PHYLOGENOMICS OF THE PULMONATE LAND SNAILS

Luisa Cinzia Teasdale

Submitted in total fulfilment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

January 2017

School of BioSciences

The University of Melbourne

Phylogenomics of the Pulmonate Land Snails

by Luisa C. Teasdale © 2017

Supervisors: Adnan Moussalli and Devi Stuart-Fox

Cover Image: *Montidelos orcardis* taken by Luisa C. Teasdale

For Kevin, and my parents: Maria and Stephen.

Abstract

The pulmonates are the most speciose gastropod lineage and are highly diverse in morphological form and habitat. The evolutionary relationships among the pulmonates have remained controversial despite a long history of scientific study. Recent molecular studies have placed traditionally pulmonate (air-breathing) and non-pulmonate taxa into Panpulmonata; however, the relationships within this new group are still poorly understood. Incongruence between molecular studies has generally resulted from a lack of informative loci but the advent of next generation sequencing technologies means it is now feasible to produce large genetic datasets for non-model organisms. The main aim of my thesis was to investigate the timing and pattern of evolutionary relationships within the Panpulmonata, at multiple taxonomic scales, using phylogenomic datasets.

The qualification of orthology is a significant challenge when developing large, multi-locus datasets for phylogenetics from transcriptome assemblies. In Chapter 2, I identified 500 orthologous single-copy genes from 21 transcriptome assemblies across the Eupulmonata (mostly terrestrial land snails and slugs) using a thorough approach to orthology determination, involving manual alignment curation and gene tree assessment. I further qualified orthology by sequencing the genes from the genomic DNA of 22 representatives of the Australian land snail family Camaenidae using exon capture. Through comparison, I also found that automated orthology determination approaches can be susceptible to transcriptome assembly errors.

I then used the orthologous genes identified in Chapter 2 to investigate the pattern and timing of evolution across Panpulmonata in Chapter 3. My dataset included representatives of all major clades within Panpulmonata including a wide representation of the stylommatophoran land snails, the most successful lineage of terrestrial molluscs. Maximum likelihood and Bayesian analyses confirm that Panpulmonata is monophyletic, and that Pulmonata is not monophyletic, implying that air-breathing has likely evolved more than once. Within Panpulmonata I show strong support for relationships previously unsupported or weakly supported in molecular analyses, including the Geophila, and the Pylopulmonata, a clade that unites the operculate panpulmonates. Molecular dating suggests a Permian or Early Triassic origin for Panpulmonata and a Triassic/Jurassic boundary origin for Eupulmonata and the freshwater Hygrophila.

In addition to investigating deep relationships within Panpulmonata, I used a similar approach to investigate phylogenetic relationships on a shallower scale within a land snail family, the Rhytididae. Australia has the highest taxonomic diversity of the Rhytididae, a carnivorous family of land snails with a Gondwanan distribution. Previous higher classifications of the Australian Rhytididae are based on limited morphological characters and have not been assessed with molecular evidence. I present a molecular phylogeny of the Australian Rhytididae based on a large multi-locus dataset comprising nuclear exons sequenced using exon capture. I identified four major monophyletic lineages within the Australian Rhytididae. I also show that there is a high amount of unrecognised diversity, particularly in the smaller rhytidids. Contrary to shell morphology, on which the current taxonomy is based, a number of currently recognised genera are either polyphyletic or paraphyletic. The Australian lineages all resulted from an apparent pulse of diversification approx. 45-30 Ma. Given the South African *Nata* and the New Zealand *Delos* and *Schizoglossa* also belong to this clade, this date suggests that cross-water dispersal has played a role in the evolution of this group.

Declaration

This is to certify that:

i. The thesis comprises only my original work towards the PhD except where indicated in the preface,

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

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

LUisa T

Signature: ________________________ Date: ________6/01/2017________

Luisa C. Teasdale

Preface

This thesis includes a series of independent publications; consequently there may be some repetition between chapters. Chapters 2 – 4 comprise co-authored manuscripts that have either been published or will be submitted for publication. This thesis includes the largely unchanged reprint of the following previously published journal article, which has been included as a chapter with the permission of all co-authors:

Teasdale, L.C., Koehler, F., Murray, K.D., O'Hara, T. and Moussalli, A. 2016. Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon-capture. Molecular Ecology Resources, DOI: http://dx.doi.org/10.1101/035543 (**Chapter 2**)

I am the primary author of this work and was responsible for experimental design, performing the research, analysing the data, and writing the manuscript. Adnan Moussalli contributed to experimental design, specimen collection, and data analysis. Frank Köhler assisted with specimen collection, and Kevin Murray and Tim O'Hara provided bioinformatics assistance. All co-authors provided feedback on the manuscript. Additionally the following manuscripts, for which I am also primary author, are in preparation for publication:

Teasdale, L.C., Hugall, A., Köhler, F., Herbert, D., Barker, G., O'Hara, T. and Moussalli, A. In prep. The pace and pattern of pulmonate evolution. *Molecular Biology and Evolution*. (**Chapter 3**)

Teasdale, L.C. and Moussalli, A. In prep. Phylogenetic investigation of the Australian Rhytididae using exon capture. *Molecular Phylogenetics and Evolution*. (**Chapter 4**)

Adnan Moussalli contributed to experimental design, specimen collection, and data analysis for Chapters 3 and 4. For Chapter 3, Frank assisted with specimen collection, Tim O'Hara assisted with the bioinformatics, Andrew Hugall provided assistance with the phylogenetic analyses, Gary Barker provided the morphological data, and David Herbert assisted with the biological interpretation. All co-authors provided feedback on the manuscripts.

Acknowledgements

First and foremost I would like to thank my supervisor Adnan Moussalli for all his help, patience, and support throughout my PhD. I have valued our rigorous scientific discussions and without his help this PhD would not have been possible. I also thank my co-supervisor, Devi Stuart-Fox, for her support, mentorship, and guidance. I thank the members of my advisory committee, Michael Keogh and Ary Hoffman for their time and expertise.

I thank my collaborators for their contribution to this work. Specifically, I thank Andrew Hugall for invaluable discussions as well as mentoring me through the tangled world of phylogenetic analysis. I thank Frank Köhler for his assistance with field work and the biological interpretation of my results, but perhaps not for his sense of direction. I also thank Gary Barker for his guidance through the complexities of snail morphology. I am gratefully for Kevin Bonham, Francesco Criscione, and Winston Ponder who generously collected and provided samples. I would also like to thank the service providers who conducted a lot of the lab work and assisted with troubleshooting, namely Roger Neilsen and Travis Glenn at the Georgia Genomics Facility, and Jake Enk and Alison Devault at MYcroarray. I would also like to thank a number of people at Museum Victoria who have provided assistance, support, and friendship during my PhD, including Stella Claudius, Maggie Haines, Melanie Mackenzie, Claire McLean, Michela Mitchell, Tim O'Hara, Chris Rowley, Jo Sumner, and Jo Taylor. I would particular like to thank Martin Gormon for his mentorship, which I have valued more than I think he realises. I would also like to thank Craig Moritz and his lab at ANU for adopting me while my supervisors were on sabbatical. I thank Maddie and Tom for their friendship and support at Moor St and my Canberra friends Panda, Brian, and Claire. I thank my best friend Kim for her support, friendship, and ability to distract. I would also like to thank my family for their support and patience when listening to explanations of why snails are interesting.

Most of all, I would like to thank Kevin for his patience, support, assistance, and friendship. Lastly, I would like to thank the snails for being so interesting and David Attenborough and Quantum for igniting my interest in zoology in the first place.

