigated, *L. palmulaoides*, was collected and kept in the laboratory for approximately six months before being euthanised. In the microbiome of this captive specimen, there was an increase in abundance of all bacterial genera that were found in the wild caught *L. palmulaoides* specimens, except for two genera: *Wolbachia,* which was equally abundant in both captive and wild-caught specimens, and *Spiroplasma,* which was more abundant in the wild caught individuals (Fig 4.3). However, no unique bacterial genera were present in the laboratory individual, however, with only a single individual studied, it is difficult to extrapolate from these results.

Two of the three beetle species from Laverton possessed *Wolbachia* in their microbiome (Table 4.3). This genus comprises obligate intracellular bacteria that infect invertebrates, which occurs in both a parasitic and mutualistic relationship depending on its host (Weeks *et al*., 2007). *Wolbachia* is vertically transmitted and is estimated to infect between 25-70% of insect species (Kozek and Rao, 2007). *Wolbachia* infections have been recorded in epigean species of dytiscid beetles (Küchler *et al*., 2009), and subterranean insects such as termites (Salunke *et al*., 2010). The presence of *Wolbachia* in the beetles here is, therefore, interesting largely because of its total absence in the Sturt Meadows calcrete and from one of the species in the Laverton calcrete. *Wolbachia* is found in all the individuals sequenced from Laverton and two of the three species present except the one *L. windarraensis* individual. The latter is the medium-sized beetle, and it is possible that *Wolbachia* is found in this species and just that the particular individual was not infected. Due to its role in reproduction, *Wolbachia* has been implicated in speciation of multiple species of insects such as in *Drosophila* (Telschow *et al*., 2005). The lack of *Wolbachia* in one of the species in the calcrete presents an interesting mechanistic hypothesis on how speciation might have occurred. A comprehensive survey into the presence of *Wolbachia* in the Laverton calcrete and whether *L. windarraensis* also harbours *Wolbachia* is required. Additionally, more

research into the different strains of the bacterium that are present is needed to determine if they could have caused a reproductive barrier in the past between the species.

Spiroplasma was also found in beetle species of the Laverton calcrete and is another vertically inherited intracellular bacterial genus that is known to alter the reproduction of insects (Duron *et al*., 2008). Unlike *Wolbachia*, *Spiroplasma* was found in all three beetle species in the Laverton calcrete, although it was once again absent in the Sturt Meadows calcrete. In other beetle species, *Spiroplasma* is known to be lethal to male embryos (Tinsley and Majerus, 2006). In mosquitoes, *Spiroplasma* has been shown to protect against the sterilising effects of parasitic nematodes (Jaenike *et al*., 2010b). Therefore, this bacterial genus may be playing a defensive symbiotic role for the beetle species within the Laverton calcrete. Additionally, in some insects, a positive association between *Wolbachia* and *Spiroplasma* has been identified, and it has been suggested that they may have a complex cooperative association (Jaenike *et al*., 2010a).

4.5 Conclusion

Here we show distinct taxonomic differences between the gut microbiomes of stygobiontic diving beetles from different life stages and locations. Specifically, life stage led to drastic alterations in the microbiome between individuals in the species *P. macrosturtensis*, which suggests a possible alteration in trophic position during development. This study also provides evidence for chemoautotrophic bacteria, which may provide an additional source of energy into the subterranean ecosystem. At the Laverton calcrete, endosymbiotic bacteria were also found which may be influencing the reproduction of their hosts. These results provide the first investigation into the microbiome of subterranean diving beetles and provide insights into the wild insect microbiome.

Acknowledgements

The authors would like to thank Flora, Peter and Paul Axford of the Sturt Meadows Station for providing both accommodation and access to their property. We would also like to thank Dr Robert Edwards and Dr Elizabeth Dinsdale for their advice during this project. This research was funded by Australian Research Council linkage grants (LP0348753, LP 100200494 and LP140100555) to SJBC, WFH, ADA and PM, with industry partners Newmont Australia, Placer Dome Asia Pacific, Minara Resources Limited, the Western Australian Museum, the South Australian Museum, Rio Tinto,

Biota Environmental Sciences, Bennelongia, Department of Biodiversity, Conservation and Attractions (WA).

References

Andreotti, R., de León, A.A.P., Dowd, S.E., Guerrero, F.D., Bendele, K.G., and Scoles, G.A. (2011) Assessment of bacterial diversity in the cattle tick Rhipicephalus (Boophilus) microplus through tag-encoded pyrosequencing. *BMC Microbiology* 11: 6.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S. *et al*. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455-477.

Baulina, O.I. (2012) Ultrastructural plasticity of cyanobacteria under dark and high light intensity conditions. In *Ultrastructural plasticity of Cyanobacteria*: Springer, Heidelberg pp. 11-63.

Blaser, M.J. (2014) The microbiome revolution. *The Journal of Clinical Investigation* 124: 4162-4165.

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

Bradford, T.M., Humphreys, W.F., Austin, A.D., and Cooper, S.J.B. (2014) Identification of trophic niches of subterranean diving beetles in a calcrete aquifer by DNA and stable isotope analyses. *Marine and Freshwater Research* 65: 95-104.

Chandler, J.A., Lang, J.M., Bhatnagar, S., Eisen, J.A., and Kopp, A. (2011) Bacterial communities of diverse Drosophila species: ecological context of a host–microbe model system. *PLoS Genetics* 7: e1002272.

Clarke, K.R. (1993) Non-parametric multivariate analyses of changes in community structure. *Austral Ecology* 18: 117-143.

Clarke, K.R., and Warwick, R. (2001) Change in marine communities. An approach to statistical analysis and interpretation. 2nd Edition PRIMER-E Ltd: Plymouth, United Kingdom.

Cooper, S.J.B., Hinze, S., Leys, R., Watts, C.H.S, and Humphreys, W.F. (2002) Islands under the desert: molecular systematics and evolutionary origins of stygobitic water beetles (Coleoptera: Dytiscidae) from central Western Australia. *Invertebrate Systematics* 16: 589-590.

Degli Esposti, M., and Romero, E.M. (2017) The functional microbiome of arthropods. *PLoS One* 12: e0176573.

Dillon, R., and Charnley, K. (2002) Mutualism between the desert locust Schistocerca gregaria and its gut microbiota. *Research in Microbiology* 153: 503-509.

Douglas, A.E. (2015) Multiorganismal insects: diversity and function of resident microorganisms. *Annual Review of Entomology* 60: 17-34.

Duron, O., Bouchon, D., Boutin, S., Bellamy, L., Zhou, L., Engelstädter, J., and Hurst, G.D. (2008) The diversity of reproductive parasites among arthropods: Wolbachia do not walk alone. *BMC Biology* 6: 27.

Eren, A.M., Esen, Ö.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., and Delmont, T.O. (2015) Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3: e1319.

Gill, S.R., Pop, M., DeBoy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S. *et al*. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* 312: 1355-1359.

Harvey, M. S. (2002). Short-range endemism amongst the Australian fauna: some examples from non-marine environments. *Invertebrate Systematics*, 16(4), 555-570.

Hoffmann, B., Kahmen, A., Cernusak, L.A., Arndt, S.K., and Sachse, D. (2013) Abundance and distribution of leaf wax n-alkanes in leaves of Acacia and Eucalyptus trees along a strong humidity gradient in northern Australia. *Organic Geochemistry* 62: 62-67.

Humphreys, W.F, Watts, C.H.S., Cooper, S.J.B., and Leijs, R. (2009) Groundwater estuaries of salt lakes: buried pools of endemic biodiversity on the western plateau, Australia. *Hydrobiologia* 626: 79-95.

Hyde, J., Cooper, S.J.B., Humphreys, W.F., Austin, A.D., and Munguia, P. (2018) Diversity patterns of subterranean invertebrate fauna in calcretes of the Yilgarn Region, Western Australia. *Marine and Freshwater Research* 69: 114-121.

Jaenike, J., Stahlhut, J.K., Boelio, L.M., and Unckless, R.L. (2010a) Association between Wolbachia and Spiroplasma within Drosophila neotestacea: an emerging symbiotic mutualism? *Molecular Ecology* 19: 414-425.

Jaenike, J., Unckless, R., Cockburn, S.N., Boelio, L.M., and Perlman, S.J. (2010b) Adaptation via symbiosis: recent spread of a Drosophila defensive symbiont. *Science* 329: 212-215.

Juan, C., Guzik, M.T., Jaume, D., and Cooper, S.J.B. (2010) Evolution in caves: Darwin's 'wrecks of ancient life'in the molecular era. *Molecular Ecology* 19: 3865-3880.

Kang, D.D., Froula, J., Egan, R., and Wang, Z. (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3: e1165.

Kim, D., Song, L., Breitwieser, F.P., and Salzberg, S.L. (2016) Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Research* 26: 1721-1729.

Kim, D., Baik, K.S., Kim, M.S., Jung, B.-M., Shin, T.-S., Chung, G.-H. *et al*. (2007) Shewanella haliotis sp. nov., isolated from the gut microflora of abalone, Haliotis discus hannai. *International Journal of Systematic and Evolutionary Microbiology* 57: 2926-2931.

Kinross, J.M., von Roon, A.C., Holmes, E., Darzi, A., and Nicholson, J.K. (2008) The human gut microbiome: implications for future health care. *Current Gastroenterology Reports* 10: 396-403.

Kostic, A.D., Howitt, M.R., and Garrett, W.S. (2013) Exploring host–microbiota interactions in animal models and humans. *Genes and Development* 27: 701-718.

Kozek, W.J., and Rao, R.U. (2007) The discovery of Wolbachia in arthropods and nematodes–A historical perspective. (Eds. Hoerauf, A., and Rao, R.U.) In *Wolbachia: a bug's life in another bug*: Karger Publishers, Sydney, NSW pp. 1-14.

Küchler, S.M., Kehl, S., and Dettner, K. (2009) Characterization and localization of Rickettsia sp. in water beetles of genus *Deronectes* (Coleoptera: Dytiscidae). *FEMS Microbiology Ecology* 68: 201-211.

Langmead, B., and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357.

Lee, K.-A., Kim, S.-H., Kim, E.-K., Ha, E.-M., You, H., Kim, B. *et al*. (2013) Bacterial-derived uracil as a modulator of mucosal immunity and gut-microbe homeostasis in *Drosophila*. *Cell* 153: 797-811.

Leigh, B., Karrer, C., Cannon, J.P., Breitbart, M., and Dishaw, L.J. (2017) Isolation and characterization of a Shewanella phage–host system from the gut of the tunicate, *Ciona intestinalis*. *Viruses* 9: 60.

Leijs, R., Van Nes, E.H., Watts, C.H.S., Cooper, S.J.B., Humphreys, W.F., and Hogendoorn, K. (2012) Evolution of blind beetles in isolated aquifers: a test of alternative modes of speciation. *PLoS One* 7: e34260.

Leys, R., Watts, C.H.S., Cooper, S.J.B., and Humphreys, W.F. (2003) Evolution of subterranean diving beetles (Coleoptera: Dytiscidae Hydroporini, Bidessini) in the arid zone of Australia. *Evolution* 57: 2819-2834.

Minard, G., Tran, F.H., Raharimalala, F.N., Hellard, E., Ravelonandro, P., Mavingui, P., and Valiente Moro, C. (2012) Prevalence, genomic and metabolic profiles of Acinetobacter and Asaia associated with field-caught *Aedes albopictus* from Madagascar. *FEMS Microbiology Ecology* 83: 63-73.

Montagna, M., Gómez-Zurita, J., Giorgi, A., Epis, S., Lozzia, G., and Bandi, C. (2015) Metamicrobiomics in herbivore beetles of the genus *Cryptocephalus* (Chrysomelidae): toward the understanding of ecological determinants in insect symbiosis. Insect *Science* 22: 340-352.

Prosdocimi, E.M., Mapelli, F., Gonella, E., Borin, S., and Crotti, E. (2015) Microbial ecology-based methods to characterize the bacterial communities of non-model insects. *Journal of Microbiological Methods* 119: 110-125.

Ryu, J.-H., Kim, S.-H., Lee, H.-Y., Bai, J.Y., Nam, Y.-D., Bae, J.-W. *et al*. (2008) Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in *Drosophila*. *Science* 319: 777-782.

Salunke, B.K., Salunkhe, R.C., Dhotre, D.P., Khandagale, A.B., Walujkar, S.A., Kirwale, G.S. *et al*. (2010) Diversity of *Wolbachia* in *Odontotermes spp*.(Termitidae) and *Coptotermes heimi* (Rhinotermitidae) using the multigene approach. *FEMS Microbiology Letters* 307: 55-64.

Schmieder, R., and Edwards, R.A. (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27: 863-864.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30: 2068-2069.

Shelomi, M., Lo, W.-S., Kimsey, L.S., and Kuo, C.-H. (2013) Analysis of the gut microbiota of walking sticks (Phasmatodea). *BMC Research Notes* 6: 368.

Shin, S.C., Kim, S.-H., You, H., Kim, B., Kim, A.C., Lee, K.-A. *et al*. (2011) *Drosophila* microbiome modulates host developmental and metabolic homeostasis via insulin signaling. *Science* 334: 670-674.

Silva, G.G.Z., Cuevas, D.A., Dutilh, B.E., and Edwards, R.A. (2014) FOCUS: an alignment-free model to identify organisms in metagenomes using non-negative least squares. *PeerJ* 2: e425.

Silva, G.G.Z., Green, K.T., Dutilh, B.E., and Edwards, R.A. (2015) SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data. *Bioinformatics* 32: 354-361.

Telschow, A., Hammerstein, P., and Werren, J.H. (2005) The effect of *Wolbachia* versus genetic incompatibilities on reinforcement and speciation. *Evolution* 59: 1607-1619.

Tiedje, J.M. (2002) *Shewanella*—the environmentally versatile genome. *Nature Biotechnology* 20: 1093-1094Tinsley, M.C., and Majerus, M.E.N. (2006) A new male-killing parasitism: *Spiroplasma* bacteria infect the ladybird beetle *Anisosticta novemdecimpunctata* (Coleoptera: Coccinellidae). *Parasitology* 132: 757-765.

Watts, C.H.S., and Humphreys, W.F. (2004) Thirteen new Dytiscidae (Coleoptera) of the genera *Boongurrus* Larson, *Tjirtudessus* Watts and Humphreys and *Nirripirti* Watts and Humphreys, from underground waters in Australia. *Transactions of the Royal Society of South Australia* 128: 99-129.

Watts, C.H.S., and Humphreys, W.F. (2009) Fourteen new Dytiscidae (Coleoptera) of the genera *Limbodessus* Guignot, *Paroster* Sharp, and *Exocelina* Broun from underground waters in Australia. *Transactions of the Royal Society of South Australia* 133: 62-107.

Weeks, A.R., Turelli, M., Harcombe, W.R., Reynolds, K.T., and Hoffmann, A.A. (2007) From parasite to mutualist: rapid evolution of *Wolbachia* in natural populations of *Drosophila*. *PLoS Biology* 5: e114.

Yun, J.-H., Roh, S.W., Whon, T.W., Jung, M.-J., Kim, M.-S., Park, D.-S. *et al*. (2014) Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Applied and Environmental Microbiology* 80: 5254-5264.

Supplementary Figures

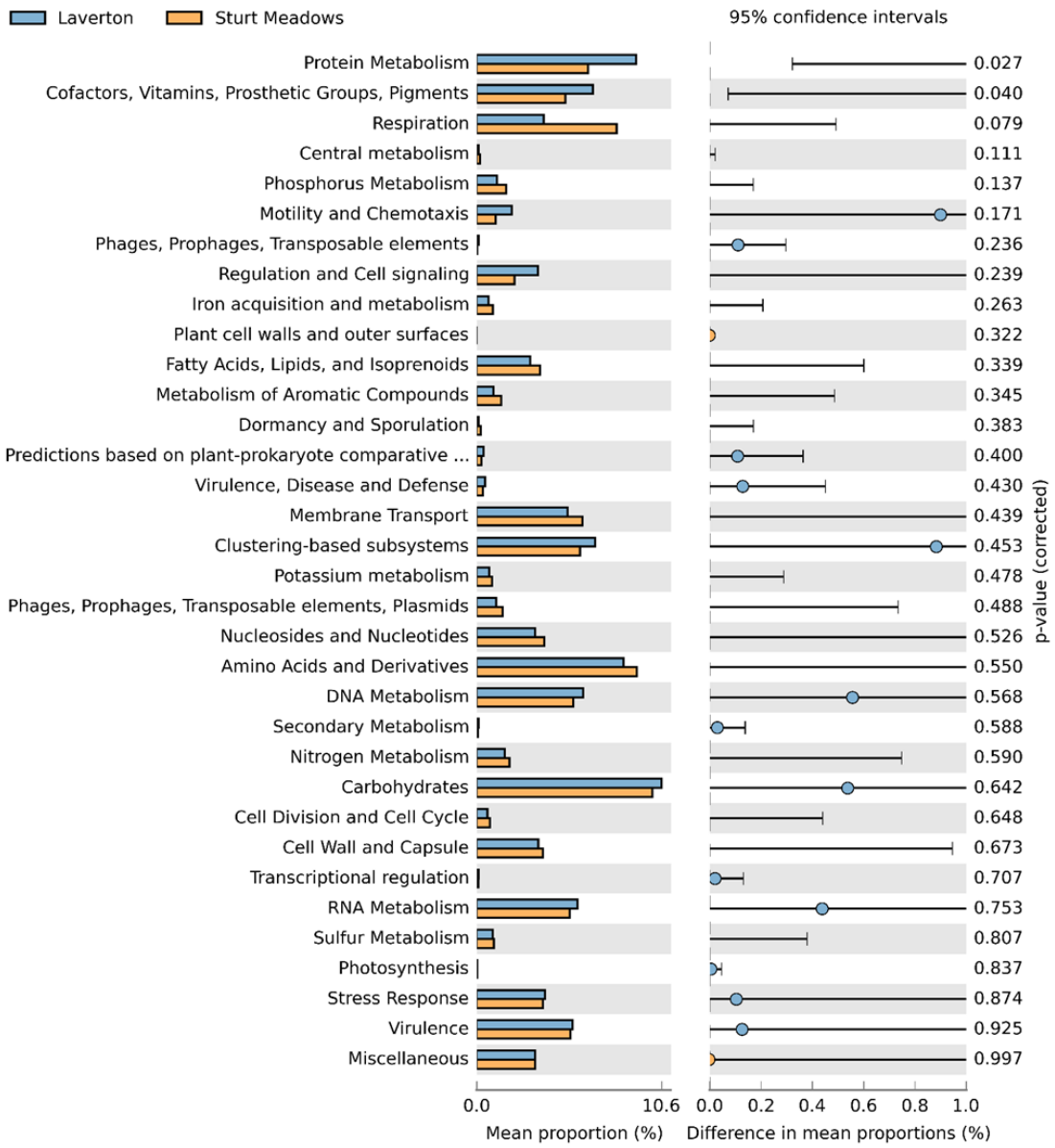

Fig S4.1 Pairwaise post hoc plot of the mean proportion of all functional gene groups at SEED level 1 in the bacterial species by beetle species between Laverton and Sturt Meadows.