The fieldwork conducted in this thesis was carried out under a permit to Take, Use, Keep or interfere with Cultural or Natural Resources (ScientificPurpose) from the Queensland government (WITK12586613) and a Scientific License from the New South Wales government (SL101119). Financial support for this research was provided by the Holsworth Wildlife Research Endowment, the National Environmental Research Program (NERP), Museum Victoria (1854 Student Scholarship) and the University of Melbourne (Drummond Award, MacBain Research Scholarship). This research was undertaken with the assistance of resources from the National Computational Infrastructure (NCI), which is supported by the Australian Government, and the Victorian Life Sciences Computation Initiative (VLSCI).

Table of contents

List of Figures

List of Tables

CHAPTER 1:

General introduction

1.1 PULMONATE SYSTEMATICS

The pulmonates are a major lineage of snails and slugs within the Gastropoda, representing over 25,000 described species (Lydeard et al., 2010; Ponder and Lindberg, 2008). They are found globally (except Antarctica), in a wide range of habitats including marine, freshwater, and terrestrial environments, and are morphologically diverse, ranging from snails, to limpets and other forms with reduced shells, and slugs where the shell reduction is most advanced (Ponder and Lindberg, 2008). Many pulmonates are economically important as major agricultural pests, invasive species, and vectors of parasites. Conversely many species are highly endangered with limited distributions, and are susceptible to threats such as land clearing, pollution, and predation (Lydeard et al., 2010). The terrestrial pulmonates in particular, are potentially good indicators of conservation priorities for other organisms such as vertebrates but not vice versa (Moritz et al., 2001). There are many important evolutionary questions that can be addressed using the pulmonates as a study system. These processes include limacisation – the process of the reduction and internalization of the shell to form a slug, the evolution of carnivory, and understanding adaptations that have allowed major habitat transitions; however, a robust phylogeny is needed to investigate these processes. Despite their importance and diversity, the evolutionary relationships among the pulmonates have remained controversial despite a long history of scientific study (Ponder and Lindberg, 2008; Schrödl, 2014).

Classically, the pulmonates have been considered a monophyletic lineage within the Heterobranchia – the ‘different-gilled’ snails and slugs within Gastropoda. The Pulmonata was one of three major groups included in the Heterobranchia, the other two being the Opisthobranchia (sea slugs and related snails), and the ‘Lower Heterobranchia’ (several lineages regarded as basal or primitive) (Haszprunar, 1985). Morphological and molecular studies, however, have suggested that all three lineages are not monophyletic (Dinapoli and Klussmann-Kolb, 2010; Haszprunar, 1985; Klussmann-Kolb et al., 2008; Schrödl, 2014; Schrödl et al., 2011). A recent molecular study by Jörger et al. (2010) formally proposed the group Panpulmonata, uniting all pulmonate taxa with several lineages traditionally belonging to the Opisthobranchia and the 'Lower Heterobranchia'. There are no clear morphological synapomorphies for the Panpulmonata, although the double-rooted rhinophoral nerve has been suggested (Schrödl, 2014). While the monophyly of the Panpulmonata has been supported by subsequent molecular studies (Kocot et al., 2013b; Romero et al., 2016; Zapata et al., 2014), the relationships between the major lineages within Panpulmonata remain uncertain. Incongruence between molecular studies may be due to a lack of suitable phylogenetic loci (Ponder and Lindberg, 2008) and the use of the mitochondria, which may not be a suitable marker to address deep relationships, given potentially higher substitution rates (Romero et al., 2016; Thomaz et al., 1996). Different systematic hypotheses from these incongruent studies imply very different interpretations of morphological evolution within Panpulmonata.

A monophyletic Pulmonata within the Panpulmonata would imply that air-breathing has only evolved once. The lineages within Pulmonata were originally grouped together as they were hermaphroditic snails and slugs that did not have an operculum and breathed air through a contractile pneumostome (Cuvier, 1817; Ponder and Lindberg, 2008). The definition of Pulmonata was later expanded to comprise all air-breathing Heterobranchia including terrestrial, freshwater, and intertidal lineages. Accordingly, pulmonate lineages include the Stylommatophora (terrestrial snails and slugs), the Systellommatophora (mostly intertidal and terrestrial slugs), and the Basommatophora, which comprise the Hygrophila (freshwater snails), the Siphonariidae (intertidal false limpets), and Amphiboloidea (intertidal and estuarine snails) (Hubendick, 1979; Solem, 1979). A number of studies based on small sets of nuclear genes or the mitochondria have shown Pulmonata to be polyphyletic (Dayrat et al., 2011; Dinapoli and Klussmann-Kolb, 2010; Holznagel et al., 2010; Jörger et al., 2010; Klussmann-Kolb et al., 2008; Romero et al., 2016). The two phylogenomic studies to address these relationships to date, however, could not reject a monophyletic Pulmonata due to a lack of resolution (Zapata et al., 2014) or insufficient taxonomic sampling (Kocot et al., 2013b).

Several studies have suggested that the Siphonariidae (marine false limpets) – traditionally classified as basommatophoran – and the Sacoglossa (sap-sucking sea slugs) – traditionally classified as opisthobranchs, are the basal lineages within Panpulmonata (Dinapoli and Klussmann-Kolb, 2010; Jörger et al., 2010; Klussmann-Kolb et al., 2008; Kocot et al., 2013b; White et al., 2011). While most studies are unresolved, there is molecular support for two alternative topologies. A sister relationship between the Sacoglossa and the Siphonariidae, termed Siphoglossa (Medina et al., 2011), is supported by molecular analysis based on 18S rRNA and 28S rRNA and the mitochondrial 16S rRNA and COI (Klussmann-Kolb et al., 2008). If correct, the Siphoglossa relationship implies that Pulmonata is not monophyletic and that the narrowed opening (pneumostome) of the modified pallial cavity (‘lung’) in the Siphonariidae evolved independent of the pneumostome in the rest of the pulmonates. Both the Siphonariidae and Sacoglossa share an opisthobranch-type gill, but the homology of this organ has been questioned (Jensen, 2011). The alternative placement of the Sacoglossa is as the basal lineage within Panpulmonata. This relationship has support from a phylogenomic analysis, albeit with limited taxonomic sampling, and is not inconsistent with a monophyletic Pulmonata (Kocot et al., 2013b).

There is also evidence to suggest that the Amphiboloidea, another traditionally basommatophoran lineage, is more closely related to non-air-breathing lineages. The Amphiboloidea inhabit mudflats, saltmarshes, and mangroves, and were included in the Pulmonata because they breathe air through a narrowed pneumostome (Golding et al., 2010; Solem and Yochelson, 1979). Recent molecular studies, however, have suggested that the Amphiboloidea are more closely related to the non-air-breathing, freshwater Glacidorbidae and the marine Pyramidellidae (Dinapoli and Klussmann-Kolb, 2010; Jörger et al., 2010). When first described, the Glacidorbidae were included in the Pulmonata based on morphological characters including features of the reproductive and nervous systems (Ponder, 1986). The Glacidorbidae, however, have also been considered as belonging to the ‘Lower Heterobranchia’ due to a lack of a pneumostome and a lack of several features of the nervous system typical of the pulmonates (Haszprunar and Huber, 1990; Haszprunar, 1988). The Pyramidellidae are minute marine snails that are ectoparasites of marine invertebrates including other molluscs and annelid worms (Dinapoli et al., 2011). The Pyramidellidae have been included in the ‘Lower Heterobranchia’ since the taxon Heterobranchia was erected (Haszprunar, 1985). Only two molecular studies (Dinapoli and Klussmann-Kolb, 2010; Jörger et al., 2010), based on small sets of nuclear and mitochondrial loci, show support for a monophyletic clade comprising the Amphiboloidea, Glacidorbidae, and Pyramidellidae, and this relationship has not been assessed with a phylogenomic dataset. These three families are the only three panpulmonate lineages to retain the ancestral operculate as adults. However, support for this clade would also imply that air-breathing evolved independently multiple times across the pulmonates.

Several traditionally basommatophoran lineages, including the Ellobiidae, were grouped with the Stylommatophora by Haszprunar and Huber (1990) to form the

Eupulmonata. The Eupulmonata (sensu Bouchet and Rocroi, 2005) share characteristics of the nervous system and all breathe through a contractile pneumostome (Haszprunar and Huber, 1990). While the monophyly of Eupulmonata has been supported by a number of molecular studies (Dinapoli and Klussmann-Kolb, 2010; Jörger et al., 2010; Klussmann-Kolb et al., 2008), the two most recent studies to address these relationships rejected a monophyletic Eupulmonata (Dayrat et al., 2011; Romero et al., 2016). A non-monophyletic Eupulmonata would imply that the morphological characters that unite the eupulmonates evolved independently (or that multiple reversals have occurred). Dayrat et al. (2011) suggested that the freshwater non-air-breathing Glacidorbidae had a sister relationship with terrestrial Stylommatophora, whereas Romero et al. (2016) showed a sister relationship between the Stylommatophora and the pulmonate freshwater Hygrophila. These alternative hypotheses imply very different scenarios for the transitions to freshwater and terrestrial habitats in the pulmonates. An additional alternative hypothesis, termed the Geophila by Férussac (1819), suggests a sister relationship between the eupulmonate Systellommatophora and Stylommatophora. A number of morphological studies have supported Geophila (Barker, 2001; Dayrat and Tillier, 2002) but to date no molecular phylogeny has supported this clade.

1.2 Stylommatophoran Systematics

The Stylommatophora are the most speciose lineage within the pulmonates and the most successful molluscan lineage on land. The Stylommatophora share the ability to retract and invaginate (rather than simply contract) the cephalic tentacles, the presence of a membrane covering the pedal gland, and the acquisition of a secondary ureter that aids water retention (Little, 1983). While the monophyly of the Stylommatophora has been supported by molecular evidence (Tillier et al. 1996; Wade et al. 2001, 2006), the relationships within the clade remain largely unresolved (Tillier et al. 1996; Wade et al. 2001, 2006). The Stylommatophora were originally divided into four separate groups based on the structure of the excretory system – the Sigmurethra, the Mesurethra, Heterurethra, and the Orthurethra (Baker, 1955; Pilsbry, 1900). Only the Heterurethra (often termed Elasmognatha) and the Orthurethra are supported by molecular evidence, but only represent a small proportion of the stylommatophoran families (Wade et al., 2006, 2001). The only major relationship within the Stylommatophora that has received strong support in detailed phylogenetic analyses is the primary split between the achatinoid (named for the Achatinidae and related families) and non-achantinoid lineages (Wade et al., 2006, 2001). Morphological analyses have suggested that the Elasmognatha, a stylommatophoran lineage comprising the triangle and leaf-vein slugs (Athoracophoridae) and the amber snails (Succineidae), are the basal stylommatophoran lineage (Barker, 2001); however, this relationship has not been supported in molecular analysis (Wade et al., 2001; 2006).

Given the fossil record, it has been suggested that the lack of resolution may be due to a relatively rapid diversification event within the Stylommatophora (Tillier et al., 1996). Fossils from the Carboniferous, for genera such as *Dendropupa*, were once regarded as stylommatophorans (Solem and Yochelson, 1979), suggesting that the pulmonates transitioned to land not long after the arthropods (Engel and Grimaldi, 2004) and the tetrapods (Ahlberg and Milner, 1994). Most of the terrestrial carboniferous fossils, however, have since been assigned to the Caenogastropoda (Bouchet and Rocroi, 2005; Gordon and Olson, 1995). The earliest unambiguous fossil record for the Stylommatophora is from the late Cretaceous (~85 Ma) (Dayrat et al., 2011). A better understanding of the phylogeny and timing of evolution is needed to investigate whether rapid diversifications have occurred in the Stylommatophora and to place such events in the context of the fossil record. No molecular dating analysis has been performed for the Stylommatophora.

1.3 RHYTIDIDAE SYSTEMATICS

The Stylommatophora contains over 100 families (Bouchet and Rocroi, 2005). The evolutionary relationships between and within many of these families are still poorly understood (Wade et al., 2006), and in many cases the species level taxonomy is yet to be assessed with molecular evidence. One such example is the Rhytididae, a family of carnivorous land snail (Stylommatophora) with a Gondwanan distribution: they are found in South Africa, Australia, New Zealand, Papua New Guinea, and some Pacific islands, however, the centre of taxonomic diversity is Australia. Based on shell morphology, including shell sculpture and size – the Australian Rhytididae range from minute (2mm) to large (45mm) – Solem (1959) suggested that the Australian rhytidids formed several major groups that included both large and small Rhytididae from Australia, New Zealand, and the Pacific islands; although he noted that these groups may not represent phylogeny. The most recent study to address the species level taxonomy of the Australian Rhytididae (Stanisic et al., 2010) described 60 new species and 15 new genera based on shell morphology. Phylogenetic studies based on mitochondrial markers and/or the nuclear gene 28S have greatly advanced and revised the taxonomy of the major South African (Moussalli and

Herbert, 2016; Moussalli et al., 2009) and New Zealand (Efford et al., 2002; Spencer et al., 2006) lineages but no study has examined the phylogenetic relationships of the Australian Rhytididae in any detail.

1.4 PHYLOGENOMICS

Most molecular phylogenetic studies in non-model systems have had to rely on a limited number of readily sequenced genes due to the effort and cost restrictions of Sanger sequencing and the availability of suitable phylogenetic markers. Both theoretical and empirical studies, however, have shown that a greater number of independently evolving loci are often needed to resolve difficult phylogenetic questions (Gontcharov et al., 2004; Leaché and Rannala, 2011; Wortley et al., 2005). This need has been addressed by the advent of high-throughput sequencing technologies that, in conjunction with developments in bioinformatics, have made the acquisition of large phylogenomic datasets possible. High-throughput sequencing is used to sequence genomes, however, these are still relatively expensive and usually not necessary for phylogenetics. Instead, reduced representation methods such as RNAseq (i.e. sequencing of the transcriptome) and exon capture, offer a cost efficient method for producing large phylogenomic datasets for large numbers of samples in non-model systems.

Transcriptome datasets are increasingly being used in phylogenomic analyses for non-model taxa (e.g. Kocot et al., 2013b, 2011; Misof et al., 2014; O’Hara et al., 2014; Oakley et al., 2012; Zapata et al., 2014). A transcriptome represents the total pool of mRNA in a tissue. Typically, sequence information for over 10,000 genes can be obtained, despite only sequencing a few percent of the whole genome, making transcriptome sequencing highly cost effective. Additionally, the development of transcriptome-specific assembly algorithms, such as Trinity (Haas et al., 2013), means analyses of transcriptomes do not require a reference genome. However, mRNA can only be sequenced from samples where the RNA is preserved (e.g. using RNAlater or liquid nitrogen).