Supplementary Tables

Table S4.1 List of sequences (whole genome and mitochondrial genomes and other genetic sequences of calcrete inhabitants, potential prey species and potential contaminants) used to remove non-bacterial sequences from the raw data during pre-processing.

Species	Genbank Accession number	Host-associated	Prey associated	Other
Abacion magnum	NC_021932.1		X	
Acropora tenuis	AF338425.1			X
Ammothea carolinensis	NC_014671.1			X
Ancylostoma caninum	FJ483518.1		X	
Aplysia californica	AY569552.1		X	
Argyroneta aquatica	NC_026863.1		X	
Artemia franciscana	NC_001620.1		X	
Asterias amurensis	AB183559.1			X
Austropotamobius pallipes	NC_026560.1		X	
Balanoglossus clavigerus	NC_013877.1		X	
Branchiostoma lanceolatum	AB478564.1			X
Calanus hyperboreus	NC_019627.1		X	
Campodea lubbocki	DQ529237.1		X	
Caprella mutica	GU130250.1		X	
Caridina gracilipes	KM023648.1		X	
Chabertia ovina	KF660604.1		X	
Corbicula fluminea	KX254564.1			X
Crassostrea gigas	EU672831.1			X
Cristaria plicata	FJ986302.1			X
Cyclopoida sp.	Morph ID		X	
Damithrax spinosissimus	NC_025518.1		X	
Daphnia magna	NC_026914.1		X	
Dendronephthya castanea	NC_023343.1			X
Dermatophagoides farinae	NC_013184.1			X
Didemnum vexillum	KM259617.1		X	
Duplodicodrilus schmardae	NC_029867.1		X	
Ephydatia muelleri	EU237481.1		X	
Fenneropenaeus chinensis	DQ518969.1		X	
Fierscyclops fiersi	Morph ID		X	
Galba pervia	NC_018536.1			X
Gammarus duebeni	NC_017760.1		X	
Geodia neptuni	AY320032.1		X	
Gonodactylus chiragra	DQ191682.1		X	
Halicyclops cf. ambiguus	Morph ID		X	
Halicyclops kieferi	Morph ID		X	
Haliotis discus hannai	KF724723.1			X
Halocynthia roretzi	NC_002177.1		X	
Harpacticoida sp.	Morph ID		X	
Heliocidaris crassispina	NC_023774.1			X
Heterometrus longimanus	KR190462.1		X	
Homarus gammarus	KC107810.1		X	

Ixodes persulcatus	KU935457.1			X	
Japyx solifugus	AY771989.1			X	
Lampsilis ornata	AY365193.1				X
Lepeophtheirus salmonis	EU288200.1			X	
Ligia oceanica	DQ442914.1			X	
Limbodessus palmulaoides	MG912994.1	X			
Lithobius forficatus	AF309492.1			X	
Litopenaeus vannamei	EF584003.1			X	
Macrobrachium nipponense	HQ830201.1			X	
Marsupenaeus japonicus	AP006346.1			X	
Mytilus chilensis	NC_030633.1				X
Narceus annularus	NC_003343.1			X	
Naticarius hebraeus	NC_028002.1				X
Onchidella celtica	AY345048.2				X
Ornithoctonus huwena	AY309259.1				X
Palaemon gravieri	KU899135.1			X	
Panulirus versicolor	NC_028627.1			X	
Paracyclopina nana	NC_012455.1			X	
Parasesarma tripectinis	NC_030046.1			X	
Parhyale hawaiiensis	AY639937			X	
Paroster macrosturtensis	MG912995.1	X			
Paroster mesosturtensis	MG912996.1	X			
Paroster microsturtensis	MG912997.1	X			
Penaeus monodon	NC_002184.1			X	
Phalangium opilio	EU523757.1			X	
Phyxioschema suthepium	NC_020322.1			X	
Pinctada margaritifera	NC_021638.1				X
Pista cristata	EU239688.1			X	
Placopecten magellanicus	DQ088274.1				X
Portunus trituberculatus	AB093006.1			X	
Procambarus clarkii	KT036444.1			X	
Saccoglossus kowalevskii	AY336131.1				X
Schizopera cf. austindownsi	Morph ID			X	
Schmidtea mediterranea	NC_022448.1				X
Scutachiltonia axfordi	KT958022.1, KT958075.1			X	
Scylla paramamosain	FJ827761.1			X	
Scyllarides latus	KC107814.1			X	
Sipunculus nudus	NC_011826.1			X	
Spadella cephaloptera	AY545549.1			X	
Squilla mantis	NC_006081.1			X	
Strongylocentrotus intermedius	NC_023772.1				X
Sturt Meadows Oligochaeta sp. 1	N/A			X	
Sturt Meadows Oligochaeta sp. 2	N/A			X	
Sturt Meadows Oligochaeta sp. 3	N/A			X	
Sturt Meadows Oligochaeta sp. 4	N/A			X	
Sturt Meadows Oligochaeta sp. 5	N/A			X	

Stygochiltonia bradfordae	KT958023.1, KT958077.1		X	
Taenia asiatica	AF445798.2			X
Temnopleurus hardwickii	NC_026200.1			X
Tigriopus californicus	DQ913891.2		X	
Tribolium castaneum	NC_007418.3	X		
Trichuris trichiura	KT449826.1		X	
Triops cancriformis	AB084514.1		X	
Uncultured crAssphage	NC_024711.1			X
Unio douglasiae	KM657954.1			X
Unionicola foili	EU856396.1		X	
Urechis caupo	AY619711.1		X	
Uroctonus mordax	EU523756.1		X	
Yilganiella sturtensis	KT958021.1, KT958073.1, KT958074.1		X	
Bos taurus	AC_000158 .1			X
Homo sapiens	NC_000001.11-NC_000001.11, NC_000001.11			X

Table S4.1 Number of sequences removed during the processing of the raw sequences through the quality control pipeline. Sequences with Ns, low quality sequences and those that matched the host DNA were removed in the intial quality control steps using three different bioinformatics programs; Prinseq, Trimmomatic and Bowtie2.

		Prinseq		Trimmomatic		Bowtie 2	
Sequence number	Raw (pairs)	Ns removed	Pairs surviving	Both surviving (%)	Pairs surviving	Host removed (pairs)	Pairs surviving
25542	4,417,482	36	4,417,451	94.53	4,176,024	29,699	4,146,325
25543	3,056,701	15	3,056,688	95.78	2,927,669	8,858	2,918,811
25544	3,732,270	13	3,732,259	88.39	3,299,016	15,676	3,283,340
25545	1,302,320	8	1,302,314	97.85	1,274,353	17,900	1,256,453
27820	1,082,334	0	1,082,334	99.84	1,080,581	2,993	1,077,588
27821	1,320,946	0	1,320,946	99.90	1,319,659	2,672	1,316,987
27822	1,716,381	0	1,716,381	99.85	1,713,727	9,055	1,704,672
27823	958,312	0	958,312	99.89	957,245	8,852	948,393
27824	1,614,519	1	1,614,518	99.85	1,612,049	8,287	1,603,762
27825	1,476,630	1	1,476,629	99.85	1,474,432	9,035	1,465,397
28084	2,536,527	162	2,536,367	92.28	2,340,492	2,959	2,337,533
28085	2,524,576	194	2,524,384	93.58	2,362,197	2,773	2,359,424
28086	3,598,466	364	3,598,104	96.87	3,485,372	7,349	3,478,023
28088	2,897,727	256	2,897,475	96.12	2,785,015	7,699	2,777,316
28089	2,784,735	226	2,784,511	93.42	2,601,357	2,728	2,598,629

Table S4.3 The taxonomic classification of the bacterial bins (from MetaBAT) from Laverton through the two taxonomic classification programs Centrifuge and FOCUS.Centrifuge output was only at genus level while FOCUS gave classification information from Phylum to Genus.

		Laverton				
	Centrifuge	Focus				
Bin	Genus	Phylum	Class	Order	Family	Genus
Bin_18	*Achromobacter*	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	*Achromobacter*
Bin_13	*Acidovorax*	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Acidovorax*
Bin_22	*Acinetobacter*	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*
Bin_24	*Acinetobacter*	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*
Bin_10	*Alcanivorax*	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Alcanivoracaceae	*Alcanivorax*
Bin_5	*Dyadobacter*	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Cytophagaceae	*Dyadobacter*
Bin_25	*Methyloversatilis*	Proteobacteria	None	None	None	*None*
Bin_4_2	*Methyloversatilis*	Proteobacteria	Betaproteobacteria	None	None	*None*
Bin_7	*Microbacterium*	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	*Microbacterium*
Bin_16	*Mycobacterium*	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	*Mycobacterium*
Bin_19	*None*	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	*Muricauda*
Bin_11	*None*	Bacteroidetes	Flavobacteriia	Flavobacteriales	Cytophagaceae	*Niastella*
Bin_1_1	*None*	None	None	None	None	*None*
Bin_1_2	*None*	None	None	None	None	*None*
Bin_1_3	*None*	None	None	None	None	*None*
Bin_1_4	*None*	None	None	None	None	*None*
Bin_1_5	*None*	None	None	None	None	*None*
Bin_1_6	*None*	None	None	None	None	*None*
Bin_12	*None*	None	None	None	None	*None*
Bin_17	*None*	Proteobacteria	Betaproteobacteria	None	None	*None*
Bin_2	*None*	Proteobacteria	None	None	None	*None*
Bin_20	*None*	None	Gammaproteobacteria	None	None	*None*
Bin_21	*None*	Proteobacteria	None	None	None	*None*
Bin_23	*None*	Proteobacteria	Sphingobacteriia	Sphingobacteriales	None	*None*
Bin_26	*None*	None	None	None	None	*None*
Bin_3	*None*	None	None	None	None	*None*
Bin_14	*Ochrobactrum*	Proteobacteria	Alphaproteobacteria	Rhizobiales	None	*Ochrobactrum*
Bin_15	*Pannonibacter*	Proteobacteria	Alphaproteobacteria	None	None	*None*
Bin_9_2	*Sphingopyxis*	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	*Sphingopyxis*
Bin_6	*Spiroplasma*	Tenericutes	Mollicutes	Entomoplasmatales	Spiroplasmataceae	*Spiroplasma*
Bin_4_1	*Unknown*	Proteobacteria	Alphaproteobacteria	None	None	*None*
Bin_4_3	*Unknown*	Proteobacteria	Alphaproteobacteria	None	None	*None*
Bin_4_4	*Unknown*	Proteobacteria	Alphaproteobacteria	None	None	*None*
Bin_4_5	*Unknown*	Proteobacteria	Alphaproteobacteria	None	None	*None*
Bin_9_1	*Unknown*	Proteobacteria	Alphaproteobacteria	None	None	*None*
Bin_8	*Wolbachia*	Proteobacteria	Alphaproteobacteria	Rickettsiales	Anaplasmataceae	*Wolbachia*

Table S4.4 The taxonomic classification of the bacterial bins (from MetaBAT) from Sturt Meadows through the two taxonomic classification programs Centrifuge and FOCUS.Centrifuge output was only at genus level while FOCUS gave classification information from Phylum to Genus.

	Sturt Meadows					
	Centrifuge	Focus				
Bin	Genus	Phylum	Class	Order	Family	Genus
Bin_12_2	*Acinetobacter*	Proteobacteria	Gammaproteobacteria	None	None	*None*
Bin_12_1	*Acinetobacter*	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*
Bin_12_3	*Acinetobacter*	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*
Bin_19_1	*Acinetobacter*	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*
Bin_19_3	*Acinetobacter*	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*
Bin_19_4	*Acinetobacter*	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*
Bin_21_1	*Acinetobacter*	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*
Bin_21_2	*Acinetobacter*	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*
Bin_26	*Acinetobacter*	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*
Bin_29_2	*Acinetobacter*	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*
Bin_31	*Acinetobacter*	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*
Bin_29_3	*Acinetobacter*	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	*Vibrio*
Bin_27	*Bacillus*	Firmicutes	Bacilli	Bacillales	Bacillaceae	*Bacillus*
Bin_19_2	*Bacillus*	Firmicutes	None	None	None	*None*
Bin_18_1	*Mycobacterium*	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	*Mycobacterium*
Bin_23	*Mycobacterium*	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	*Mycobacterium*
Bin_25	*Mycobacterium*	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	*Mycobacterium*
Bin_10	*None*	Bacteroidetes	None	None	None	*None*
Bin_16	*None*	Bacteroidetes	None	None	None	*None*
Bin_8	*None*	Bacteroidetes	None	None	None	*None*
Bin_1_1	*None*	None	None	None	None	*None*
Bin_1_2	*None*	None	None	None	None	*None*
Bin_1_3	*None*	None	None	None	None	*None*
Bin_1_4	*None*	None	None	None	None	*None*
Bin_11	*None*	None	None	None	None	*None*
Bin_13	*None*	None	None	None	None	*None*
Bin_14	*None*	None	None	None	None	*None*
Bin_15	*None*	None	None	None	None	*None*
Bin_2	*None*	None	None	None	None	*None*
Bin_24	*None*	None	None	None	None	*None*
Bin_29_1	*None*	None	None	None	None	*None*
Bin_3	*None*	None	None	None	None	*None*
Bin_30	*None*	None	None	None	None	*None*
Bin_32	*None*	None	None	None	None	*None*
Bin_6	*None*	None	None	None	None	*None*
Bin_28	*None*	Proteobacteria	Alphaproteobacteria	Rhizobiales	None	*None*
Bin_9	*None*	Proteobacteria	None	None	None	*None*
Bin_33	*Pseudomonas*	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*
Bin_5_1	*Pseudomonas*	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*

Bin_18_2	*Pseudonocardia*	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae	*Pseudonocardia*
Bin_22	*Serratia*	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	*Serratia*
Bin_20	*Shewanella*	Proteobacteria	Gammaproteobacteria	Alteromonadales	Shewanellaceae	*Shewanella*
Bin_5_3	*Stenotrophomonas*	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	*Stenotrophomonas*
Bin_5_4	*Stenotrophomonas*	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	*Stenotrophomonas*
Bin_5_5	*Stenotrophomonas*	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	*Stenotrophomonas*
Bin_17_1	*Unknown*	Proteobacteria	Alphaproteobacteria	None	None	*None*
Bin_17_2	*Unknown*	Proteobacteria	Alphaproteobacteria	None	None	*None*
Bin_17_3		Proteobacteria	Alphaproteobacteria	None	None	*None*

Table S4.5 Functional genes abundance at SEED subsystem level 1 for each of the different bacterial phyla found at at the Sturt Meadows and Laverton calcretes