Targeted enrichment techniques, including exon capture, are an alternative form of reduced representation whereby a specific set of genes are targeted and sequenced from genomic DNA rather than the RNA. Specific genes are selectively sequenced by designing probes for the respective genes from a reference set of genomes or transcriptomes. Such targeted enrichment techniques allow the sequencing of ethanol preserved specimens such as those preserved in museum collections. A number of studies have used targeted enrichment to investigate deep relationships across major lineages (e.g. Hugall et al., 2016; Lemmon et al., 2012; McCormack et al., 2013) and shallower relationships within families and genera (e.g. Bi et al., 2012; Bragg et al., 2016; Eytan et al., 2015). Most targeted enrichment protocols, however, can only tolerate up to ~12% sequence divergence (e.g. Hugall et al., 2016). To successfully capture sequence across highly divergent lineages some studies have targeted highly conserved regions of the genome (e.g. UCEs - Faircloth et al., 2012; anchored enrichment - Lemmon et al., 2012). However, to target large sets of exons that are not highly conserved, access to genomic-scale resources for the clade of interest is needed.

Despite the pulmonates being a highly speciose, major gastropod lineage, the genomic resources available to address the molecular systematics of this group are relatively limited. While genome projects for a number of gastropods are currently in progress, few of these datasets are currently publicly available or well assembled and annotated. At the time this research was performed, the highest quality molluscan genome, in terms of assembly and annotation, was that of the owl limpet (*Lottia gigantea*; Simakov et al., 2013), a basal gastropod divergent from the eupulmonates (Kocot et al., 2011). Data for the California sea hare genome, *Aplysia californica*, is available; however, the gene models are not as well annotated as in the *L. gigantea* genome. Transcriptome datasets for the pulmonates are steadily becoming available; typically as studies on the transcriptomes of individual species (e.g. Feldmeyer et al., 2011; Sadamoto et al., 2012; Wägele et al., 2011) or as part of phylogenomic studies conducted for the Gastropoda (Zapata et al., 2014) or the Mollusca as a whole (Kocot et al., 2011; Smith et al., 2011). These phylogenomic datasets for broader Molluscan groups are potentially an important starting point for identifying genes appropriate for phylogenetics in the pulmonates (e.g. Kocot et al., 2013b). The eupulmonates were a key gap in pulmonate genomic resources at the time this research was conducted.

1.5 ORTHOLOGY

While next generation sequencing allows access to the sequences of potentially thousands of genes, not all genes are suitable for phylogenetics. It is relatively straight forward to determine whether genetic sequences are homologous – generally through alignment using algorithms such as BLAST (e.g. Camacho et al., 2009) – however, sequences also need to be orthologous to be useful for phylogenetic analysis. Two sequences are orthologous if they diverged through speciation rather than duplication within the genome (Fitch, 1970) (Figure 1.1). If a duplication event occurs in the genome, through polyploidy or

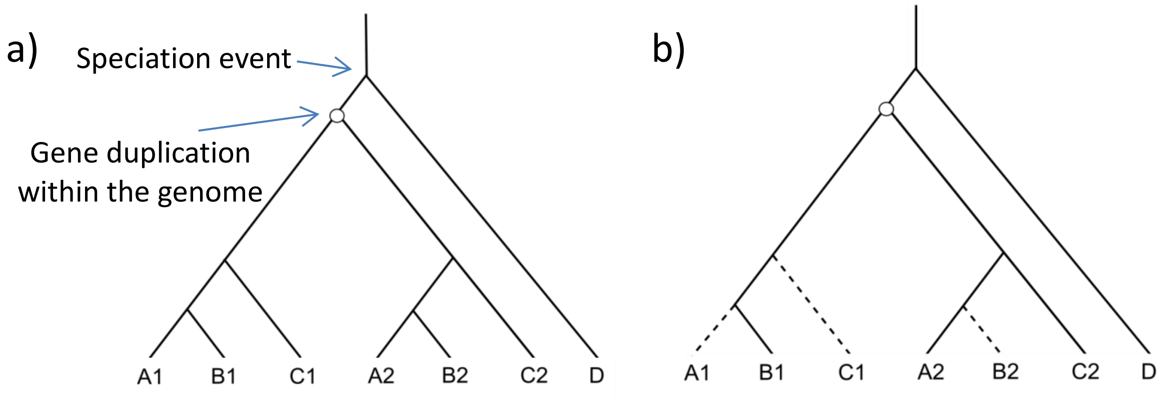


Figure 1.1. The consequences of using non-orthologous sequences for phylogenetic reconstruction. a) shows the evolutionary relationships between four species (A, B, C and D) for a gene that underwent a duplication in the genome of the common ancestor of species A, B, and C. There are now two copies of the gene in species A, B, and C. If some of the subsequent gene copies are subsequently lost, as in b), phylogenetic reconstructions show A and C are sister species when in truth they are not.

partial duplication, there will subsequently be two copies of certain genes within the genome (Figure 1.1). The two (or more) copies, referred to as paralogs, in most cases will subsequently undergo independent evolution (Fitch, 2000). Paralogous sequences can mislead phylogenetics as the divergence between the two sequences will not reflect subsequent speciation (Fitch, 2000). Paralogs are especially misleading and hard to recognise if one of the copies is missing, either because it was subsequently lost from the genome or because it was not sequenced (see Figure 1.1). Several studies have shown that even a small number of paralogs in phylogenomic datasets can have a significant impact on biological interpretation (Dávalos et al., 2012; Qiu et al., 2012; Struck, 2013). Most phylogenomic studies therefore perform analyses to select orthologous genes before phylogenetic analyses.

There are a number of different approaches to determine whether sequences are orthologous. Once homologous sequences across the samples of interest are identified, orthology is qualified using similarity based approaches, including best-hit reciprocal blasts (Ebersberger et al., 2009; Ward and Moreno-Hagelsieb, 2014; Waterhouse et al., 2013), and/or tree based methods, where gene trees are used to identify sequences with purely orthologous relationships (e.g., Agalma, Dunn et al., 2013; PhyloTreePruner, Kocot et al., 2013a; TreSpEx, Struck, 2014). Despite rapid advances in automated approaches to homolog identification and qualifying orthology, there are many characteristics of transcriptome assemblies that challenge such automated methods. These include frameshifts, mis-indexing, transcript fragmentation, and the presence of multiple isoforms. Not accounting for these issues can lead to erroneous inclusion of paralogous sequences and/or the inadvertent removal of appropriate orthologous sequences (Martin and Burg, 2002; Philippe et al., 2011; Pirie et al., 2007).

1.6 THESIS OUTLINE

The overall aim of my thesis is to use phylogenomic datasets to address fundamental evolutionary questions in the pulmonates. My thesis comprises three data chapters, each structured as independent manuscripts; therefore there is some repetition among chapters. In **Chapter 2**, "*Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon capture*", I sequenced 21 transcriptomes and screened for orthologous genes appropriate for phylogenetics across the Eupulmonata. I compared manual and automated approaches to orthology determination and further qualified the orthologous genes by sequencing them from 22 Australian representatives of the land snail family Camaenidae using exon capture. In the subsequent chapters of my thesis I use the genes identified in Chapter 2 to investigate pulmonate relationships at two scales.

In **Chapter 3**, "*Pattern and pace of pulmonate evolution*", I investigate the timing and pattern of diversification within Panpulmonata with a particularly focus on the Stylommatophora. Specifically I test whether Pulmonata forms a monophyletic clade within Panpulmonata, to determine whether air-breathing has evolved more than once, and provide a fossil calibrated phylogeny of Panpulmonata. In **Chapter 4**, "*Phylogenetic investigation of the Australian Rhytididae using exon capture*", I designed an exon capture probe set for the Australian Rhytididae, a family of carnivorous land snails. Using this dataset I address the phylogenetic relationships of this group in a biogeographic context and test the validity of the current taxonomy. In **Chapter 5**, "*General discussion*" I provide a synthesis of the major findings of my thesis and highlight areas for future research.