	Laverton														
Subsystem Level 1	Achromobacter	Acidovorax	Acinetobacter	Alcanivorax	Dyadobacter	Methyloversatilis	Microbacterium	Mycobacterium	None	Ochrobactrum	Pannonibacter	Sphingopyxis	Spiroplasma	Unknown	Wolbachia
Carbohydrates	9.52	10.45	9.96	14.28	15.02	8.42	18.69	10.10	11.88	15.30	12.88	9.11	12.25	14.28	7.30
Amino Acids and Derivatives	10.28	12.18	9.25	3.93	3.75	8.71	10.42	11.74	8.31	8.53	10.09	7.45	1.52	11.05	6.43
Protein Metabolism	4.67	6.30	9.53	10.51	6.31	5.16	6.51	4.89	14.02	5.97	5.23	9.52	18.64	5.12	17.00
Cofactors, Vitamins, Prosthetic Groups, Pigments	6.23	7.26	4.71	7.59	5.04	5.40	5.17	13.55	7.47	5.38	4.87	4.77	7.07	5.81	5.96
Clustering-based subsystems	6.84	6.25	2.41	2.60	3.74	4.38	5.18	5.39	5.21	6.40	5.98	5.20	5.56	6.93	6.34
Nucleosides and Nucleotides	2.45	2.81	3.96	3.80	4.03	2.62	3.43	2.84	2.48	1.43	3.29	2.37	8.48	3.12	8.24
Respiration	3.34	3.75	3.73	3.44	4.23	6.78	3.51	2.07	4.67	3.11	3.35	4.82	1.52	4.06	5.53
Virulence	3.56	4.91	7.68	16.30	10.24	5.95	4.18	6.46	2.85	2.72	3.69	9.95	2.78	5.98	4.51
DNA Metabolism	2.59	4.37	6.45	3.35	7.48	4.49	4.43	3.05	3.79	6.52	3.38	7.03	21.72	3.77	8.69
Cell Wall and Capsule	5.02	3.72	4.00	2.61	3.63	5.19	3.77	2.06	3.09	5.57	4.78	4.97	0.00	3.59	4.71
Membrane Transport	7.36	5.89	8.73	4.58	19.21	6.72	6.62	2.51	3.55	5.00	6.18	8.96	0.30	6.37	1.18
Stress Response	4.76	3.87	5.39	2.67	2.74	8.29	4.05	3.73	4.54	5.92	7.07	3.38	0.61	4.33	3.98
Metabolism of Aromatic Compounds	2.95	2.63	1.08	0.04	2.48	1.14	1.76	2.66	0.45	1.86	1.98	1.30	0.00	2.00	0.00
Nitrogen Metabolism	1.81	2.66	1.55	5.17	2.24	3.41	0.73	1.38	0.69	2.16	2.95	1.92	0.00	1.70	1.54
Miscellaneous	3.91	2.86	1.04	0.12	0.03	2.42	2.73	2.55	5.07	5.00	2.43	2.89	4.92	3.71	1.97
Sulfur Metabolism	2.42	1.30	2.05	0.00	0.15	1.03	1.39	2.93	0.66	1.82	1.56	1.47	0.00	1.91	0.59
Phosphorus Metabolism	2.04	1.56	2.41	2.23	0.53	1.84	1.09	1.11	0.98	0.91	1.96	1.20	1.52	1.21	0.00
Fatty Acids, Lipids, and Isoprenoids	4.06	2.78	4.15	0.47	0.88	3.79	3.06	8.82	4.29	1.67	3.23	3.92	4.92	2.50	2.94
RNA Metabolism	3.52	5.52	2.57	4.64	0.59	5.09	4.11	2.70	8.53	3.86	2.32	2.45	2.65	3.58	5.10
Iron acquisition and metabolism	2.75	0.31	0.62	0.00	0.18	0.36	1.62	1.11	0.39	0.91	2.30	0.47	0.00	0.87	3.53
Motility and Chemotaxis	2.72	3.91	0.00	4.90	0.00	2.40	0.69	0.41	0.84	4.39	3.64	1.25	0.00	1.80	0.00
Phages, Prophages, Transposable elements, Plasmids	0.86	0.10	2.06	0.00	3.80	0.72	0.92	1.07	0.45	1.59	0.92	0.23	0.00	1.50	1.18

Potassium metabolism	1.26	1.55	0.67	2.03	0.08	1.20	0.00	1.33	0.68	0.68	1.38	1.60	0.00	0.51	0.00
Cell Division and Cell Cycle	0.67	0.68	0.01	1.96	1.06	0.48	0.31	0.49	0.39	0.27	0.17	0.62	2.53	0.61	1.01
Predictions based on plant-prokaryote comparative analysis	0.31	0.31	0.77	0.00	0.40	0.24	0.26	0.44	0.52	0.11	0.31	0.27	0.00	0.42	1.18
Central metabolism	0.08	0.00	0.06	0.00	0.00	0.07	0.21	0.33	0.21	0.00	0.27	0.07	0.00	0.30	0.00
Transcriptional regulation	0.08	0.00	0.27	0.00	1.59	0.12	0.00	0.00	0.01	0.15	0.17	0.47	0.00	0.08	0.52
Regulation and Cell signaling	3.80	1.95	2.51	2.80	0.58	2.36	2.60	3.21	3.03	2.31	2.98	2.20	1.52	2.49	0.59
Virulence, Disease and Defense	0.00	0.00	2.32	0.00	0.00	0.74	1.39	0.44	0.31	0.45	0.39	0.00	0.00	0.15	0.00
Secondary Metabolism	0.00	0.10	0.05	0.00	0.01	0.00	0.25	0.00	0.38	0.00	0.08	0.10	0.00	0.03	0.00
Photosynthesis	0.00	0.00	0.00	0.00	0.00	0.00	0.23	0.22	0.08	0.00	0.00	0.00	0.00	0.07	0.00
Dormancy and Sporulation	0.16	0.00	0.00	0.00	0.00	0.24	0.70	0.24	0.15	0.00	0.17	0.00	0.00	0.15	0.00
Plant cell walls and outer surfaces	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Phages, Prophages, Transposable elements	0.00	0.00	0.00	0.00	0.00	0.24	0.00	0.22	0.00	0.00	0.00	0.00	1.52	0.00	0.00

Table S4.5 Functional genes abundance at SEED subsystem level 1 from the bacteria at the Sturt Meadows and Laverton calcretes (cont.)

	Sturt									
Subsystem Level 1	Acinetobacter	Bacillus	Mycobacterium	None	Pseudomonas	Pseudonocardia	Serratia	Shewanella	Stenotrophomonas	Unknown
Carbohydrates	8.08	13.78	13.00	12.21	10.84	13.54	23.04	12.78	11.49	13.50
Amino Acids and Derivatives	11.71	10.81	12.31	9.93	8.39	13.32	5.07	7.42	8.52	11.63
Protein Metabolism	7.53	7.69	4.93	10.55	11.50	6.25	6.28	5.36	6.92	7.71
Cofactors, Vitamins, Prosthetic Groups, Pigments	5.28	6.34	11.09	6.85	3.93	10.13	6.01	4.76	5.94	6.64
Clustering-based subsystems	4.49	6.57	4.61	5.17	4.88	4.31	2.34	5.63	3.64	6.33
Nucleosides and Nucleotides	3.28	3.00	3.29	3.13	4.43	2.43	0.35	0.00	2.03	5.38
Respiration	2.87	2.69	3.96	5.32	2.35	4.96	8.86	6.46	2.28	4.98
Virulence	5.00	3.77	5.31	3.91	5.58	3.65	8.60	7.56	6.67	4.96
DNA Metabolism	5.93	5.01	3.49	3.94	6.52	3.62	0.00	7.51	5.60	4.93
Cell Wall and Capsule	5.42	5.20	2.39	4.14	4.89	2.84	1.81	1.03	4.02	4.06
Membrane Transport	7.84	3.49	3.39	4.28	6.33	3.64	0.00	3.58	11.09	3.99
Stress Response	4.90	2.89	3.83	3.61	5.77	3.89	3.42	5.64	4.79	3.32
Metabolism of Aromatic Compounds	1.83	0.50	2.33	0.56	0.41	3.01	4.82	0.00	0.48	2.42
Nitrogen Metabolism	1.92	1.19	3.05	0.75	3.32	2.48	16.37	9.36	1.11	2.31
Miscellaneous	1.66	3.39	3.20	4.33	1.35	3.78	0.00	0.00	2.94	2.22
Sulfur Metabolism	1.91	1.21	1.43	1.19	0.29	1.37	0.00	0.00	0.58	2.11
Phosphorus Metabolism	2.23	1.82	1.11	1.57	1.67	1.77	0.00	6.11	1.62	1.83
Fatty Acids, Lipids, and Isoprenoids	4.38	1.10	8.42	3.59	1.15	7.00	2.43	0.41	3.74	1.76
RNA Metabolism	3.96	3.76	2.30	5.60	4.82	2.49	1.06	7.54	4.01	1.66
Iron acquisition and metabolism	1.86	1.91	0.62	1.11	1.83	0.99	0.00	0.00	1.49	1.51
Motility and Chemotaxis	1.13	2.17	0.75	1.03	2.66	0.28	0.00	4.43	3.61	1.27
Phages, Prophages, Transposable elements, Plasmids	1.44	1.40	0.87	0.58	1.08	0.14	0.04	2.94	1.09	1.24
Potassium metabolism	0.99	1.02	0.80	1.04	2.73	0.84	2.31	0.25	1.68	1.17
Cell Division and Cell Cycle	0.76	1.60	0.26	0.17	0.27	0.54	0.00	0.00	0.33	1.11
Predictions based on plant-prokaryote comparative analysis	0.25	0.55	0.24	0.38	0.01	0.55	0.00	0.00	0.13	0.64

Central metabolism	0.29	0.08	0.13	0.26	0.05	0.24	0.00	0.00	0.08	0.43
Transcriptional regulation	0.05	0.09	0.02	0.19	0.08	0.00	0.00	0.00	0.10	0.37
Regulation and Cell signaling	2.20	2.01	2.25	3.19	2.79	1.50	7.11	1.22	3.21	0.31
Virulence, Disease and Defense	0.62	1.83	0.00	0.84	0.02	0.14	0.00	0.01	0.55	0.14
Secondary Metabolism	0.03	0.00	0.48	0.23	0.04	0.08	0.00	0.00	0.00	0.09
Photosynthesis	0.00	0.00	0.13	0.05	0.00	0.00	0.00	0.00	0.26	0.00
Dormancy and Sporulation	0.00	3.13	0.02	0.28	0.03	0.17	0.07	0.00	0.00	0.00
Plant cell walls and outer surfaces	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Phages, Prophages, Transposable elements	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table S4.6 Average SIMPER percent dissimilarity between the different beetle species to each other

	P. macrosturtensis	*P. mesosturtensis*	*P. microsturtensis*	*L. palmulaoides*	*L. windarraensis*	*L. lapostaae*
P. macrosturtensis						
P. mesosturtensis	78.54					
P. microsturtensis	74.33	56.85				
L. palmulaoides	99.01	100	100			
L. windarraensis	100	100	100	82.9		
L. lapostaae	100	100	100	71.87	29.9	

Statement of Authorship

Title of Paper	The first complete mitochondrial genomes of subterranean dytiscid diving beetles (Limbodessus and Paroster) from calcrete aquifers of Western Australia
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Published in the Australian Journal of Zoology

Principal Author

Name of Principal Author (Candidate)	Josephine Hyde
Contribution to the Paper	Collection of samples, molecular work, analysing the data, interpretation of the data, writing the manuscript
Overall percentage (%)	80%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a th ision in this thesis. I am the primary author of this paper.
Signature	Date 15/6/18

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Steven Cooper
Contribution to the Paper	Collection of samples, editing the manuscript
Signature	Date 15/6/18.

Name of Co-Author	Pablo Munguia
Contribution to the Paper	Editing the manuscript
Signature	Date 5 - Jun 2018

Name of Co-Author	William Humphreys

Contribution to the Paper	Collection of samples, editing the manuscript
Signature	Date 5/6/2018

Name of Co-Author	Andrew Austin
Contribution to the Paper	Editing the manuscript
Signature	Date 15/6/18

Please cut and paste additional co-author panels here as required.

Chapter 5: The first complete mitochondrial genomes of subterranean dytiscid diving beetles (*Limbodessus* and *Paroster*) from calcrete aquifers of Western Australia

Josephine Hyde[A,F], *Steven J. B. Cooper*[A,B], *Pablo Munguia*[C], *William F. Humphreys*[D,E] *and Andrew D. Austin*[A]

[A]Australian Centre for Evolutionary Biology and Biodiversity, School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia.

[B]Evolutionary Biology Unit, South Australian Museum, North Terrace, Adelaide, SA 5000, Australia.

[C]RMIT Studios, RMIT University, Melbourne, Vic. 3000, Australia.

[D]Western Australian Museum, Welshpool DC, WA 6986, Australia.

[E]School of Animal Biology, University of Western Australia, Crawley, WA 6009, Australia.

[F]Corresponding author. Email: josephine.hyde@adelaide.edu.au

Abstract

Comparative analyses of mitochondrial (mt) genomes may provide insights into the genetic changes, associated with metabolism, that occur when surface species adapt to living in underground habitats. Such analyses require comparisons among multiple independently evolved subterranean species, with the dytiscid beetle fauna from the calcrete archipelago of central Western Australia providing an outstanding model system to do this. Here, we present the first whole mt genomes from four subterranean dytiscid beetle species of the genera *Limbodessus* (*L. palmulaoides*) and *Paroster* (*P. macrosturtensis*, *P. mesosturtensis* and *P. microsturtensis*) and compare genome sequences with those from surface dytiscid species. The mt genomes were sequenced using a next-generation sequencing approach employing the Illumina Miseq system and assembled *de novo*. All four mt genomes are circular, ranging in size from 16 504 to 16 868 bp, and encode 37 genes and a control region. The overall structure (gene number, orientation and order) of the mt genomes is the same as that found in eight sequenced surface species, but with genome size variation resulting from length variation of intergenic regions and the control region. Our results provide a basis for future investigations of adaptive evolutionary changes that may occur in mt genes when species move underground.

5.1 Introduction

Stygobionts, invertebrates that complete their life cycle entirely in subterranean groundwater habitats, are highly adapted to their environment (Galassi 2001). These habitats are characterised by darkness, a reduction in primary production, a lack of oxygen and, often, fragmented microhabitats (Hüppop 2000; Culver *et al*. 2005). Stygobionts are characteristically blind and depigmented, and show adaptations to living underground that include elongated appendages, reduced metabolism and reproduction rates, loss of wings, as well as an extended life span (Jeffery 2001; Di Lorenzo *et al*. 2015). Understanding the evolution of mitochondrial (mt) genomes of these animals could provide insights into a key part of their metabolic processes. However, there are currently very few mt genomes sequenced from subterranean animals, making such evolutionary analyses difficult to perform.

The arid Yilgarn region in central Western Australia is a biodiversity hotspot for subterranean invertebrates (Guzik *et al*. 2011). The region has hundreds of isolated calcrete (carbonate) bodies, with those examined having their own array of endemic stygofaunal species. Of particular note are the aquatic diving beetles (Dytiscidae), of which ~100 stygobitic species have been described from two tribes, Bidessini and Hydroporini (Balke and Ribera 2004; Leys and Watts 2008; Watts and Humphreys 2009). In each calcrete with stygobitic beetles present, there are generally between two and four species from non-overlapping size classes. Several these are sympatric sister species, suggesting that they have evolved from a stygobitic ancestral species within the calcrete (Cooper *et al*. 2002; Leys *et al*. 2003; Leijs *et al*. 2012). However, most species have evolved independently from surface ancestors, providing a powerful system for exploring the adaptive and regressive changes that occur during the evolution of subterranean animals.

Advances in DNA sequencing technology in recent years now make it possible to obtain whole mt genomes of diverse animal groups. Most of the genomes that have been published to date have relied on long-range PCR, which can be both challenging and time-consuming (Hahn *et al*. 2013) because it requires high-molecular-weight DNA, and available primers may not work on the target animal group. With next-generation sequencing of total genomic DNA, some of these issues can be resolved. Due to the small size of the mt genome and high copy number, only relatively shallow

sequencing is required to reconstruct the complete genome (Cameron 2014; Kocher *et al*. 2014; Linard *et al*. 2016).

In this study, we present four new mt genomes from subterranean diving beetle species from the genera *Limbodessus* and *Paroster*. Additionally, we compare the overall structure (gene content, order, orientation and size) of these genomes with those of eight epigean (surface) dytiscid species.

5.2 Materials and methods

5.2.1 Specimen collection

The study sites included two calcretes found in the Yilgarn region of Western Australia, at Laverton Downs and Sturt Meadows pastoral stations. Thirteen specimens from four species were sequenced in this study (collection details are listed in Table 5.1). Species included *Limbodessus palmulaoides*, the largest beetle species found in the Laverton calcrete (Watts and Humphreys 2009), and *Paroster macrosturtensis*, *P. mesosturtensis* and *P. microsturtensis*, three sympatric sister species from the Sturt Meadows calcrete (Guzik *et al*. 2009; Watts and Humphreys 2009). Adult beetles were identified on the basis of morphological characters, and larval beetles by COI barcoding and BLAST comparison with COI data from GenBank. All specimens, except two, were preserved by snap freezing in liquid nitrogen and stored at –80°C. An additional specimen (25545) was preserved in 100% ethanol and then stored at –20°C, while the other (25542) was killed with 100% ethanol just prior to DNA extraction.

Table 5.1 Collection information for the beetles sequenced in this study. All collection localities are in Western Australia. Collectors: KKJ, K. K. Jones; SJBC, S. J. B. Cooper; BL, B. Langille; WFH, W. F. Humphreys; JH, J. Hyde; AA, A. Allford

Genus	Species	Life stage	Location	Collection date	Collector	BPA catalogue no.
Limbodessus	*palmulaoides*	Adult	Laverton	Sep. 2015	KKJ, SJBC, BL	102.100.100/25542
Limbodessus	*palmulaoides*	Adult	Laverton	Apr. 2015	WFH, SJBC, JH	102.100.100/27821
Limbodessus	*palmulaoides*	Larva	Laverton	Apr. 2015	WFH, SJBC, JH	102.100.100/28086
Limbodessus	*palmulaoides*	Larva	Laverton	Apr. 2015	WFH, SJBC, JH	102.100.100/28088
Paroster	*macrosturtensis*	Adult	Sturt Meadows	Sep. 2015	KKJ, SJBC, BL	102.100.100/25544
Paroster	*macrosturtensis*	Adult	Sturt Meadows	Nov. 2006	WFH, SJBC, AA	102.100.100/25545
Paroster	*macrosturtensis*	Adult	Sturt Meadows	Apr. 2015	WFH, SJBC, JH	102.100.100/27822
Paroster	*macrosturtensis*	Larva	Sturt Meadows	Apr. 2015	WFH, SJBC, JH	102.100.100/28084
Paroster	*macrosturtensis*	Larva	Sturt Meadows	Apr. 2015	WFH, SJBC, JH	102.100.100/28085
Paroster	*macrosturtensis*	Larva	Sturt Meadows	Apr. 2015	WFH, SJBC, JH	102.100.100/28089
Paroster	*mesosturtensis*	Adult	Sturt Meadows	Apr. 2015	WFH, SJBC, JH	102.100.100/27823
Paroster	*microsturtensis*	Larva	Sturt Meadows	Apr. 2015	WFH, SJBC, JH	102.100.100/25543
Paroster	*microsturtensis*	Adult	Sturt Meadows	Apr. 2015	WFH, SJBC, JH	102.100.100/27824

5.2.2 DNA extraction and sequencing

Genomic DNA was extracted from whole specimens using a modified Gentra Pure-Gene DNA purification kit protocol (Gentra Systems, Minneapolis, MN, USA). All extractions were performed inside an Aura PCR cabinet (EuroClone, Pero, Italy). Genomic DNA was sent for library construction, and sequencing at the Australian Genome Research Facility; libraries were prepared using a Nextera DNA library prep kit (Caruccio 2011). Each library contained a single specimen and three Illumina Miseq runs were performed. The first Miseq run (300 bp paired end sequencing) included five libraries, four of which were included in the current study. A second Miseq run (150 bp paired end sequencing) included six libraries, four of which were included in the current study. A final Miseq run (300 bp paired end sequencing) had five libraries that were all included in the current study.

5.2.3 Analysis and annotation

Raw sequences were initially analysed to filter out low-quality sequences and those that contained Ns using Trimmomatic (Bolger *et al*. 2014) and Prinseq (Schmieder and Edwards 2011). A reference database of all complete beetle mt genomes on GenBank (n = 172, March 2017) was assembled. The dytiscid sequence data were mapped to the reference file using bowtie2, and default parameter settings (Langmead and Salzberg 2012), to separate the mt genome sequences from the rest of the sequence data. The mt genomes were assembled using a combination of MIRA4 (using the default settings for mirabait) (Chevreux *et al*. 1999) and the circular genome assembly tool in Geneious 8.1.9 (Kearse *et al*. 2012). The De Novo assembly tool used in Geneious was set to custom sensitivity with minimum overlap set to 100 bp, minimum overlap ID 95%, word length 50 bp and maximum mismatch set to 5%. The mt genomes were annotated using MITOS (Bernt *et al*. 2013), applying the invertebrate mt genetic code. Any tRNA genes (tRNAs) not found using MITOS were checked against the beetle mt genomes from Linard *et al*. (2016). The protein-coding genes (PCGs) and rRNA genes (rRNAs) were verified by using Blast+ searches (Camacho *et al*. 2009) and then the 5’ and 3’ ends of the genes were refined by comparing the sequences against beetles in the suborder Adephaga (available on GenBank, March 2017) (Table 5.2) and refining the regions by eye.

5.2.4 Sequence analysis

For comparison to published genomes, eight complete mtDNA genomes from surface dytiscid species were downloaded from GenBank (21/02/2017) (Table 5.2). The AT Skew of each of the four completed genomes, and those for the additional eight dytiscids were calculated using (A–T)/(A+T) and the GC skew was calculated using (G–C)/(G+C); both skew calculations were based on the majority strand sequence (Grigoriev 1998). AT skews were calculated for the whole genome, the PCGs, tRNAs, rRNAs and the control region (CR). AT percentage ratios were also calculated for the coding region, the CR and the rRNAs. The relative synonymous codon usage was also calculated using MEGA6 (Tamura *et al*. 2013).

Table 5.2 List of complete mt genomes of species in the suborder Adephaga used in this study and their genome sizes

Species	Family	Size (bp)	Accession no.	Reference
Trachypachus holmbergi	Trachypachidae	15 722	NC_011329	Sheffield *et al*. (2008)
Aspidytes niobe	Aspidytidae	14 257	NC_012139	Pons *et al*. (2010)
Macrogyrus oblongus	Gyrinidae	16 643	NC_013249	Cameron *et al*. (2009)
Damaster mirabilissimus mirabilissimus	Carabidae	16 823	NC_016469	Wan *et al*. (2012)
Calosoma sp. BYU-CO241	Carabidae	16 462	NC_018339	Song *et al*. (2010)
Abax parallelepipedus	Carabidae	17 701	NC_030592	Linard *et al*. (2016)
Hygrobia hermanni	Hygrobiidae	16 336	NC_030593	Linard *et al*. (2016)
Acilius sp.	Dytiscidae	20 689	KT876878	Linard *et al*. (2016)
Hygrotus sp.	Dytiscidae	16 730	KT876899	Linard *et al*. (2016)
Hygrotus sp.	Dytiscidae	17 968	KT876900	Linard *et al*. (2016)
Hygrotus sp.	Dytiscidae	17 071	KT876901	Linard *et al*. (2016)
Colymbetes sp.	Dytiscidae	16 211	KT876885	Linard *et al*. (2016)
Liopterus sp.	Dytiscidae	16 541	KT876902	Linard *et al*. (2016)
Hydroporus sp.	Dytiscidae	23 380	KT876896	Linard *et al*. (2016)
Hydroporus sp.	Dytiscidae	17 698	KT876897	Linard *et al*. (2016)
Limbodessus palmulaoides	Dytiscidae	16 868	MG912994	This study
Paroster macrosturtensis	Dytiscidae	16 676	MG912995	This study
Paroster mesosturtensis	Dytiscidae	16 663	MG912996	This study
Paroster microsturtensis	Dytiscidae	16 504	MG912997	This study

5.3 Results and discussion

5.3.1 Genome organisation

Complete mt genomes were obtained for *L. palmulaoides* (16 868 bp), *P. macrosturtensis* (16 676 bp) (Fig. 5.1), *P. mesosturtensis* (16 663 bp), and *P. microsturtensis* (16 504 bp) (Table 5.3). The mean coverage for all the combined mt genomes by the raw sequence data was between 66.8× (*P. mesosturtensis*) and 2612.5×

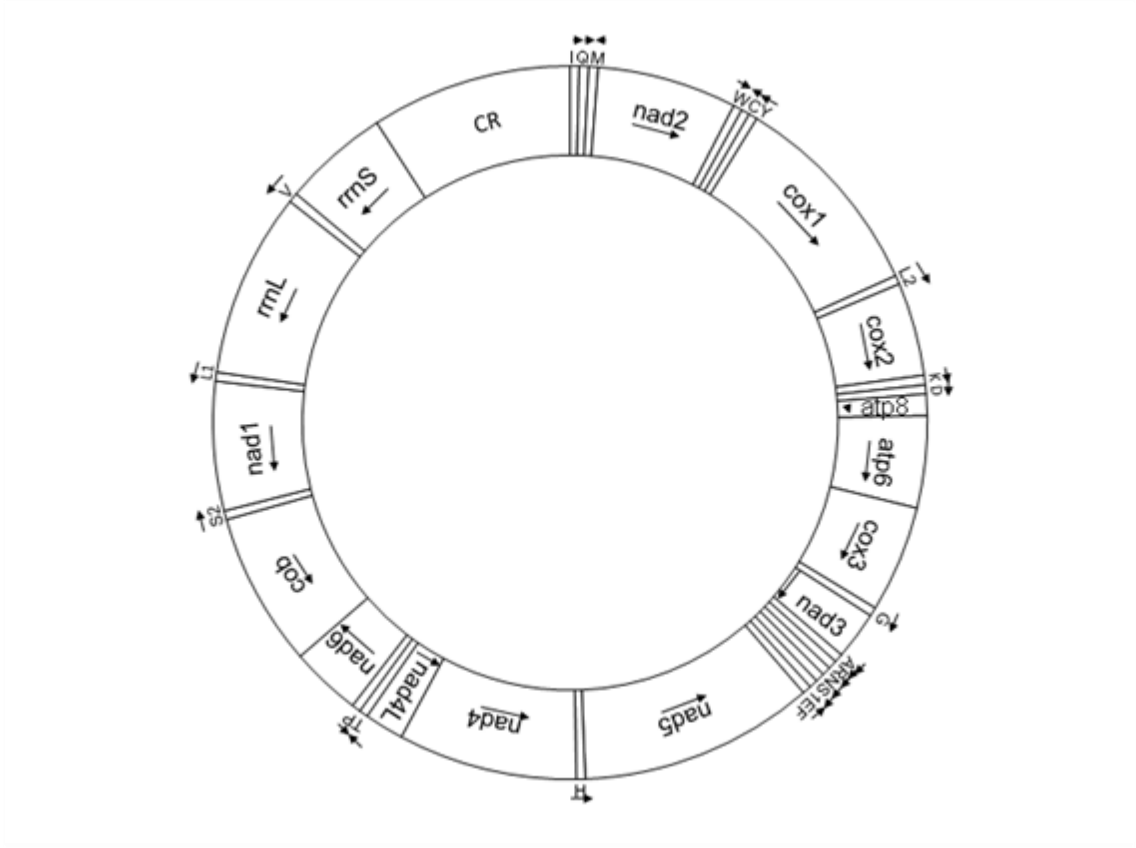

Fig. 5.1 Map of the mt genome of *Paroster macrosturtensis*. The tRNAs are labelled according to IUPAC-IUB. One-letter symbols – S1, S2, L1, L2 – denote the codons *tRNA-Ser*(AGN), *tRNA-Ser* (UCN), *tRNA-Leu*(CUN) and *tRNA-Leu*(UUR), respectively. The arrow direction indicates if the gene is on the majority or minority strand. An identical mt genome structure was found for all the dytiscid species that were sequenced in the current study.

(*L. palmulaoides*) and every nucleotide was covered a minimum of 2× (*P. microsturtensis*) to 23× (*P. macrosturtensis*) (Table S5.1, Fig. S5.1, Supplementary Material). None of the sequences had a Q30 score below 81.5 and the mean confidence for the genomes was 1116.61 (Table S5.2, Supplementary Material).

Previous studies have suggested that the coding region of the coleopteran mt genome is reasonably stable at an average of 14 700 bp in length (Sheffield *et al*. 2008). However, the mean coding region length for the four new mt genomes is 15 370.8 bp (with a standard deviation of 284.2 bp) and this only decreases to a mean of 15 145.5 bp (with a standard deviation of 334.1 bp) when the eight other genomes from the surface species are included (Linard *et al*. 2016). There is little variation in length among the three *Paroster* sister species with a standard deviation between the three genomes of 5.6 bp compared with the two epigean *Hydroporus spp*., which had a standard deviation of 81.5 bp and the three epigean *Hygrotus sp*., which had a standard deviation of 220.3 bp.

Table 5.3 The annotation and gene organisation of the four mt genomes sequenced

Gene	Direction	Location	Length	Anticodon	Codon start	Codon stop	Intergenic
L. palmulaoides							
Transfer RNA-Ile	F	1–68	68	gat	–	–	0
Transfer RNA-Gln	R	1170–1238	69	ttg	–	–	1101
Transfer RNA-Met	F	1238–1306	69	cat	–	–	–1
nad2	F	1307–2332	1026	–	ATT	TAA	0
Transfer RNA-Trp	F	2333–2396	64	tca	–	–	0
Transfer RNA-Cys	R	2428–2492	65	gca	–	–	31
Transfer RNA-Tyr	R	2493–2557	65	gta	–	–	0
cox1	F	2550–4094	1545	–	ATT	TAA	–8
Transfer RNA-Leu(UUR)	F	4090–4154	65	taa	–	–	–5
cox2	F	4155–4842	688	–	ATG	T	0
Transfer RNA-Lys	F	4843–4913	71	ctt	–	–	0
Transfer RNA-Asp	F	4914–4979	66	gtc	–	–	0
atp8	F	4980–5138	159	–	ATT	TAA	0
atp6	F	5132–5806	675	–	ATG	TAA	–7
cox3	F	5806–6594	789	–	ATG	TAA	–1
Transfer RNA-Gly	F	6594–6659	66	tcc	–	–	–1
nad3	F	6660–7025	366	–	ATT	TAA	0
Transfer RNA-Ala	F	7012–7075	64	tgc	–	–	–14
Transfer RNA-Arg	F	7076–7139	64	tcg	–	–	0
Transfer RNA-Asn	F	7137–7203	67	gtt	–	–	–3
Transfer RNA-Ser(AGN)	F	7204–7270	67	gct	–	–	0
Transfer RNA-Glu	F	7271–7336	66	ttc	–	–	0
Transfer RNA-Phe	R	7335–7401	67	gaa	–	–	–2
nad5	R	7401–9133	1733	–	ATT	TA	–1
Transfer RNA-His	R	9131–9195	65	gtg	–	–	0
nad4	R	9196–10516	1321	–	ATA	T	–1
nad4l	R	10528–10818	291	–	ATT	TAA	11
Transfer RNA-Thr	F	10821–10886	66	tgt	–	–	2
Transfer RNA-Pro	R	10887–10953	67	tgg	–	–	0
nad6	F	10956–11471	516	–	ATT	TAA	2
cob	F	11471–12607	1137	–	ATG	TAG	–1
Transfer RNA-Ser(UCN)	F	12606–12670	65	tga	–	–	–2
nad1	R	12687–13637	954	–	TTG	TAG	16
Transfer RNA-Leu(CUN)	R	13638–13700	63	tag	–	–	0
rrnL	R	13666–15003	1338	–	–	–	–13
Transfer RNA-Val	R	15002–15072	71	tac	–	–	–1
rrnS	R	15072–15859	788		–	–	–1
CR	–	15859–16868	1009	–	–	–	–
P. macrosturtensis							
Transfer RNA-Ile	F	1–66	66	gat	–	–	0
Transfer RNA-Gln	R	485–553	69	ttg	–	–	418
Transfer RNA-Met	F	561–629	69	cat	–	–	–1

nad2	F	630–1658	1029	–	ATT	TAA	0
Transfer RNA-Trp	F	1658–1723	66	tca	–	–	–1
Transfer RNA-Cys	R	1748–1809	62	gca	–	–	24
Transfer RNA-Tyr	R	1810–1873	64	gta	–	–	0
cox1	F	1866–3410	1545	–	ATT	TAA	–8
Transfer RNA-Leu(UUR)	F	3406–3471	66	taa	–	–	–5
cox2	F	3473–4160	688	–	ATG	T	1
Transfer RNA-Lys	F	4161–4231	71	ctt	–	–	0
Transfer RNA-Asp	F	4232–4297	66	gtc	–	–	0
atp8	F	4298–4456	159	–	ATG	TAG	0
atp6	F	4450–5127	678	–	ATG	TAA	–7
cox3	F	5127–5915	789	–	ATG	TAA	–1
Transfer RNA-Gly	F	5915–5980	66	tcc	–	–	–1
nad3	F	5981–6334	354	–	ATC	TAG	0
Transfer RNA-Ala	F	6333–6397	65	tgc	–	–	–2
Transfer RNA-Arg	F	6398–6462	65	tcg	–	–	0
Transfer RNA-Asn	F	6463–6526	64	gtt	–	–	0
Transfer RNA-Ser(AGN)	F	6527–6593	67	gct	–	–	0
Transfer RNA-Glu	F	6594–6657	64	ttc	–	–	0
Transfer RNA-Phe	R	6656–6720	65	gaa	–	–	–2
nad5	R	6720–8453	1734	–	ATT	TAA	–1
Transfer RNA-His	R	8451–8516	66	gtg	–	–	0
nad4	R	8517–9850	1334	–	ATA	T	–1
nad4l	R	9849–10139	291	–	ATT	TAA	11
Transfer RNA-Thr	F	10142–10206	65	tgt	–	–	2
Transfer RNA-Pro	R	10207–10272	66	tgg	–	–	0
nad6	F	10274–10795	522	–	ATC	TAA	1
Cob	F	10795–11931	1137	–	ATG	TAG	–1
Transfer RNA-Ser(UCN)	F	11930–11995	66	tga	–	–	–2
nad1	R	12012–12959	1184	–	TTG	TAG	16
Transfer RNA-Leu(CUN)	R	12960–13024	65	tag	–	–	0
rrnL	R	12990–14340	1351	–	–	–	–13
Transfer RNA-Val	R	14339–14409	71	tac	–	–	–1
rrnS	R	14411–15193	783		–	–	–1
CR	–	15194–16676	1482	–	–	–	–
P. mesosturtensis							
Transfer RNA-Ile	F	1–65	65	gat	–	–	0
Transfer RNA-Gln	R	482–550	69	ttg	–	–	416
Transfer RNA-Met	F	558–626	69	cat	–	–	7
nad2	F	627–1655	1029	–	ATT	TAA	0
Transfer RNA-Trp	F	1655–1721	67	tca	–	–	–1
Transfer RNA-Cys	R	1745–1806	62	gca	–	–	23
Transfer RNA-Tyr	R	1807–1871	65	gta	–	–	0
cox1	F	1864–3408	1545	–	ATT	TAA	–8
Transfer RNA-Leu(UUR)	F	3404–3469	66	taa	–	–	–5
cox2	F	3472–4159	688	–	ATG	T	2