CHAPTER 2:

Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon capture

2.1 ABSTRACT

The qualification of orthology is a significant challenge when developing large, multi-loci phylogenetic datasets from assembled transcripts. Transcriptome assemblies have various attributes, such as fragmentation, frameshifts, and mis-indexing, which pose problems to automated methods of orthology assessment. Here, I identify a set of orthologous single-copy genes from transcriptome assemblies for the land snails and slugs (Eupulmonata) using a thorough approach to orthology determination involving manual alignment curation, gene tree assessment and sequencing from genomic DNA. I qualified the orthology of 500 nuclear, protein coding genes from the transcriptome assemblies of 21 eupulmonate species to produce the most complete gene data matrix for a major molluscan lineage to date, both in terms of taxon and character completeness. Exon capture targeting 490 of the 500 genes (those with at least one exon > 120 bp) from 22 species of Australian Camaenidae successfully captured sequences of 2,825 exons (representing all targeted genes), with only a 3.7% reduction in the data matrix due to the presence of putative paralogs or pseudogenes. The automated pipeline Agalma retrieved the majority of the manually qualified 500 single-copy gene set and identified a further 375 putative single-copy genes, although it failed to account for fragmented transcripts resulting in lower data matrix completeness. This could potentially explain the minor inconsistencies I observed in the supported topologies for the 21 eupulmonate species between the manually curated and Agalma-equivalent dataset (sharing 458 genes). Overall, my study confirms the utility of the 500 gene set to resolve phylogenetic relationships at a broad range of evolutionary depths, and highlights the importance of addressing fragmentation at the homolog alignment stage for probe design.

2.2 INTRODUCTION

Robust and well resolved phylogenies document the evolutionary history of organisms and are essential for understanding spatio-temporal patterns of phylogenetic diversification and phenotypic evolution. Despite the central role of phylogenies in evolutionary biology, most phylogenetic studies in non-model systems have relied on a limited number of readily sequenced genes due to cost restrictions and availability of phylogenetic markers. However, both theoretical and empirical studies have shown that a greater number of independently evolving loci are needed to resolve difficult phylogenetic questions (Gontcharov et al., 2004; Leaché and Rannala, 2011; Wortley et al., 2005). This need has been addressed by rapid advances in phylogenomics, which capitalise on high-throughput sequencing to acquire large multi-loci datasets. In particular, both transcriptome sequencing and targeted-enrichment strategies are increasingly being employed to reconstruct phylogenetic relationships across a wide range of taxonomic levels (e.g. Bi et al., 2012; Faircloth et al., 2012; Lemmon et al., 2012; Misof et al., 2014; O'Hara et al., 2014; Zapata et al., 2014). A common aim of these studies, especially targeted enrichment based studies, has been to identify universal sets of orthologous loci that can readily be captured and sequenced across a broad taxonomic spectrum (Faircloth et al., 2012; Hugall et al., 2016; e.g. Lemmon et al., 2012). Obtaining such universal sets of orthologous genes allows for consistency and comparison across studies, and ultimately contributes towards a more comprehensive Tree of Life (ToL) meta-analysis.

One of the greatest challenges associated with developing large, multi-loci phylogenomic datasets is the qualification of orthology. In the context of phylogenetic analysis, genes need to be orthologous and single-copy across all taxa under study (Fitch, 2000; Philippe et al., 2011; Struck, 2013). To this end, a number of automated pipelines have been developed to identify single-copy orthologous genes from assembled transcriptomes. These methods generally involve two main steps. The first step is to identify and cluster homologous sequences, either by direct reference to annotated genomes (e.g., O'Hara et al., 2014) or by reference to ortholog databases, which themselves are derived from genome comparisons (Altenhoff et al., 2015; Ranwez et al., 2007; e.g., Tatusov et al., 2003; Waterhouse et al., 2013). Alternatively, non-reference methods have been employed such as all-by-all and reciprocal BLAST comparisons (Dunn et al., 2013; Li et al., 2003) followed by clustering (Enright et al., 2002). In the second step, orthology is qualified using either similarity based approaches, including best-hit reciprocal blasts (Ebersberger et al., 2009; Ward and Moreno-Hagelsieb, 2014; Waterhouse et al., 2013), and/or tree based methods, where gene trees are used to identify sequences with purely orthologous relationships (e.g., Agalma, Dunn et al., 2013; PhyloTreePruner, Kocot et al., 2013a; TreSpEx, Struck, 2014).

Despite rapid advances in automated approaches to homolog clustering and qualifying orthology, there are many characteristics of transcriptome assemblies that challenge such automated methods. These include frameshifts, mis-indexing, transcript fragmentation and the presence of multiple isoforms. Not accounting for these issues can lead to erroneous inclusion of paralogous sequences and/or the inadvertent removal of appropriate orthologous sequences (Martin and Burg, 2002; Philippe et al., 2011; Pirie et al., 2007). To address these issues O'Hara *et al.* (2014) placed greater emphasis on careful manual curation and editing of homolog alignments prior to orthology qualification. A key aspect of this approach was the concatenation of transcript fragments into a single consensus sequence prior to tree-based ortholog qualification, leading to a more complete final data matrix. This, in turn, allowed a more robust probe design for subsequent exon capture (Hugall et al., 2016). With the same objective of deriving a gene set appropriate for exon capture in future studies, here I implement this approach to identify and qualify 500 single-copy orthologous genes for the Eupulmonata, a major lineage of air breathing snails and slugs within the class Gastropoda.

Eupulmonata comprises over 20,000 species, with an evolutionary depth spanning over 150 million years (Jörger et al., 2010; Lydeard et al., 2010). The evolutionary relationships of the Eupulmonata, however, remain incompletely understood despite many morphological and molecular phylogenetic studies over the last two decades (Dayrat et al., 2011; Dinapoli and Klussmann-Kolb, 2010; Grande et al., 2004; Holznagel et al., 2010; e.g., Ponder and Lindberg, 1997; Wade et al., 2006, 2001). The lack of congruence between studies is largely due to a combination of using insufficient genetic markers (Schrödl, 2014), with many studies relying on 28S rRNA or mitochondrial sequences, and widespread morphological convergence (Dayrat and Tillier, 2002). Therefore to resolve the 'tree of life' of the eupulmonates, it is essential to identify more independently evolving markers, with a greater range of substitution rates, to better estimate relationships across all evolutionary depths. To achieve this, I sequenced and assembled transcriptomes for representatives of 15 families across Eupulmonata. I used the owl limpet genome, *Lottia gigantea*, as a reference to identify and cluster homologous sequences and visually assessed and manually edited candidate homolog alignments accounting for transcript fragmentation, mis-indexing and frameshifts. I then further qualified orthology by assessing individual gene trees and by sequencing the orthologous gene set from genomic DNA using exon capture as unexpressed paralogs or pseudogenes will not be detected in transcriptome datasets. Lastly, as a comparison and qualification of my approach I also analysed my transcriptome dataset using the fully automated orthology determination pipeline Agalma (Dunn *et al.* 2013).

2.3 MATERIALS AND METHODS

2.3.1 Transcriptome sequencing and assembly

I sequenced transcriptomes for 21 species of terrestrial snails and slugs representative of 15 families across Eupulmonata (Table 2.1). Total RNA was extracted from foot or whole body tissue stored in RNAlater (Ambion Inc, USA) using the Qiagen RNeasy extraction kit (Qiagen, Hilden, Germany). Library preparations were conducted using the TruSeq RNA sample preparation kit v2 (Illumina Inc., San Diego, CA), and sequenced on the Illumina HiSeq 2000 platform (100 bp paired end reads). I used the program Trimmomatic v0.22 (Lohse et al., 2012) to remove and trim low quality reads and adaptor sequences, and the program Trinity v2012-06-08 (Grabherr et al., 2011; Haas et al., 2013) with default settings to assemble the transcriptomes.