Transfer RNA-Lys	F	4160–4230	71	ctt	–	–	0
Transfer RNA-Asp	F	4231–4295	65	gtc	–	–	0
atp8	F	4296–4454	159	–	ATG	TAA	0
atp6	F	4448–5125	678	–	ATG	TAA	–7
cox3	F	5125–5913	789	–	ATG	TAA	–1
Transfer RNA-Gly	F	5913–5978	66	tcc	–	–	–1
nad3	F	5979–6332	354	–	ATC	TAG	0
Transfer RNA-Ala	F	6331–6395	65	tgc	–	–	–2
Transfer RNA-Arg	F	6395–6458	64	tcg	–	–	–1
Transfer RNA-Asn	F	6459–6522	64	gtt	–	–	0
Transfer RNA-Ser(AGN)	F	6523–6589	67	gct	–	–	0
Transfer RNA-Glu	F	6590–6653	64	ttc	–	–	0
Transfer RNA-Phe	R	6652–6716	65	gaa	–	–	–2
nad5	R	6716–8449	1734	–	ATT	TAA	–1
Transfer RNA-His	R	8447–8511	65	gtg	–	–	0
nad4	R	8512–9832	1334	–	ATG	T	–1
nad4l	R	9844–10134	291	–	ATT	TAA	11
Transfer RNA-Thr	F	10137–10201	65	tgt	–	–	2
Transfer RNA-Pro	R	10202–10269	68	tgg	–	–	0
nad6	F	10271–10792	522	–	ATT	TAA	1
cob	F	10792–11928	1137	–	ATG	TAG	–1
Transfer RNA-Ser(UCN)	F	11927–11993	67	tga	–	–	–2
nad1	R	12010–12957	948	–	TTG	TAG	16
Transfer RNA-Leu(CUN)	R	12958–13022	65	tag	–	–	0
rrnL	R	12988–14333	1346	–	–	–	–13
Transfer RNA-Val	R	14332–14402	71	tac	–	–	1
rrnS	R	14404–15183	780		–	–	1
CR	–	15184–16663	1479	–	–	–	–
P. microsturtensis							
Transfer RNA-Ile	F	1–65	65	gat	–	–	0
Transfer RNA-Gln	R	307–375	69	ttg	–	–	241
Transfer RNA-Met	F	383–451	69	cat	–	–	7
nad2	F	452–1480	1029	–	ATT	TAA	0
Transfer RNA-Trp	F	1480–1547	68	tca	–	–	–1
Transfer RNA-Cys	R	1570–1631	62	gca	–	–	22
Transfer RNA-Tyr	R	1632–1698	67	gta	–	–	0
cox1	F	1691–3235	1545	–	ATT	TAA	–8
Transfer RNA-Leu(UUR)	F	3231–3296	66	taa	–	–	–5
cox2	F	3299–3986	688	–	ATG	T	2
Transfer RNA-Lys	F	3987–4057	71	ctt	–	–	0
Transfer RNA-Asp	F	4058–4123	66	gtc	–	–	0
atp8	F	4124–4282	159	–	ATG	TAA	0
atp6	F	4276–4953	678	–	ATG	TAA	–7
cox3	F	4953–5741	789	–	ATG	TAA	–1
Transfer RNA-Gly	F	5741–5806	66	tcc	–	–	–1
nad3	F	5807–6172	366	–	ATC	TAG	0

Transfer RNA-Ala	F	6159–6223	65	tgc	–	–	–14
Transfer RNA-Arg	F	6224–6289	66	tcg	–	–	0
Transfer RNA-Asn	F	6290–6353	64	gtt	–	–	0
Transfer RNA-Ser(AGN)	F	6354–6420	67	gct	–	–	0
Transfer RNA-Glu	F	6421–6484	64	ttc	–	–	0
Transfer RNA-Phe	R	6483–6547	65	gaa	–	–	–2
nad5	R	6547–8280	1734	–	ATT	TAA	–1
Transfer RNA-His	R	8278–8343	66	gtg	–	–	0
nad4	R	8344–9677	1334	–	ATA	T	–1
nad4l	R	9676–9966	291	–	ATT	TAA	11
Transfer RNA-Thr	F	9969–10033	65	tgt	–	–	2
Transfer RNA-Pro	R	10034–10099	66	tgg	–	–	0
nad6	F	10101–10622	522	–	ATT	TAA	1
cob	F	10622–11758	1137	–	ATG	TAG	–1
Transfer RNA-Ser(UCN)	F	11757–11822	66	tga	–	–	–2
nad1	R	11839–12786	948	–	TTG	TAG	16
Transfer RNA-Leu(CUN)	R	12787–12850	64	tag	–	–	0
rrnL	R	12817–14168	1352	–	–	–	–12
Transfer RNA-Val	R	14167–14237	71	tac	–	–	1
rrnS	R	14239–15018	780		–	–	1
CR	–	15019–16504	1485	–	–	–	–

The gene content in the genomes of all four subterranean species are typical of insect genomes previously reported, with 13 PCGs, 22 tRNA genes, two mt rRNA genes, and a single CR. The orientation and the order of the genes in *Limbodessus* and *Paroster* are identical to those of other beetle mt genomes previously reported and to the ancestral insect mt genome (Boore *et al*. 1998; Hwang *et al*. 2001). This lack of alteration to the gene order and the absence of any additional genes suggest that the larger sizes of the dytiscid mt genomes are likely due to an increase in the length or number of intergenic regions (IGRs) or the length of the PCGs, compared with other beetle families. When comparing the length of coding regions and the 37 mt genes from the family Dytiscidae to those in beetles from other families (Sheffield *et al*. 2008), there is evidence that the increased length may result from an increase in length of both IGRs and PCGs. Of particular note is the IGR between the tRNAs tRNA-Ile and tRNA-Gln genes, which is considerably expanded in the dytiscid beetles and is likely to account for most of the difference in length. These features are further explored in comparative analyses given below.

5.3.2 Overlapping genes, intergenic regions, and skewness

The evolution of the mt genome favours a reduction in size (Andersson and Kurland 1998) and, from an evolutionary perspective, it makes sense that there would be a reduction in the number of IGRs, potentially even to the point of gene overlap. However, gene overlap appears to be the exception rather than the rule as it is rarely the case that the end of one gene is a useful part of the next, plus overlapping genes can lead to post-transcriptional complications (Burger *et al*. 2003; Sheffield *et al*. 2008). In all four genomes sequenced here, several PCGs (cox1, nad3, and cob) overlap tRNA genes (Table 5.3). Additionally, in all three *Paroster* species *nad2* overlaps by 1 bp with the tRNA-trp gene.

The mt genome of *L. palmulaoides* contains 1163 bp of IGRs spread over six regions, ranging from 1101 to 2 bp; the longest region is between tRNAs tRNA-Ile and tRNA-Gln. The mt genomes of the three *Paroster* species contain IGRs of 482 bp (*P. macrosturtensis*), 480 bp (*P. mesosturtensis*), and 304 bp (*P. microsturtensis*). The longest region (418, 416, 241 bp) in each species (respectively) occurs between tRNAs tRNA-Ile and tRNA-Gln. In all three species, the shortest IGR is 1 bp. Except for KT876896 (*Hydroporus* sp.) and KT876902 (*Liopterus* sp.), all other dytiscid mt genomes sequenced have their longest region between tRNAs tRNA-Ile and tRNA-Gln and range between 1563 and 100 bp (Linard *et al*. 2016). This region contained no tandem repeats and did not return any significant blast results. Additionally, it did not fold like tRNAs or contain any open reading frames, suggesting that the region is non-coding and non-functional. While most IGRs are unique to each species, there is one well known IGR common to Coleoptera and other insect orders. It is a small IGR between tRNA-Ser(UCN) and nad1, with a 5-bp conserved region (TACTA) (Cameron and Whiting 2008; Sheffield *et al*. 2008). All 12 sequenced dytiscids have a 16-bp conserved region including a 5-bp TACTA motif between tRNA-Ser(UCN) and *nad1* with only two species having a single T → A point mutation.

The overall AT content of the genomes of the four subterranean beetles ranged from 75.8% to 77.0% (Table S5.3, Supplementary Material), which is within the range of the eight surface species' genomes (75.1–81.2%). It is also within the ranges that have been reported previously for other Coleoptera (65.6–78.2%) (Sheffield *et al*. 2008). Three of the previously sequenced mt genomes (KT876878 *Acilius* sp., KT876896 *Hydroporus* sp., and KT876897 *Hydroporus* sp.) have overall AT contents

that are above the range previously reported for Coleoptera and all the dytiscid beetles sequenced are at the top end of the range. The four new genomes, like the eight previously reported, have a weak positive AT skew and a negative GC skew (Table S5.3, Supplementary Material).

5.3.3 Protein-coding genes

The 13 PCGs in all four mt genomes presented use standard start codons and both complete and incomplete stop codons with the exception of *nad1*, which uses the atypical start codon TTG (Table 5.3). They also have incomplete stop codons to terminate cox2 (T) and nad4 (T). Additionally, *L. palmulaoides* has an incomplete stop codon at the end of nad5 (TA). Other beetle mt genomes, including those of dytiscids, have been found to include both the atypical start codon and incomplete stop codons (Sheffield *et al*. 2008; Linard *et al*. 2016). Relatively synonymous codon usage values of each of the four mt genomes are summarised in Table S5.4 (Supplementary Material); in all three of the *Paroster* genomes, all of the codons are present. In L. palmulaoides only the codons CGC and CGG, which both code for Arg, are not represented in the coding sequence. The most frequent amino acids are leucine 2 (Leu(UUR) amino acid present 401–433 times), isoleucine (Ile amino acid present 379–393 times) and phenylalanine (Phe amino acid present 369–385 times); these amino acids are also abundant in the other dytiscid beetles as well as other insects (Sheffield *et al*. 2008; Dai *et al*. 2017). These frequencies are consistent with the range observed in the surface dytiscid beetles sequenced (Fig. S5.3, Supplementary Material). The average AT content of the 13 PCGs for *L. palmulaoides* is 73.91%, and in the three *Paroster* species it ranges from 72.30% (*P. macrosturtensis*) to 74.79 (*P. mesosturtensis*). The AT skew in PCGs of all four subterranean dytiscids was slightly negative, indicating a higher content of T than of A (Table S5.1, Supplementary Material). The GC skew in *P. macrosturtensis* and *P. mesosturtensis* was also slightly negative, showing a higher content of C than G present in the PCGs. Similar results were found in the other epigean dytiscids investigated. However, the GC skew of both *P. microsturtensis* and *L. palmulaoides* is zero, indicating that an equal number of Gs and Cs are present in the 13 PCGs overall (Table S5.3, Supplementary Material).

5.3.4 Transfer RNAs

It was found that for the structure of the tRNA genes in the subterranean species, 14 are encoded on the major strand, and the remaining eight are encoded on the minor

strand. The total length of the tRNAs of *L. palmulaoides* is 1460 bps, while the three *Paroster* species total tRNA lengths range from 1454 bp (*P. macrosturtensis*) to 1458 bp (*P. microsturtensis*). The maximum tRNA length of all subterranean species is 71 bp. The minimum tRNA length is 62 bp in three species while in *P. mesosturtensis* the shortest tRNA is 63 bp. The tRNAs AT content for *L. palmulaoides* is 77.67%, and for the three *Paroster* species it is between 79.00% (*P. microsturtensis*) and 79.59% (*P. mesosturtensis*). All four AT (0.01–0.04) and GC (0.15–0.19) skews were slightly positive, indicating more A and G occurrences compared with T and C nucleotides.

5.3.5 Ribosomal RNAs

The two rRNA genes (*rrnS* and *rrnL*) in all four mt genomes investigated are located between tRNA-Leu(CUN) and tRNA-Val, and tRNA-Val and the CR, respectively. The lengths of *rrnL* range from 1315 bp (*L. palmulaoides*) to 1327 bp (*P. microsturtensis*) and the lengths of *rrnS* range from 780 bp (*P. mesosturtensis*, *P. microsturtensis*) to 788 bp (*L. palmulaoides*). The AT content of the two ribosomal genes is very similar, ranging between 79.25% (*L. palmulaoides*) and 80.21% (*P. microsturtensis*). The rRNA AT skew for all four species was slightly negative (–0.04 to –0.05), indicating that there were more T nucleotides than A. The GC skew was positive (0.38 to 0.40), indicating that there were more G nucleotides than C, as found in other beetle mt genomes (Friedrich and Muqim 2003; Sheffield *et al*. 2008).

5.3.6 Control region

While coding regions are, to a large degree, constrained in their length, for the genes to function properly the AT-rich CR has considerable length variation as it is the non-coding region of the mt genome and so is relatively free from these restraints (Fenn *et al*. 2007). While the size of CRs varies significantly across different beetle lineages, from less than 300 bp to over 6500 bp (Sheffield *et al*. 2008), in the Dytiscidae, especially, there appears to be a significant amount of variation in the length of the CR. It ranges in size from 1009 bp in L. palmulaoides to 8648 bp in one of the *Hydroporus* sp. (Linard *et al*. 2016), the latter being larger than any other coleopteran CR previously reported (Sheffield *et al*. 2008). The AT content of the CRs of the four mt genomes are remarkably consistent, ranging from 82.4% (*P. microsturtensis*) to 84.9% (*P. macrosturtensis*), considering the highly variable nature of the region. The AT content of surface dytiscid beetle CRs range from 84.9% (KT876899) to 94% (KT876897) (Table 5.4).

5.4 Conclusion

The mt genomes of the four subterranean diving beetles sequenced are well conserved, with no differences in the overall structure (number, order and orientation) and nucleotide composition compared with those of surface dytiscids. However, further analyses need to target the adaptive variation in individual mt genes, which would require a phylogenetic framework and additional contrasts among related surface and independently evolved subterranean taxa. These comparisons would enable tests of positive selection at the amino acid/nucleotide level in PCGs, to determine whether any metabolic changes in mt genes have evolved during the adaptation of species to subterranean life. The mt genomes reported here provide a basis for these future comparative analyses to be conducted on the ~100 subterranean dytiscids in the genera *Paroster* and *Limbodessus* from the calcrete archipelago of central Western Australia.

Acknowledgements

We thank Flora, Peter and Paul Axford of the Sturt Meadows Station for providing both accommodation and access to their property. This research was funded by Australian Research Council linkage grants (LP100200494 and LP140100555) to ADA, SJBC, WFH, and PM, with industry partners Minara Resources Limited, the Western Australian Museum, the South Australian Museum, Biota Environmental Sciences, Bennelongia Pty Ltd and the Department of Biodiversity, Conservation and Attractions (WA), and by funding from Bioplatforms Australia and the School of Biological Sciences, the University of Adelaide.

References

Andersson, S. G., and Kurland, C. G. (1998). Reductive evolution of resident genomes. *Trends in Microbiology* 6, 263–268. doi:10.1016/S0966-842X(98)01312-2

Balke, M., and Ribera, I. (2004). Jumping across Wallace's line: *Allodessus* Guignot and *Limbodessus* Guignot revisited (Coleoptera: Dytiscidae, Bidessini) based on molecular–phylogenetic and morphological data. *Australian Journal of Entomology* 43, 114–128. doi:10.1111/j.1440-6055.2004.00415.x

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

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Boore, J. L., Lavrov, D. V., and Brown, W. M. (1998). Gene translocation links insects and crustaceans. Nature 392, 667–668. doi:10.1038/33577

Burger, G., Gray, M. W., and Lang, B. F. (2003). Mitochondrial genomes: anything goes. *Trends in Genetics* 19, 709–716. doi:10.1016/j.tig.2003.10.012

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. doi:10.1186/1471-2105-10-421

Cameron, S. L. (2014). How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. *Systematic Entomology* 39, 400–411. doi:10.1111/syen.12071

Cameron, S. L., and Whiting, M. F. (2008). The complete mitochondrial genome of the tobacco hornworm, Manduca sexta (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. *Gene* 408, 112–123. doi:10.1016/j.gene.2007.10.023

Cameron, S. L., Sullivan, J., Song, H., Miller, K. B., and Whiting, M. F. (2009). A mitochondrial genome phylogeny of the Neuropterida (lace-wings, alderflies and snakeflies) and their relationship to the other holometabolous insect orders. *Zoologica Scripta* 38, 575–590. doi:10.1111/j.1463-6409.2009.00392.x

Caruccio, N. (2011). Preparation of next-generation sequencing libraries using Nextera™ technology: simultaneous DNA fragmentation and adaptor tagging by in vitro transposition. In '*High-throughput Next Generation Sequencing: Methods and Applications*'. (Eds Y. M. Kwon and S. C. Ricke.) pp. 241–255. (Humana Press: Totowa, NJ.)

Chevreux, B., Wetter, T., and Suhai, S. (1999). Genome sequence assembly using trace signals and additional sequence information. In '*German Conference on Bioinformatics*'. pp. 45–56. (Heidelberg.)

Cooper, S. J. B., Hinze, S., Leys, R., Watts, C. H. S., and Humphreys, W. F. (2002). Islands under the desert: molecular systematics and evolutionary origins of stygobitic water beetles (Coleoptera: Dytiscidae) from central Western Australia. *Invertebrate Systematics* 16, 589–590. doi:10.1071/IT01039

Culver, D. C., White, W. B., and White, W. B. (2005). '*Encyclopedia of Caves*.' (Elsevier: Amsterdam.)

Dai, L.-S., Li, S., Yu, H.-M., Wei, G., Wang, L., Qian, C., Zhang, C.-F., Li, J., Sun, Y., and Zhao, Y. (2017). Mitochondrial genome of the sweet potato hornworm, Agrius convolvuli (Lepidoptera: Sphingidae), and comparison with other Lepidoptera species. *Genome* 60, 128–138. doi:10.1139/gen-2016-0058

Di Lorenzo, T., Di Marzio, W. D., Spigoli, D., Baratti, M., Messana, G., Cannicci, S., and Galassi, D. M. P. (2015). Metabolic rates of a hypogean and an epigean species of copepod in an alluvial aquifer. *Freshwater Biology* 60, 426–435. doi:10.1111/fwb.12509

Fenn, J., Cameron, S., and Whiting, M. (2007). The complete mitochondrial genome sequence of the Mormon cricket (Anabrus simplex: Tettigoniidae: Orthoptera) and an analysis of control region variability. *Insect Molecular Biology* 16, 239–252. doi:10.1111/j.1365-2583.2006.00721.x

Friedrich, M., and Muqim, N. (2003). Sequence and phylogenetic analysis of the complete mitochondrial genome of the flour beetle Tribolium castanaeum. *Molecular Phylogenetics and Evolution* 26, 502–512. doi:10.1016/S1055-7903(02)00335-4

Galassi, D. M. P. (2001). Groundwater copepods: diversity patterns over ecological and evolutionary scales. *Hydrobiologia* 453, 227–253. doi:10.1023/A:1013100924948

Grigoriev, A. (1998). Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Research* 26, 2286–2290. doi:10.1093/nar/26.10.2286

Guzik, M. T., Cooper, S. J. B., Humphreys, W. F., and Austin, A. D. (2009). Fine-scale comparative phylogeography of a sympatric sister species triplet of subterranean diving beetles from a single calcrete aquifer in Western Australia. *Molecular Ecology* 18, 3683–3698. doi:10.1111/j.1365-294X.2009.04296.x

Guzik, M. T., Austin, A. D., Cooper, S. J. B., Harvey, M. S., Humphreys, W. F., Bradford, T., Eberhard, S. M., King, R. A., Leys, R., Muirhead, K. A. (2011) Is the Australian subterranean fauna uniquely diverse? *Invertebrate Systematics* 24, 407-418.

Hahn, C., Bachmann, L., and Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. *Nucleic Acids Research* 41, e129. doi:10.1093/nar/gkt371

Hüppop, K. (2000). How do cave animals cope with the food scarcity in caves? In '*Subterranean Ecosystems: Ecosystems of the world*'. Vol. 30. (Eds H. Wilkins, D. C. Culver, and W. F. Humphreys.) pp. 159–188. (Elsevier: Amsterdam.)

Hwang, U. W., Park, C. J., Yong, T. S., and Kim, W. (2001). One-step PCR amplification of complete arthropod mitochondrial genomes. *Molecular Phylogenetics and Evolution* 19, 345–352. doi:10.1006/mpev.2001.0940

Jeffery, W. R. (2001). Cavefish as a model system in evolutionary developmental biology. *Developmental Biology* 231, 1–12. doi:10.1006/dbio.2000.0121

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., and Duran, C. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi:10.1093/bioinformatics/bts199

Kocher, A., Kamilari, M., Lhuillier, E., Coissac, E., Péneau, J., Chave, J., and Murienne, J. (2014). Shotgun assembly of the assassin bug Brontostoma colossus mitochondrial genome (Heteroptera, Reduviidae). *Gene* 552, 184–194. doi:10.1016/j.gene.2014.09.033

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9, 357–359. doi:10.1038/nmeth.1923

Leijs, R., Van Nes, E. H., Watts, C. H. S., Cooper, S. J. B., Humphreys, W. F., and Hogendoorn, K. (2012). Evolution of blind beetles in isolated aquifers: a test of alternative modes of speciation. *PLoS One* 7, e34260. doi:10.1371/journal.pone.0034260

Leys, R., and Watts, C. H. S. (2008). Systematics and evolution of the Australian subterranean hydroporine diving beetles (Dytiscidae), with notes on Carabhydrus. *Invertebrate Systematics* 22, 217–225. doi:10.1071/IS07034

Leys, R., Watts, C. H. S., Cooper, S. J. B., and Humphreys, W. F. (2003). Evolution of subterranean diving beetles (Coleoptera: Dytiscidae Hydroporini, Bidessini) in the arid zone of Australia. *Evolution* 57, 2819–2834. doi:10.1554/03-165

Linard, B., Arribas, P., Andújar, C., Crampton-Platt, A., and Vogler, A. (2016). Lessons from genome skimming of arthropod-preserving ethanol. *Molecular Ecology Resources* 16, 1365–1377. doi:10.1111/1755-0998.12539

Pons, J., Ribera, I., Bertranpetit, J., and Balke, M. (2010). Nucleotide substitution rates for the full set of mitochondrial protein-coding genes in Coleoptera. *Molecular Phylogenetics and Evolution* 56, 796–807. doi:10.1016/j.ympev.2010.02.007

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi:10.1093/bioinformatics/btr026

Sheffield, N., Song, H., Cameron, S., and Whiting, M. (2008). A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. *Molecular Biology and Evolution* 25, 2499–2509. doi:10.1093/molbev/msn198

Song, H., Sheffield, N. C., Cameron, S. L., Miller, K. B., and Whiting, M. F. (2010). When phylogenetic assumptions are violated: base compositional heterogeneity and among-site rate variation in beetle mitochondrial phylogenomics. *Systematic Entomology* 35, 429–448. doi:10.1111/j.1365-3113.2009.00517.x

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30, 2725–2729. doi:10.1093/molbev/mst197

Wan, X., Hong, M. Y., Liao, A., Kim, M. I., Kim, K. G., Han, Y. S., and Kim, I. (2012). Complete mitochondrial genome of a carabid beetle, Damaster mirabilissimus mirabilissim (Coleoptera: Carabidae). *Entomological Research* 42, 44–54. doi:10.1111/j.1748-5967.2011.00355.x

Watts, C. H. S., and Humphreys, W. F. (2009). Fourteen new Dytiscidae (Coleoptera) of the genera *Limbodessus* Guignot, *Paroster* Sharp, and *Exocelina* Broun from underground waters in Australia. *Transactions of the Royal Society of South Australia* 133, 62–107. doi:10.1080/03721426.2009.10887112

Supplementary Figures

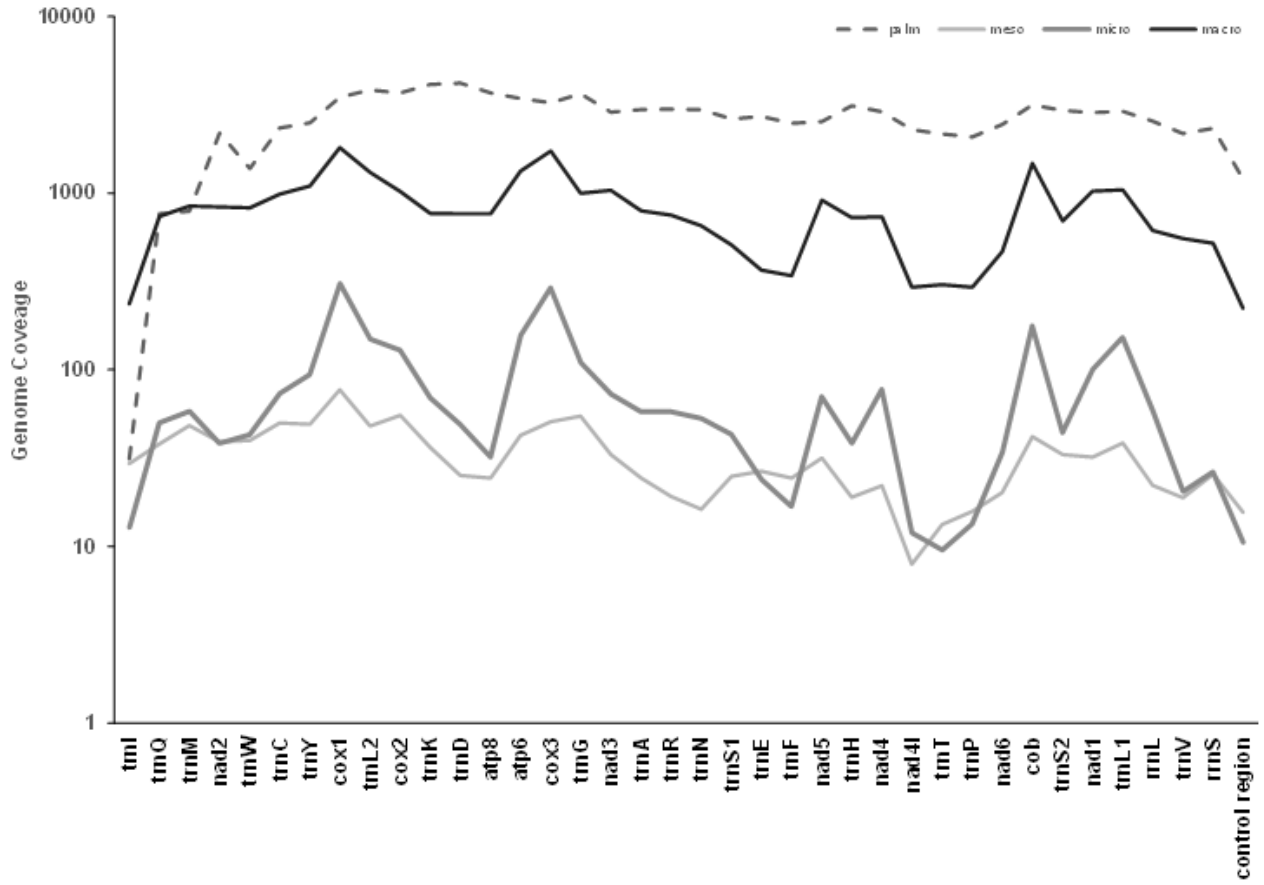

Fig. S5.1 Total genome coverage, for individual mt genes, by raw sequence data generated for the four subterranean beetle species (*L. palmulaoides* – palm; *P. macrosturtensis* – macro; *P. microsturtensis* – micro; *P. mesosturtensis* – meso). Codes for individual genes are as given in Fig. 5.1.

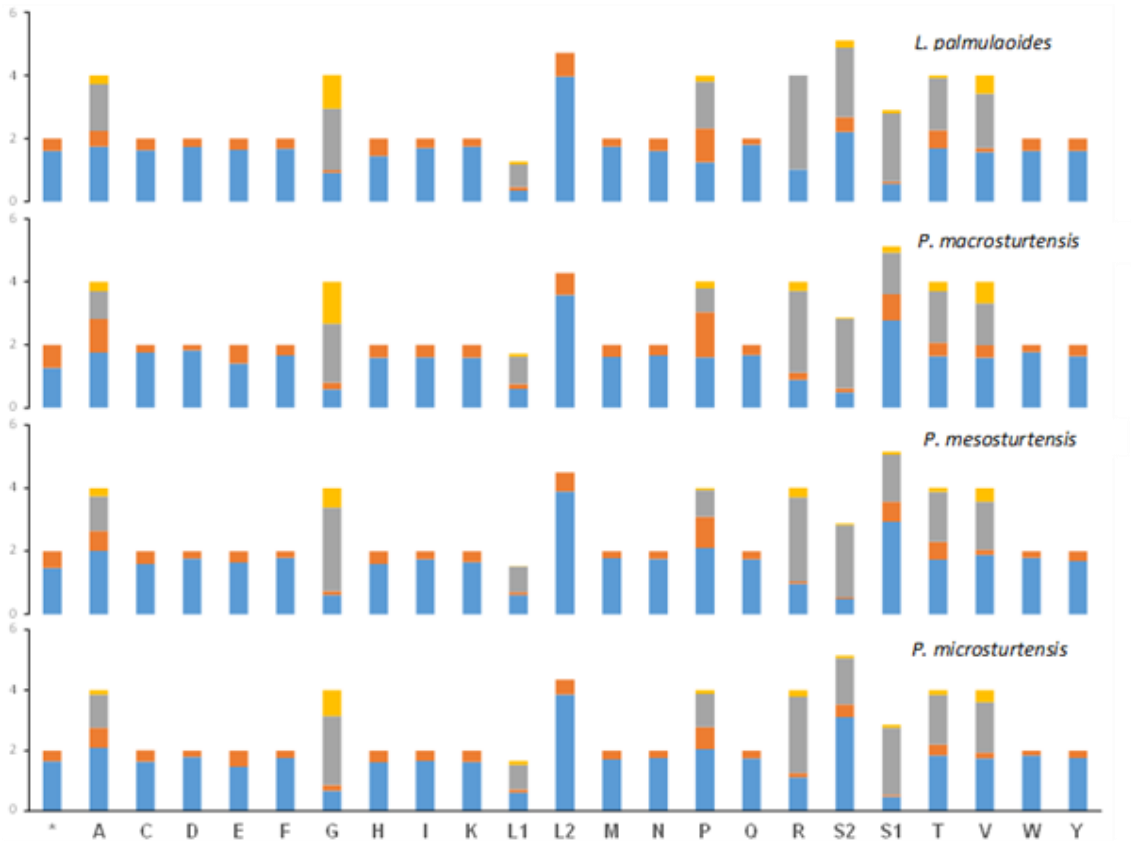

Fig S5.2 The relative synonymous codon usage in four subterranean (*P. macrosturtensis*, *P. mesosturtensis*, *P. microsturtensis and L. palmulaoides*) dytiscid beetles mt genomes. Colours indicate individual codons for each amino acid. One letter amino acid code is used.

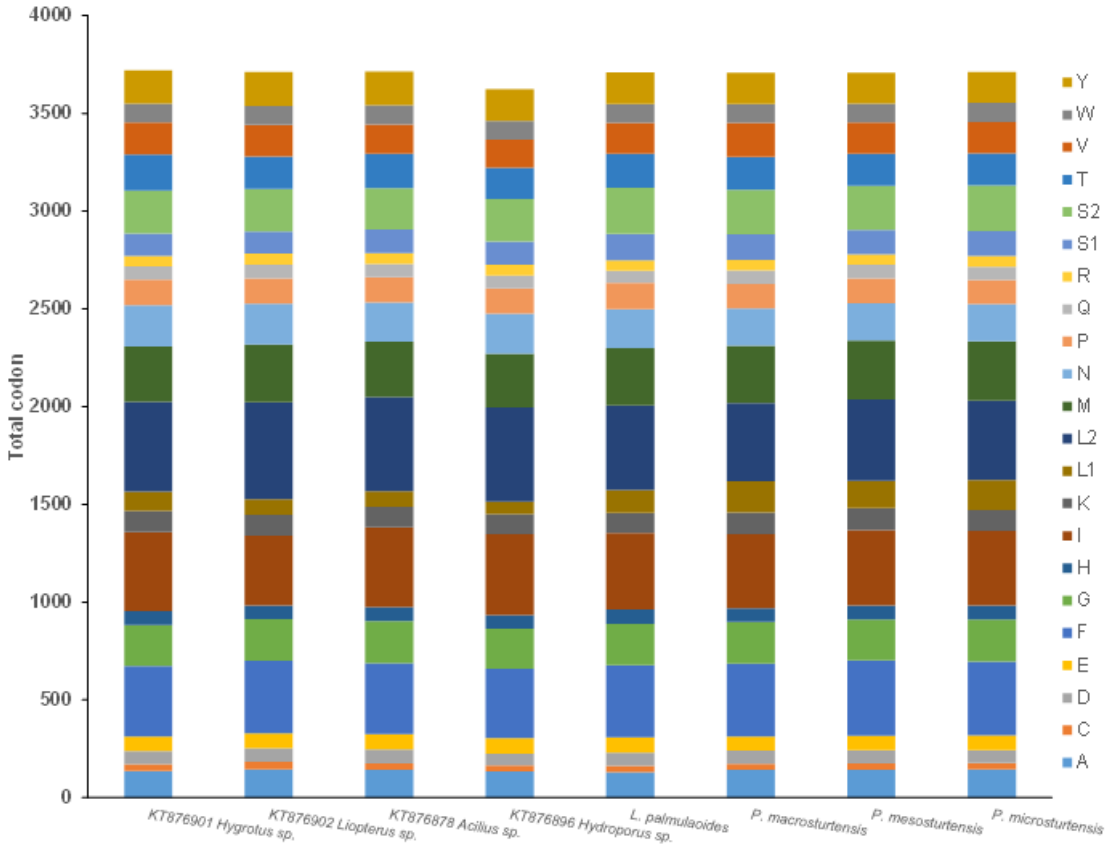

Fig S5.3 Total codons in eight dytiscid beetle species mt genomes, four surface and four subterranean species one letter amino acid code is used.

Supplementary Tables

Table S5.1 Coverage of the genomes sequenced for each of the subterranean species, including the final combined coverage for each of the sequences used in the study.