2.3.2 Homolog clustering

Our approach to homolog clustering and orthology qualification is largely consistent with that detailed in O’Hara *et al.* (2014). A schematic representation of my pipeline is provided in Figure 2.1. First, to generate clusters of putatively homologous sequences I compared each assembly to the *Lottia gigantea* predicted gene dataset (hereon referred to as the *L. gigantea* genes). The *L. gigantea* reference represents 23,851 filtered gene models annotated in the most current draft genome (Grigoriev et al., 2012; Simakov et al., 2013). Each transcriptome assembly was compared against the *L. gigantea* genes using blastx with an e-value cut off of 1e-10. This is a relatively relaxed threshold given the small size of the *L. gigantea* reference set. A relaxed e-value cutoff was used to ensure all closely related homologs were assessed without allowing through too many spurious matches with non-homologous sequences. We retained only the top hit for each assembled contig (i.e. the match with the lowest e-value).

In addition to identifying homologous contigs from each transcriptome assembly, we also identified putative paralogs within the *L. gigantea* genome itself, in order to aid the identification of paralogous sequences within the eupulmonates. I ran an all-by-all BLAST of the *L. gigantea* genes against themselves (blastp, cut off e-value of 1e-10), retaining all hits to identify *L. gigantea* genes which had hits to *L. gigantea* genes other than themselves. To qualify the all-by-all BLAST results, I also obtained orthology status for all *L. gigantea* genes classified in the Orthologous MAtrix (OMA) ortholog database (Altenhoff et al., 2015). A *L. gigantea* gene was considered to be single-copy if it was the only *L. gigantea* sequence in its respective OMA group. While this information provided guidance, I was not reliant on the *L. gigantea* orthology status when prioritising homolog clusters to assess (see below for criteria used). I considered *L. gigantea* to be sufficiently divergent from the eupulmonates (> 400 million years, Zapata *et al.* 2014) that single-copy status could differ.

The BLAST results for both the transcriptomes compared to *L. gigantea* and the *L. gigantea* all-by-all BLAST were used to produce clusters of homologous sequences linked by having a match to a specific *L. gigantea* gene. Hence, a homolog cluster represents 1) all contigs from all species transcriptomes that had a BLAST match to a given reference *L. gigantea* gene (there were often multiple contigs per taxon with hits to a given *L. gigantea* gene), and 2) all contigs having a hit to any of the closely related *L. gigantea* genes identified by the all-by-all BLAST.

2.3.3 Orthology assessment

After constructing the homologous clusters, I first visually assessed the alignments for evidence of paralogy. Sequences for each cluster were placed into the correct reading frame using coordinates output from the Blastx comparison for each transcriptome against *L. gigantea*, and were then translated and aligned in amino acids using ClustalW (Thompson et al., 1994) within the program BioEdit (Hall, 1999). I only considered the coding region (i.e. untranslated regions (UTRs) were removed) which was identified manually by reference to the *L. gigantea* protein sequence for the relevant gene, which was included in the alignments. Many of the homolog clusters contained multiple fragmented transcripts for a given species that were shorter than the coding region but which often overlapped. These fragmented transcripts were synthesised into consensus sequences by manual manipulation within BioEdit, if the overlapping regions did not differ by more than three nucleotides. Non-overlapping fragments were also concatenated if there were no competing contigs covering the same region of the alignment and both sequences displayed a high degree of similarity to non-fragmented sequences in closely related taxa.

By visually assessing each homolog alignment in both amino acid and nucleotides (in Bioedit it is straight forward to toggle between the two), I was able to identify and manually correct frameshifts. These were clearly evident as a large proportion of a contig would not align with the rest of the sequences and the site of the frameshift was usually associated with runs of adenines. I also manually edited the alignments to remove clearly erroneous sequences which could not be aligned, clear out-paralogs (i.e. sequences which are paralogous but the duplication event took place before the common ancestor of eupulmonates) and redundant sequences (identical transcripts within a species). Mis-indexing was identified as cases where, within the one assembly, two contigs were present for the same region but one (typically the shorter contig having low coverage) matched the sequence for another taxon exactly. Taxa containing paralogs were clearly evident in the alignments as they frequently had > 5% dissimilarity at the nucleotide level between overlapping contigs within the one sample. To further qualify that these sequences were paralogs I inspected genealogies constructed using the neighbour joining method in MEGA (see Appendix 2.1). Any homolog cluster containing paralogs for any species was excluded from further consideration. In certain cases paralogous sequences were closely related (3-5% dissimilarity), representing either in-paralogs (see Remm *et al.* 2001) or genes exhibiting elevated allelic diversity (see O'Hara et al., 2014). These genes were also excluded from further consideration as such genes are not optimal for exon capture.

Approximately 1,500 homologous clusters were visually assessed in order to find 500 which were orthologous across all 21 taxa assessed. This dataset size was chosen to represent a balance between phylogenetic power at varying time scales (Leaché and Rannala, 2011; Lemmon and Lemmon, 2013; Philippe et al., 2011) and a suitable size for subsequent exon capture probe design. To maintain consistency across studies, I first assessed homolog alignments corresponding to the 288 *L. gigantea* genes used in a phylogenomic study of the Mollusca (Kocot et al., 2011). Although there are two other published molluscan phylogenomic datasets (Smith et al., 2011; Zapata et al., 2014), I focussed on the final dataset of Kocot *et al.* (2011) as the *L. gigantea* gene IDs were documented in the supplementary they provided, which in turn allowed us to easily identify and assess these genes given my pipeline was based on the same reference. I then proceeded to assess and qualify additional homolog clusters until I obtained a final set of 500 single-copy orthologous genes. Accordingly, I prioritised homolog clusters with high taxonomic representation (≥ 18 taxa), as completeness of the data matrix is critical for designing probes across multiple lineages (Hugall et al., 2016; Lemmon et al., 2012). Where possible I also prioritised homolog alignments for which the corresponding *L. gigantea* gene had a coding region (CDS) ≥ 300 bp or had at least one exon ≥ 200 bp.

As a proxy measure of substitution rate variation across the final 500 gene set, I calculated uncorrected distances (p-distance) for species pairs within the families Rhytididae (*Terrycarlessia turbinata* and *Victaphanta atramentaria*) and Camaenidae (*Sphaerospira fraseri* and *Austrochloritis kosciuszkoensis*). I chose to limit this analysis to intrafamilial comparisons to avoid underestimation due to saturation. For comparison, I also calculated the p-distances for two commonly used phylogenetic markers, CO1 and 28S, for the same taxa.

2.3.4 Qualification of orthology using gene tree assessments

Although only a single copy of each gene per taxon was present in my final ortholog alignments, they may nevertheless be paralogous across taxa (see Struck 2014). To investigate ‘hidden paralogy’ I used the program TreSpEx (Struck, 2014) to assess genealogies for conflict with *a priori* taxonomic hypotheses. Gene trees for each of the 500 genes were constructed using the GTRGAMMA model, codon specific partitioning, and 100 fast bootstraps in RAxML (Stamatakis, 2006). TreSpEx then identified well supported conflicting phylogenetic signal relative to five distinct and taxonomically well-established eupulmonate clades (Limacoidea, Orthurethra, Helicoidea, the Australian Rhytididae (Table 2.1: see Hausdorf, 1998; Herbert et al., 2015; Wade et al., 2007, 2006), and the Stylommatophora). All nodes with ≥ 75 bootstrap support were first assessed for conflict with the monophyly of each of the five clades. Strongly supported sister relationships between sequences from different clades can indicate the presence of ‘hidden’ paralogous sequences. TreSpEx flags very short terminal branches (parameter blt set to 0.00001) as indicative of potential cross-contamination and internal branches which are five times greater than the average (parameter lowbl set to 5), which, in addition to strong nodal support, may indicate paralogy.