		COVERAGE			
Sample No.	no. of sequences	Mean	Std Dev	min	max
L. palmulaoides	155,473	2612.5	926.7	22	4463
25542	53,960	887.9	511.8	0	2541
27821	5,767	50.9	28.4	0	167
28086	43,414	746.0	207.5	22	1157
28088	52,335	911.7	319.4	0	1555
P. macrosturtensis	56,958	923.7	550.0	23	2476
25544	27,253	475.9	347.4	2	1395
25545	22,303	378.3	192.0	8	951
27822	7,031	62.8	39.6	7	197
28084	96	1.6	3.2	0	25
28085	416	7.2	11.0	0	96
28089	48	0.8	1.9	0	16
P. mesosturtensis	7,600	66.8	291.7	4	3555
27823	7,600	66.8	291.7	4	3555
P. microsturtensis	8,377	118.9	181.9	2	1807
25543	5,302	91.5	103.4	0	495
27824	3,076	27.4	152.4	0	1749

Table S5.2 The quality scores obtained for the genomes sequenced for each of the subterranean species, including the final combined coverage for each of the sequences used in the study.

Species	QUALITY SCORES		
	Confidence Mean	% at least Q20	% at least Q30
L. palmulaoides	34.4	94.5	85.5
25542	34.9	95.8	87.8
27821	524.4	96.9	96.1
28086	34.4	94.4	85.4
28088	33.8	93.2	82.7
P. macrosturtensis	34.3	94.2	84.8
25544	34.5	94.8	85.9
25545	33.5	93.1	81.5
27822	298.3	96.7	96.0
28084	6359.0	93.5	85.6
28085	2070.5	92.9	83.7
28089	9308.3	95.7	90.7
P. mesosturtensis	37.2	96.9	96.1
27823	37.2	96.9	96.1
P. microsturtensis	35.4	96.0	89.6
25543	34.9	95.6	87.7
27824	37.3	97.0	96.2

Table S5.3 AT content and AT and GC skews of the dytiscid genomes used in this study, calculated for the whole genome and each compartment of the genome (PCGs, tRNAs, rRNAs and the CR).

	TOTAL	T(U)	C	A	G	A+T%	AT SKEW	GC SKEW
Whole								
KT876878 Acilius sp.	20689.0	39.0	11.0	42.4	7.6	81.41	0.041	-0.179
KT876885 Colymbetes sp.	16211.0	37.2	13.4	40.2	9.2	77.34	0.039	-0.186
KT876896 Hydroporus sp.	23380.0	39.3	9.4	44.2	7.2	83.42	0.059	-0.136
KT876897 Hydroporus sp.	17693.0	38.9	11.4	42.3	7.4	81.21	0.041	-0.211
KT876899 Hygrotus sp.	16730.0	35.9	14.5	40.2	9.3	76.17	0.057	-0.216
KT876900 Hygrotus sp.	17704.0	37.6	13.0	40.2	9.2	77.81	0.034	-0.170
KT876901 Hygrotus sp.	17071.0	38.0	12.3	41.1	8.6	79.06	0.039	-0.175
KT876902 Liopterus sp.	16541.0	37.5	12.4	41.7	8.5	79.19	0.053	-0.188
L. palmulaoides	16862.0	36.4	14.7	40.2	8.7	76.63	0.050	-0.254
P. macrosturtensis	16673.0	35.8	15.5	39.3	9.4	75.09	0.046	-0.242
P. mesosturtensis	16663.0	37.0	14.3	40.1	8.7	77.03	0.040	-0.246
P. microsturtensis	16483.0	36.9	14.3	39.6	9.2	76.49	0.036	-0.221
PCGs								
KT876878 Acilius sp.	11212.0	43.9	11.0	33.3	11.8	77.19	-0.138	0.035
KT876885 Colymbetes sp.	11224.0	43.0	12.3	32.3	12.5	75.25	-0.143	0.009
KT876896 Hydroporus sp.	10946.0	44.2	10.4	34.3	11.1	78.51	-0.127	0.037
KT876897 Hydroporus sp.	11212.0	43.4	11.0	34.1	11.6	77.46	-0.121	0.026
KT876899 Hygrotus sp.	11218.0	41.6	13.6	31.9	13.0	73.45	-0.133	-0.021
KT876900 Hygrotus sp.	10979.0	42.8	12.3	32.5	12.4	75.32	-0.138	0.001
KT876901 Hygrotus sp.	11231.0	43.5	11.9	32.9	11.7	76.37	-0.138	-0.006
KT876902 Liopterus sp.	11211.0	43.7	11.0	33.4	11.9	77.10	-0.133	0.042
L. palmulaoides	11200.0	41.5	13.1	32.4	13.0	73.91	-0.123	-0.001
P. macrosturtensis	11198.0	41.5	13.9	30.8	13.8	72.30	-0.147	-0.006
P. mesosturtensis	11198.0	42.7	12.7	32.1	12.5	74.79	-0.143	-0.010
P. microsturtensis	11209.0	42.5	12.8	31.8	12.9	74.32	-0.143	0.001
tRNA								
KT876878 Acilius sp.	1461.0	39.2	8.5	40.7	11.6	79.95	0.019	0.154
KT876885 Colymbetes sp.	1467.0	38.4	9.9	39.3	12.5	77.64	0.011	0.116
KT876896 Hydroporus sp.	1453.0	39.8	7.7	41.2	11.3	81.00	0.016	0.188
KT876897 Hydroporus sp.	1452.0	39.7	7.6	41.0	11.6	80.72	0.015	0.207
KT876899 Hygrotus sp.	1453.0	38.3	9.2	40.1	12.5	78.39	0.022	0.153
KT876900 Hygrotus sp.	1445.0	38.4	9.3	40.1	12.1	78.55	0.022	0.129
KT876901 Hygrotus sp.	1463.0	39.1	8.9	40.6	11.4	79.70	0.019	0.125
KT876902 Liopterus sp.	1469.0	39.2	8.7	40.2	11.8	79.44	0.013	0.152
L. palmulaoides	1460.0	38.4	9.0	39.3	13.3	77.67	0.012	0.190
P. macrosturtensis	1454.0	38.4	8.7	41.0	11.9	79.37	0.033	0.153
P. mesosturtensis	1455.0	38.8	8.2	40.8	12.2	79.59	0.026	0.192
P. microsturtensis	1457.0	38.1	8.6	40.9	12.4	79.00	0.036	0.176
rRNA								
KT876878 Acilius sp.	2150.0	42.5	6.4	38.8	12.2	81.35	-0.045	0.312

KT876885 Colymbetes sp.	2121.0	42.5	6.5	38.8	12.2	81.33	-0.046	0.308
KT876896 Hydroporus sp.	1974.0	42.2	6.6	39.5	11.8	81.66	-0.033	0.282
KT876897 Hydroporus sp.	1984.0	42.0	6.4	39.9	11.7	81.91	-0.026	0.298
KT876899 Hygrotus sp.	2145.0	43.7	6.1	37.4	12.9	81.07	-0.078	0.360
KT876900 Hygrotus sp.	2143.0	42.6	6.3	39.2	12.0	81.75	-0.042	0.315
KT876901 Hygrotus sp.	2103.0	43.3	6.6	38.2	12.0	81.46	-0.062	0.292
KT876902 Liopterus sp.	2144.0	44.2	6.0	37.4	12.4	81.58	-0.083	0.347
L. palmulaoides	2103.0	42.1	6.3	37.6	14.0	79.65	-0.057	0.379
P. macrosturtensis	2111.0	41.7	6.3	37.6	14.4	79.25	-0.052	0.393
P. mesosturtensis	2101.0	41.8	6.0	38.3	13.9	80.06	-0.044	0.394
P. microsturtensis	2106.0	41.8	5.9	38.4	13.9	80.20	-0.042	0.400
CR								
KT876878 Acilius sp.	4991.0	41.3	7.1	47.8	3.9	89.08	0.073	-0.292
KT876885 Colymbetes sp.	1263.0	41.8	6.6	46.6	5.1	88.36	0.054	-0.129
KT876896 Hydroporus sp.	4991.0	41.3	5.5	48.6	4.5	89.96	0.082	-0.106
KT876897 Hydroporus sp.	3165.0	44.8	6.0	47.4	1.8	92.20	0.028	-0.530
KT876899 Hygrotus sp.	4903.0	42.3	9.9	42.6	5.2	84.91	0.005	-0.305
KT876900 Hygrotus sp.	4717.0	44.8	6.4	44.2	4.6	88.93	-0.006	-0.161
KT876901 Hygrotus sp.	1521.0	43.7	5.8	46.4	4.1	90.07	0.029	-0.166
KT876902 Liopterus sp.	3093.0	44.0	6.1	45.8	4.1	89.85	0.020	-0.197
L. palmulaoides	1068.0	45.4	10.2	39.0	5.3	84.46	-0.075	-0.313
P. macrosturtensis	4456.0	42.5	9.9	42.4	5.2	84.92	-0.001	-0.310
P. mesosturtensis	2978.0	42.3	9.9	42.5	5.3	84.79	0.002	-0.302
P. microsturtensis	1498.0	42.4	10.2	41.7	5.7	84.05	-0.009	-0.280

Table S5.4 Codon number and relative synonymous codon usage in four subterranean mt protein coding genes.

	P. microsturtensis		*P. macrosturtensis*		*P. mesosturtensis*		*L. palmulaoides*	
Codon	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU
UUU(F)	330	1.75	311	1.67	342	1.78	310	1.68
UUC(F)	47	0.25	61	0.33	43	0.22	59	0.32
UUA(L)	360	3.86	337	3.59	360	3.89	363	3.96
UUG(L)	46	0.49	64	0.68	56	0.61	70	0.76
CUU(L)	57	0.61	55	0.59	56	0.61	31	0.34
CUC(L)	10	0.11	16	0.17	8	0.09	11	0.12
CUA(L)	76	0.81	82	0.87	74	0.8	66	0.72
CUG(L)	11	0.12	9	0.1	1	0.01	9	0.1
AUU(I)	322	1.66	310	1.61	340	1.74	339	1.7
AUC(I)	65	0.34	75	0.39	51	0.26	61	0.3
AUA(M)	264	1.7	242	1.62	274	1.77	263	1.75
AUG(M)	46	0.3	57	0.38	36	0.23	37	0.25
GUU(V)	69	1.74	69	1.59	75	1.88	62	1.57
GUC(V)	8	0.2	17	0.39	7	0.17	5	0.13
GUA(V)	66	1.66	58	1.33	61	1.52	68	1.72
GUG(V)	16	0.4	30	0.69	17	0.42	23	0.58
UCU(S)	140	3.11	124	2.77	128	2.93	101	2.2
UCC(S)	19	0.42	37	0.83	28	0.64	22	0.48
UCA(S)	69	1.53	59	1.32	65	1.49	101	2.2
UCG(S)	4	0.09	9	0.2	4	0.09	11	0.24
CCU(P)	64	2.05	51	1.61	67	2.09	42	1.25
CCC(P)	23	0.74	45	1.42	32	1	36	1.07
CCA(P)	34	1.09	24	0.76	27	0.84	50	1.49
CCG(P)	4	0.13	7	0.22	2	0.06	6	0.18
ACU(T)	76	1.84	69	1.64	71	1.73	74	1.69
ACC(T)	15	0.36	18	0.43	23	0.56	25	0.57
ACA(T)	68	1.65	68	1.62	65	1.59	72	1.65
ACG(T)	6	0.15	13	0.31	5	0.12	4	0.09
GCU(A)	76	2.1	62	1.75	72	2.01	57	1.75
GCC(A)	24	0.66	38	1.07	23	0.64	16	0.49
GCA(A)	40	1.1	31	0.87	39	1.09	48	1.48
GCG(A)	5	0.14	11	0.31	9	0.25	9	0.28
UAU(Y)	137	1.75	130	1.64	132	1.68	129	1.61
UAC(Y)	20	0.25	29	0.36	25	0.32	31	0.39
UAA(*)	9	1.64	7	1.27	8	1.45	8	1.6
UAG(*)	2	0.36	4	0.73	3	0.55	2	0.4
CAU(H)	58	1.61	56	1.6	57	1.58	52	1.44
CAC(H)	14	0.39	14	0.4	15	0.42	20	0.56
CAA(Q)	59	1.74	58	1.68	59	1.74	57	1.81
CAG(Q)	9	0.26	11	0.32	9	0.26	6	0.19
AAU(N)	166	1.75	158	1.67	166	1.75	159	1.61

AAC(N)	24	0.25	31	0.33	24	0.25	39	0.39
AAA(K)	86	1.62	88	1.6	92	1.64	89	1.75
AAG(K)	20	0.38	22	0.4	20	0.36	13	0.25
GAU(D)	59	1.79	59	1.82	59	1.76	60	1.74
GAC(D)	7	0.21	6	0.18	8	0.24	9	0.26
GAA(E)	56	1.47	52	1.41	61	1.63	63	1.64
GAG(E)	20	0.53	22	0.59	14	0.37	14	0.36
UGU(C)	26	1.63	28	1.75	26	1.58	26	1.63
UGC(C)	6	0.38	4	0.25	7	0.42	6	0.38
UGA(W)	91	1.84	86	1.76	87	1.78	79	1.61
UGG(W)	8	0.16	12	0.24	11	0.22	19	0.39
CGU(R)	15	1.11	12	0.89	13	0.96	14	1.02
CGC(R)	2	0.15	3	0.22	1	0.07	0	0
CGA(R)	34	2.52	35	2.59	36	2.67	41	2.98
CGG(R)	3	0.22	4	0.3	4	0.3	0	0
AGU(S)	21	0.47	22	0.49	21	0.48	26	0.57
AGC(S)	3	0.07	6	0.13	2	0.05	3	0.07
AGA(S)	100	2.22	99	2.21	100	2.29	99	2.15
AGG(S)	4	0.09	2	0.04	2	0.05	5	0.11
GGU(G)	36	0.67	31	0.58	32	0.61	48	0.91
GGC(G)	11	0.2	12	0.23	7	0.13	4	0.08
GGA(G)	123	2.28	99	1.86	138	2.64	104	1.96
GGG(G)	46	0.85	71	1.33	32	0.61	56	1.06

Chapter 6: General Discussion

6.1 Introduction

In this chapter, the main findings of each empirical chapter are summarised, and general conclusions of the studies presented in the thesis are described. Additionally, the strengths and the weaknesses of the research are discussed, and suggestions for further work into areas of subterranean ecology, microbiome research and genomic structure and function are presented. This chapter concludes with recommendations for both academic and industry stakeholders.

6.2 Summary of thesis

The ecology of subterranean ecosystems, in particular in an Australian context, have to date been largely unexplored and poorly understood. This study was conducted to address the knowledge gaps in the general ecology of the calcrete aquifer, the general beetle microbiome and the trophic ecology of the beetle species present. These aims were investigated by studying six stygobiontic diving beetles from two groundwater ecosystems in the arid region of central Australia using a combination of observational and molecular techniques.

The ecological analyses in chapter 2 were undertaken by utilising collections of 9118 stygobitic macroinvertebrate specimens from 512 sampling events from 116 boreholes over an 11-year period. The results showed that the amount of rainfall and recharge that had occurred in the calcrete at Sturt Meadows was important for the diversity, distribution and abundance of stygofauna. Rainfall in the arid zone of Central Australia is highly episodic, and large events may only occur years apart (Humphreys, 2012). The average pan evaporation is 2400 mm $year^{-1}$, which far exceeds the average yearly rainfall of just over 200 mm. Water monitoring that occurred at the Sturt Meadows calcrete showed that high rainfall events (>30 mm) led to significant recharge of the aquifer, while low rainfall events had no influence on groundwater levels. Rainfall patterns influenced different taxon groups differently in terms of their individual distribution, abundance and dominance. However, there were some overall patterns. High rainfall events led to high overall taxon diversity and low evenness, while low rainfall events, or long periods without recharge events, led to low overall taxon diversity and high evenness.

The microbiome analyses in chapters 3 and 4 were undertaken using random shotgun sequencing (metagenomic analyses) of 15 beetle individuals comprising six species (*Limbodessus lapostaae*, *L. windarraensis* and *L. palmulaoides* (Watts and Humphreys, 1999,

2006), *Paroster macrosturtensis*, *P. mesosturtensis*, and *P. microsturtensis* (Watts and Humphreys, 2006)) from two different calcretes (Sturt Meadows and Laverton). The results showed that both the gut bacterial and viral communities of these beetles vary significantly by site and by life stage. The differences in microbial communities between the adult and larval stages of the same species are most likely indicative of trophic differences between the life stages. For both the bacteria and viruses, the Laverton calcrete showed a higher diversity (H') compared to the Sturt Meadows calcrete.

Prior to the current study, there had not been any published genomic analyses of subterranean dytiscid beetles, except for a small number of nuclear and mt genes used specifically for phylogenetic analyses (Cooper *et al*., 2002; Leys *et al*., 2003; Leijs *et al*., 2012; Tierney *et al*., 2015). Therefore, in order to identify the source of beetle DNA recovered from the shotgun analyses and classify it as either host or prey, in chapter 5, mitochondrial genomes were assembled for four of the six beetle species (*Limbodessus palmulaoides*, *Paroster macrosturtensis*, *P. mesosturtensis* and *P. microsturtensis*) from Sturt Meadows and Laverton calcretes. Unfortunately, for the final two species only partial assemblies could be achieved. The overall structure (gene number, orientation and order) and the mt genome length was consistent with other surface dytiscid beetle species already sequenced (Sheffield *et al*., 2008). The extra length, compared to other beetles, of the mt genomes in all four genomes, came from extra bases in both the intergenic and control regions. These results allow the potential identification of beetle species as prey items in other beetle individuals (see below), but they also provide a basis for future investigations of the molecular evolution of mt genes and whether there is evidence of adaptive evolution for living underground.

6.3 Contributions of this research

Many calcrete aquifers in the Yilgarn Region are impacted by anthropogenic activities, such as mining and pastoral activities. Groundwater may be utilised for either processing of minerals or is removed from the mine site to allow better access to the minerals in question. Additionally, at pastoral stations, the water is extracted for use by stock. As each calcrete aquifer is a semi-closed underground ecosystem accessible from a limited number of boreholes or wells, direct observation of the species within the calcretes is difficult, and indirect means are needed, i.e. by sampling via a limited number of boreholes or wells. Using this approach, the results of chapter 2 suggest a highly dynamic ecosystem that is influenced by recharge. It is possible, even likely, that recharge into the calcrete is causing an increase in

connectivity within the calcrete due to an increase in the water table, thus opening up new habitats and resources and leading to population booms of species in the lower trophic levels. In addition, the recharge could also be bringing in nutrients into the aquifer directly, causing the population booms of the copepods and amphipods.

Subterranean ecosystems are generally considered to be low in energy (Huppop, 2000; Culver and White, 2005), with most of the energy thought to be mediated by the amount and nutrient quality of the water that enters the ecosystem; the remaining energy is generally provided by plant roots and chemoautotrophic bacteria (Deharveng and Bedos, 2000; Humphreys, 2012). Therefore, most energy in the calcrete aquifer system is likely to be affected by seasonal rainfall and plant growth (Humphreys, 2012). In subterranean ecosystems, which are continuously dark, trophic levels are usually truncated with a reduction in both primary producers and top predators. Nutrients in these systems are not necessarily reduced. However, the distribution is generally uneven (Deharveng and Bedos, 2000). In chapter 2, evidence that the Sturt Meadows calcrete connectivity was affected by changes in the water table, with species' distributions increasing with high rainfall and recharge events, further supports the idea that some areas of the calcrete may be better suited to supporting stygofauna than others and that this distribution could shift over time.

Investigation of the trophic levels and diet of the beetles in the Laverton and Sturt Meadows calcretes by direct observation of animals feeding was problematic due to limited access through boreholes. Additionally, the larvae of these beetles are fluid feeders, so there are no remains to identify, making traditional microscopic examinations of gut contents futile. Also lack of information about the identity of the beetle's prey, particularly at the Laverton calcrete, with all possible prey species not having been collected and sequenced, made multiplex PCR and DNA barcoding impractical. Another likely issue is the presence of sister beetle species in the Sturt Meadows calcrete which might be preying on each other, and the need in DNA barcoding techniques to design blocking primers to prevent the host DNA being amplified preferentially over the prey DNA. Thus, due to the high similarity of their sequences, and the short fragments of DNA typically involved in diet analysis, blocking the host DNA while still allowing for the amplification of the closely related beetle species would be very difficult.

In order to overcome the above limitations in investigating the diets and trophic levels of the beetles at Laverton and Sturt Meadows, shotgun sequencing (metagenomic analyses) of whole specimens was trialled. During the processing stage of the shotgun data, all

potentially eukaryotic sequences were removed. For multiple individuals, non-beetle invertebrate mitochondrial DNA was recovered, including amphipod, other beetle species, and some epigean species (Table 6.1). As it was only possible to obtain diet analyses for a limited number of beetle individuals, they were not included as a separate publication on the potential prey items of the different beetle species. However, these results do provide a proof of concept that metagenomics and specifically random shotgun sequencing could be used instead of more traditional methods such as DNA barcoding or metabarcoding when target primer sequences are unknown. One beetle individual was kept in the laboratory before being fed a local epigean amphipod *Austrochiltonia australis* and then immediately killed and preserved in 100% ethanol, while other individuals were processed in the field. In the beetle that was fed before being euthanised in the laboratory, *COI* sequences from the amphipod were recovered. In the field, one beetle from Laverton had *COI* sequences from an amphipod. Although none of the of the amphipods from the Laverton calcrete have been sequenced, the closest match was *Yilgarniella sturtensis,* which is from the Sturt Meadows calcrete, and two beetles from Sturt Meadows had *COI* sequences from gnats, Mycetophilidae and Cecidomyiidae, the former family having a known host association with fungi while the later is associated with plant galls. These epigean species potentially are falling into the calcrete during rainfall events or when the cap on the boreholes is missing. In two *P. macrosturtensis* beetles, mtDNA sequences from other beetle species within the calcrete were recovered. All beetle species were kept in separate tubes once collected in the field, and extracted in a UV hood with only one individual in each library preparation and duel barcodes used to reduce cross-contamination of sequences. Overall, these results also suggest that individuals containing DNA of other beetle species are not due to contamination, but that they are preying on each other. Further, finding amphipod DNA is consistent with previous observational and molecular results (Bradford *et al*., 2014). However, this is the first time that molecular evidence has been provided to support field observations that beetle species are eating each other.

Table 6.4 Prey results from the random shotgun sequencing by host individual. All prey species searched by blastn are the closest match from Genbank, searched on the 14/06/17. Prey species identified by pairwise identity, sequence length. All remaining prey were identified using the map to reference option in Geneious 8.1.9.

	Host Species	Prey Species	Gene	Match Method	Pairwise identity
25542	*Limbodessus palmulaoides*	*Austrochiltonia australis*	*COI*	Blastn	96%
25544	*Paroster macrosturtensis*	*Mycomya affinis*	*COI*	Blastn	95%
		Paroster microsturtensis	*nad5*	Map to reference	99%
27822	*Paroster macrosturtensis*	*Paroster mesosturtensis*	*12S*	Blastn	98%
25545	*Paroster macrosturtensis*	Cecidomyiidae sp.	*COI*	Blastn	90%
27820	*Limbodessus windarraensis*	*Yilgarniella sturtensis*	*COI*	Map to reference	85%

It has been suggested in the past that in subterranean ecosystems there is a reduction in obligate predators compared to epigean ecosystems and many species that belong to families that are normally predatory in surface environments, have diversified in subterranean habitats to become either omnivorous or polyphagous (Gers, 1995; Gibert and Deharveng, 2002). Microbiome diversity is found to differ depending on diet, with omnivores generally having a higher diversity index than predators or herbivores (Yun *et al*., 2014). The fact that there are different beetle size classes within many of the calcretes across the Yilgarn Region has led to the suggestion that the different species have trophic differences (Leijs *et al*., 2012). Therefore, comparing their microbiomes may help to determine if there are differences in their diets.

The viral family results could also provide insights into other animals present in the calcrete food web and into potential prey of the beetles investigated. Using the known host associations of the detected viral families, different interactions between species could potentially be inferred. For example, the presence of viruses specific to plants in the viromes of the beetles, suggests an interaction between the beetles and plant roots, most likely indirectly (i.e. beetles consuming arthropods that feed on plant roots). This interaction is most likely to occur via the consumption of amphipods, which are thought to have plant roots as one of the main parts of their diet (Wildish, 1988). The identification of viruses specific to algae, amoeba and fungi suggest the possible presence of a biofilm in the groundwater. The higher abundance of these viruses in larval beetles compared to the adult stage, suggests they play a larger role in the diet of larvae indirectly, through the partitioning of diet via the beetle's life stage.

6.4 Limitations

The main limitation of studying the trophic ecology of these beetle species was the sample size of the metagenomic analyses, with some beetle species only having one individual sequenced and only 15 beetles sequenced in total. The sequencing in this study was limited by time and the expense of performing the metagenomic analyses. With sequencing costs continually decreasing, and the amount of data generated on the different sequencing platforms rapidly increasing, the ability to multiplex many more individuals together for analyses in the future means this approach is becoming increasingly cost-effective. With the results here, preliminary analyses can be undertaken into their natural microbiomes, but how much these results can be extrapolated to the species level, rather than the individual level, is currently unknown. Another limitation was the small number of beetles, with less than 50% that showed evidence for prey DNA; this is most likely due to the beetles' significantly slower metabolic rate than that found in surface dytiscid species and their ability to go long periods without feeding (K. Jones pers. comm).

By extracting DNA from beetles of different species at the same time and sequencing them in the same sequencing lane, it is impossible for two of the three sequencing runs to entirely rule out contamination as a potential source of DNA when DNA of another beetle species was found in an individual. In only one run was it possible to be confident that the beetle DNA was coming from a prey item, rather than from contamination. Another issue was that in analysing closely related sister species from Sturt Meadows and sometimes having short 150 bp fragments of DNA sequenced, it was not always possible to be sure which species a fragment of DNA was coming from if the fragment came from regions without species-diagnostic sites. Having the mitochondrial genome sequences from each species helped with this situation but, for a large number of sequenced fragments, it was still difficult to identify which species a fragment of DNA originated from.

6.5 Future Work

To improve understanding of calcrete aquifer ecosystems, specifically in the Yilgarn Region of WA, these ecological, genetic and microbial analyses must be expanded beyond the beetle species from just two calcretes. A comprehensive understanding of the ecology and trophic interactions of these calcretes would require a survey of all life within the calcrete; however, current sampling methods are biased against microorganisms, as well as sessile animals. Additionally, sequencing the microbiomes of other species present in the calcrete, plus DNA extracted from the water (eDNA) and plant roots, may provide a much clearer

overall picture as to what species are present and how they are interacting with each other within the calcrete. If one or a few calcretes could be comprehensively investigated, then a complete interaction model of the species present could be constructed. Using this information, keystone species and their trophic position within calcrete ecosystems can be identified. This information could lead to a greater understanding of the natural dynamics within the calcrete, particularly how the loss of one component could affect other parts of the ecosystem. This information could also be applied to other calcretes in the Yilgarn, which have similar ecosystems. Groundwater is both vital for supporting the animals that live in it and the health of the epigean ecosystem (e.g., trees and shrubs) that, in turn, also contributes resources/energy to the calcrete ecosystem and relies on the groundwater for its survival. Therefore, these calcretes are an important ecosystem to conserve due to their vital importance in the arid landscape of Central Australia.

Another important reason for sequencing the microbiomes of species in as many calcretes as possible, particularly those that have been undisturbed by humans, is due to their long history of isolation. The similarity of the ecosystems among the different calcretes provides a valuable opportunity for comparative analyses. Many of the stygobiont species are extremely short-range endemics, completely confined to an individual calcrete, and they have been isolated from their congeners for millions of years (Cooper *et al*., 2002; Leys *et al*., 2003; Cooper *et al*., 2007; Cooper *et al*., 2008). As a result, many of the bacteria and viruses present within these species will be unknown to science, providing an excellent opportunity for bioprospecting and future biotechnological research. These viruses and bacteria could provide a hotspot for the discovery of novel enzymes, antibiotics and signal mimics. For example, some of the sequences found (chapters 3 and 4) appear to be from ancient lineages; for example, the viral sequences that match to species in the family Polydnavirus, or are currently unknown, such as the bacterial sequences that do not match anything on the current sequence databases.

As shown in chapters 3 and 4, the adult and larval beetles have different microbiome diversity indexes with the latter having a higher diversity index than that found in the adult beetles. This higher diversity suggests that the larval beetles have a more diverse diet with a shift away from being an obligate predator, or they are predating on animals that directly feed on the biofilm, while the adult beetles maintain a predatory lifestyle associated with alternative prey species. Also, the presence of specific microbiome signatures for different life stages may, in future, allow for the identification of prey by life stage. The microbiome

results lend support to the idea that the different beetle species within the calcrete are occupying different ecological niches, with this difference potentially providing the selective force that originally led to the size divergence of species within the calcrete, and in the case of the species at Sturt Meadows, contributed to their speciation in sympatry.

Recent work suggests that the three subterranean beetle species from Sturt Meadows have a significantly slower metabolic rate than that found in surface dytiscid species (K. Jones pers. comm). Coupling the already sequenced mt genomes with metabolic gene sequence and expression data in many more independently evolved subterranean lineages may identify the signatures of adaptation. While subterranean ecosystems are thought to be low in energy and reduced in trophic complexity (Gibert and Deharveng, 2002), my study suggests that the calcretes investigated may have a variety of energy sources, and may be less energy deficient than subterranean environments are traditionally considered. Chapters 2, 3 and 4 suggest that energy is being supplied to the calcrete via water recharge, plant roots and chemoautotrophic bacteria. Furthermore, chapters 3 and 4 suggest that in both calcretes predators are still present, with the beetles both having DNA of other invertebrates present and an H' index similar to predators found in other studies (Yun *et al*., 2014), further suggesting that there is less of a reduction in trophic complexity than has been thought to occur in other subterranean environments. The reduction in metabolic rate of the beetles may be an adaptation to the episodic nature of high nutrient availability, or it could be an evolutionary adaptation to their altered breathing mechanism and the limitations of oxygen transfer, as recent experiments suggest they do not need to surface to breathe and take up oxygen from the water through their cuticle (K. Jones pers. comm).

6.6 Conclusion

Subterranean fauna in the Yilgarn region of Australia provides an excellent and unique resource for studying insect microbiomes, subterranean ecology and trophic systems. Studying the role of rainfall and recharge, as well as the bacterial and viral microbiomes of these subterranean dytiscid beetles, has provided insights into how these species are distributed in the calcretes and how they interact with other species in their subterranean environment. This research also highlights how metagenomics can be used to complement traditional methods such as stable isotope analysis and DNA barcoding to investigate food webs and trophic interactions. Extending this research to include more species within the ecosystem, and more calcretes across the Yilgarn Region, and to include more individuals,

would provide for a powerful investigation into the unknown world of the subterranean environment and its complex ecosystem.

References

Bradford, T.M., Humphreys, W.F., Austin, A.D., and Cooper, S.J.B (2014) Identification of trophic niches of subterranean diving beetles in a calcrete aquifer by DNA and stable isotope analyses. *Marine and Freshwater Research* **65**: 95-104.

Cooper, S.J.B., Hinze, S., Leys, R., Watts, C.H.S., and Humphreys, W.F. (2002) Islands under the desert: molecular systematics and evolutionary origins of stygobitic water beetles (Coleoptera: Dytiscidae) from central Western Australia. *Invertebrate Systematics* **16**: 589-590.

Cooper, S.J.B., Bradbury, J.H., Saint, K.M., Leys, R., Austin, A.D., and Humphreys, W.F. (2007) Subterranean archipelago in the Australian arid zone: mitochondrial DNA phylogeography of amphipods from central Western Australia. *Molecular Ecology* **16**: 1533-1544.

Cooper, S.J.B., Saint, K.M., Taiti, S., Austin, A.D., and Humphreys, W.F. (2008) Subterranean archipelago: mitochondrial DNA phylogeography of stygobitic isopods (Oniscidea : Haloniscus) from the Yilgarn region of Western Australia. *Invertebrate Systematics* **22**: 195-203.

Culver, D.C., and White, W.B. (2005) *Encyclopedia of Caves*: Elsevier Amsterdam (The Netherlands).

Deharveng, L., and Bedos, A. (2000) The cave fauna of Southeast Asia. Origin, evolution and ecology. *Ecosystems of the World*: 603-632.

Gers, C. (1995) Stratégie alimentaire de Coléoptères troglobies du genre Aphaenops (Coleoptera, Trechinae). *Mémoires de Biospéologie* **22**: 35-45.

Gibert, J., and Deharveng, L. (2002) Subterranean Ecosystems: A Truncated Functional Biodiversity *BioScience* **52**: 473-481.

Humphreys, W. (2012) Diversity patterns in Australia. *Encyclopedia of caves, 2nd edn (eds White WB, Culver DC)*: 203-219.

Huppop, K. (2000) How do cave animals cope with the food scarcity in caves? *Ecosystems of the World*: 159-188.

Leijs, R., Van Nes, E.H., Watts, C.H.S., Cooper, S.J., Humphreys, W.F., and Hogendoorn, K. (2012) Evolution of blind beetles in isolated aquifers: a test of alternative modes of speciation. *PLoS One* **7**: e34260.

Leys, R., Watts, C.H.S., Cooper, S.J.B., and Humphreys, W.F. (2003) Evolution of subterranean diving beetles (Coleoptera: Dytiscidae : Hydroporini, Bidessini) in the arid zone of Australia. *Evolution* **57**: 2819-2834.

Sheffield, N., Song, H., Cameron, S., and Whiting, M. (2008) A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. *Molecular Biology and Evolution* **25**: 2499-2509.

Tierney, S.M., Cooper, S.J.B., Saint, K.M., Bertozzi, T., Hyde, J., Humphreys, W.F., and Austin, A.D. (2015) Opsin transcripts of predatory diving beetles: a comparison of surface and subterranean photic niches. *Royal Society Open Science* **2**: 140386.

Watts, C.H.S., and Humphreys, W.F. (1999) Three new genera and five new species of Dytiscidae (Coleoptera) from underground waters in Australia. *Records of the South Australian Museum* **32**: 121-142.

Watts, C.H.S, and Humphreys, W.F. (2006) Twenty-Six New Dytiscidae (Coleoptera) of the Genera *Limbodessus* Guignot and *Nirripirti* Watts and Humphreys, from Underground Waters in Australia. *Transactions of the Royal Society of South Australia* **130**: 123-185.

Wildish, D.J. (1988) Ecology and natural history of aquatic Talitroidea. *Canadian Journal of Zoology* **66**: 2340-2359.

Yun, J.-H., Roh, S.W., Whon, T.W., Jung, M.-J., Kim, M.-S., Park, D.-S. *et al*. (2014) Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Applied and Environmental Microbiology* **80**: 5254-5264.