2.3.5 Qualification of orthology using exon capture

To further qualify orthology and identify unexpressed paralogs and pseudogenes, I designed an exon capture probe set to enrich and sequence exons from my 500 gene dataset. As the divergence across the eupulmonates is too large for a single probe design I designed a probe set for the Australian Camaenidae as a test case. It would be feasible, however, to design a probe set from my alignments for any of the taxa I have assessed in this study. I designed the baits based on two species of Australian Camaenidae, *Sphaerospira fraseri* and *Austrocholritis kosciuszkoensis*, which represent two divergent lineages of the Australian camaenids (Hugall and Stanisic, 2011). Specifically, I included sequences from both taxa for each gene in the probe design. The divergence between these taxa ranges up to ~12% (Figure 2.3) which is about the level of divergence tolerated be the probes (Hugall et al., 2016). Including both taxa in the design increases the likelihood that we will capture sequences from more divergent lineages within the Camaenidae for which we don't yet have transcriptome sequences. Exon boundaries were first delineated using the program Exonerate v2.2.0 (Slater and Birney, 2005) in reference to the *L. gigantea* genomic sequences and then manually qualified using the boundaries detailed in the *L. gigantea* genome annotation (JGI, Grigoriev et al., 2012). All exons shorter than 120bp (the probe length) were excluded. This resulted in a target consisting of 1,646 exons from 490 of the 500 genes (ten genes contained only exons shorter than 120bp and were excluded from the bait design). Probes for the target sequences were designed and produced by MYcroarray (Ann Arbor, Michigan) using MYbaits custom biotinylated 120bp RNA baits at 2X tiling.

I tested the probe set on 22 camaenid species spanning much of the phylogenetic breadth of the Australasian camaenid radiation, representing up to 30 million years (My) of evolution (Hugall and Stanisic, 2011) (Table 2.2). DNA was extracted using the DNeasy blood and tissue kit (Qiagen) and sheared using the Covaris S2 (targeting a fragment size of 275bp). Libraries where then constructed using the Kapa DNA Library Preparation Kit (Kapa Biosystems, USA), modified to accommodate dual-indexing using the i7 and i5 index sets (see Hugall et al., 2016). Up to eight libraries (normalised to 100 ng each) were pooled per capture, and hybridised to the baits (at one-quarter dilution) for 36 hours, following the MYbait protocol v1. A second hybridisation was then carried out on the fragments retained from the first hybridisation to further enrich the capture. Several captures were then multiplexed and sequenced on the Illumina MiSeq platform (v2), obtaining 150bp paired-end reads.

I used FastUniq v1.1 (Xu et al., 2012) to remove duplicates, and Trimmomatic v0.22 (Bolger et al., 2014) to trim and remove low quality reads and adaptor sequences (minimum average quality score threshold of 20 per 8 bp window). Reads shorter than 40 bases after trimming were discarded. The trimmed reads were then mapped onto the transcriptome sequences used for the probe design using BFAST v0.7.0a (Homer et al., 2009) with a single index of 22 bp without mismatch. After creating pileup files using Samtools v0.1.19 (Li et al. 2009), VarScan v2.3.7 (Koboldt et al., 2012) was used to call variants and produce a final consensus sequence for each taxon per exon. Viewing the initial BAM alignments showed that exon boundaries were often not conserved between *L. gigantea* and the Camaenidae. In these cases (Appendix 2.5) the reference exons were split to reflect the actual exon boundaries in the Camaenidae. The reads where then mapped to the revised exon reference and consensus sequences made as outlined above. To flag potential pseudogenes and paralogs I identified consensus sequences with an elevated proportion of variable sites (> 3% heterozygote sites) and reviewed the corresponding read alignments (BAM files) using the Integrative Genomics Viewer (IGV: Thorvaldsdóttir et al., 2013). All sequences with greater than 3% ambiguous sites where removed from the final dataset. Exons where more than 10% of the taxa contained greater than 3% ambiguous sites were discarded entirely.

I again used TreSpEx to assess conflicting phylogenetic signal. I screened for hidden paralogs based on five *a priori* phylogenetic hypotheses representing well supported clades (≥75% bootstrap support) within the Australasian camaenid radiation as delineated by Hugall and Stanisic (2011), namely the Hadroid group (clade 1 – 4 inclusive), the far-northern (sister clades 5 and 6) and north-eastern (clade 7) Chloritid groups, a group dominated by arid and monsoonal camaenids (clade 11) previously recognised as the subfamily Sinumeloninae (e.g. Solem, 1992), and a phenotypically and ecologically diverse group dominated by eastern Australian wet forest taxa (sister clades 8 and 9). Gene trees for each of the 490 genes (exons from the same gene were combined as one partition) were constructed using the GTRGAMMA model and 100 fast bootstraps in RAxML (Stamatakis, 2006). TreSpEx was run using the same settings as the analysis for the transcriptome dataset (i.e. TreSpEx considered nodes for strong conflict, long branches, and short branches in that order with parameters upbl and lowbl set to 5 and blt 0.00001).

2.3.6 Comparison to the Agalma pipeline

As an independent qualification of the manually curated 500 gene set I ran the fully automated orthology determination pipeline Agalma (Dunn et al., 2013) (Figure 2.1). I commenced this pipeline from the ‘postassemble’ step which first identified open reading frames and putative coding regions (Dunn et al., 2013). Homolog clusters were then identified using an all-by-all tblastx, followed by clustering using the Markov Clustering algorithm (MCL) (Figure 2.1). Homolog clusters were then translated and aligned using MAFFT (Katoh and Standley, 2013) and gene trees estimated using RAxML. To identify orthologous sequences, the genealogies were then screened for ‘optimally inclusive subtrees’ which contain only a single representative of each species. Multiple orthologous subtrees can be delineated per homolog cluster, potentially allowing paralogs to be separated and retained.

The surviving subtrees were filtered based on the number of taxa (set to greater than four taxa) and realigned for subsequent phylogenetic analysis. I then identified Agalma homologous clusters that corresponded to the manually curated 500 gene set using BLAST (blastp, e-value cut off of 1e-10).

2.3.7 Phylogenetic analysis

After removal of paralogs or sequences with excessive polymorphism (>3% dissimilarity), my phylogenomic datasets were refined by removing any regions of ambiguous alignment through the use of Gblocks (Castresana, 2000), which is built into the Agalma pipeline, and manual masking. I reconstructed maximum likelihood trees using the program RAxML (Stamatakis, 2006) for datasets resulting from both the manual curation and the Agalma pipeline. PartitionFinder (Lanfear et al., 2014, 2012) was used to identify suitable models and partitioning schemes, implemented with 1% heuristic r-cluster searches, optimized weighting, RAxML likelihood calculations, and model selection based on BIC scores. In all cases, nodal support was assessed by performing 100 full non-parametric bootstraps.

I analysed two datasets resulting from the Agalma pipeline. The first dataset comprised ortholog clusters that corresponded to the manually curated 500 gene set (here on referred to as the ‘Agalma equivalent dataset’). The second dataset consisted of all ortholog clusters which had high taxon coverage (≥18), and were derived from homolog clusters containing only a single ortholog cluster (from here on referred to as the ‘Agalma best dataset’); that is, Agalma homolog clusters containing multiple copies, albeit diagnosable, were not considered further. Finally, I reconstructed a phylogeny for the camaenid dataset obtained through exon capture and included sequences from the five camaenid transcriptomes presented herein, as well as sequences of *Cornu aspersum* as an outgroup.

2.4 RESULTS

2.4.1 Transcriptome assembly and homolog clustering

The number of paired reads obtained for each of the 21 eupulmonate species sequenced ranged from 7.8M to 31.6M (Table 3). Trimming and de novo assembly statistics are presented in Table 3. The number of *L. gigantea* reference genes with BLAST matches ranged from 7,011 to 9,699 per assembly (Table 3), 5,490 of which had homologous sequences in at least 18 of the 21 transcriptome assemblies.

Of the 288 genes used in a previous molluscan phylogenomic study (Kocot et al., 2011), 130 were single-copy for all eupulmonates considered here, while 146 contained paralogs in at least one species (mean p-distance between paralogs within a sample was 0.28, ranging from 0.16-0.46). I could not unambiguously qualify the remaining 12 genes from the Kocot *et al.* study as they were poorly represented in my transcriptomes. Prioritising genes with high taxon coverage and long exon length, I assessed additional alignments of candidate homolog clusters until I reached a total of 500 single-copy genes. In addition to the 146 Kocot genes shown to be paralogous within the eupulmonates, I identified and qualified 62 multi-copy genes during the course of this work. The resulting manually curated 500 single-copy gene set is 98.5% taxa complete (i.e. sequence present for each gene and taxon) and 93.1% character complete (Figure 2.4 d), with an average gene length of 1,190 bp, ranging from 228 bp to 6,261 bp. In total, the final alignment of this gene set represents 512,958 bp. Approx. 12% of the sequences in the final gene-by-species matrix were derived by merging fragmented transcripts.

Based on the all-by-all BLAST comparison of the *L. gigantea* genes, 347 of my final 500 genes had a single hit at an e-value threshold of 1e-10 (i.e. single copy status was consistent between the *L. gigantea* reference and the eupulmonates), while the remainder had multiple hits, indicative of the presence of close paralogs in the reference. Conversely, of the 208 genes qualified as multiple-copy for the eupulmonates (146 from the Kocot gene set plus 62 from this study), 134 only had one hit within the *L. gigantea* gene set (i.e. just over half of the multiple-copy gene set are potentially single copy for patellogastropods). These results broadly correspond to the orthology designation in the OMA (Orthologous MAtrix) database.

Across the 500 single-copy genes, the p-distance between the two rhytidids, *Terrycarlessia turbinata* and *Victaphanta atramentaria*, ranged from 0.02 to 0.13 (average of 0.06; Figure 2.3). This family is thought to have originated 120 Ma (Bruggen, 1980; Upchurch, 2008). However, the Australian rhytidids probably represent a more recent radiation (Herbert et al. 2015, Moussalli and Herbert 2016). Similarly, p-distance between the two camaenids, *Sphaerospira fraseri* and *Austrochloritis kosciuszkoensis*, ranged from 0.01 to 0.13 (average of 0.04). This group is thought to have originated in the Oligo-Miocene approximately 30 Ma (Hugall and Stanisic, 2011). All genes had a higher relative substitution rate than the commonly used phylogenetic marker 28S, and were on average approximately four times slower than COI (Figure 2.3).

2.4.2 Qualification of orthology using gene tree assessments

TreSpEx analyses of all 500 genes found no well supported conflict with the *a priori* phylogenetic hypotheses, suggesting that hidden paralogs (i.e., genes represented by a single sequence per taxon yet paralogous across multiple taxa) were absent from my dataset. Furthermore, this analysis also showed no evidence of cross sample contamination, nor any evidence of suspect long internal branches within the Stylommatophora.

2.4.3 Qualification of orthology using exon capture

I enriched and sequenced all 1,646 targeted exons, from 490 genes, when considering all 22 samples collectively. I first mapped reads to the original reference used in the probe design with exon boundaries delineated based on the *L. gigantea* genome. Examination of the resulting read alignments (BAM files) identified 437 exons which contained multiple internal exon boundaries within the Camaenidae (Appendix 2.4). Accordingly, the mapping reference was modified to account for exon-splitting (including the removal of 163 exons that were shorter than 40 bp after splitting), with the final revised reference comprising 2,648 exons representing 417,846 bp (Appendix 2.9). I targeted an average of five exons per gene.

I then remapped reads to the revised reference (coverage and specificity statistics presented in Table 2.4) and flagged resulting consensus sequences which exhibited elevated polymorphism (> 3% heterozygote sites). There were 508 exons where at least one taxon exhibited elevated polymorphism. Of these, 105 exons had greater than 10% of the taxa (typically two or more taxa, taking into account missing taxa) exhibiting elevated polymorphism. Based on an examination of the corresponding read alignments, 95 exons were classified as having lineage specific pseudogenes or paralogs, four contained evidence of processed pseudogenes, and six where the alignment was complicated by the mapping of unrelated reads containing small, highly similar domains (see Appendix 2.4-8 for examples of each case). These 105 exons were removed prior to phylogenetic analyses. For the remaining 403 exons only the consensus sequences for the taxa with elevated polymorphism were removed from the final alignment. In total, 3.7% of the sequences were removed from final data matrix due to elevated polymorphism. The final exon capture data matrix was 98% taxa complete and 95% character complete.

Based on the TreSpEx analyses, four genes did not support the monophyly of the ‘Far North Chloritid’ group, but rather placed (*Nannochloritis layardi* and *Patrubella buxtoni*) as sister to the ‘North-East Chloritid’ group (Figure 2.5). I concluded that this was not the result of hidden paralogy, but rather due to insufficient lineage sorting of relatively conserved genes. An additional five genes were in conflict with the *a priori* taxonomic hypotheses, however, these represented cases where the genes were small and the proportion of phylogenetically informative sites was low. Five genes were flagged as having at least one internal branch which was greater than five times the average. Assessment of the alignments and corresponding genealogies indicated that they represented deep basal divergence between well supported major clades, and was not reflective of hidden paralogy.

Finally, I enriched another representative of *Sphaerospira fraseri*, one of the reference species used in the probe design. Comparing the mapped consensus genomic sequence to the transcriptome reference I found only minor mismatch, reflective of intraspecific variation as the two samples came from different populations (the exons had a median p-distance of 0.8%). Furthermore, for this species at least, all reference genes constructed from multiple transcript fragments were consistent with those captured from genomic DNA (i.e. chimeras of unrelated fragments were not created) and showed no evidence of paralogy or elevated heterozygosity.

2.4.4 Comparison to Agalma pipeline

Using the Agalma pipeline I identified 11,140 ortholog clusters. Of these ortholog clusters 635 corresponded to 457 of my 500 single-copy gene set. We refer to this dataset as the "Agalma equivalent" dataset, and is 61% taxa complete and 54% character complete. Many of the genes were represented by multiple ortholog clusters in the Agalma analysis, many of which contained fewer taxa relative to that obtained via manual curation (Figure 2.2). Rather than paralogs, in all cases fragmentation in the transcriptome assemblies resulted in the splitting of homolog clusters into multiple ortholog clusters, each representing the same locus but containing a different subset of taxa (see example in Appendix 2.3). Of the 43 single-copy genes not picked up by Agalma, five were not annotated in the 'postassemble' step, 12 were annotated but not recovered by the all-by-all BLAST, 18 were recovered by the all-by-all BLAST but dropped during the clustering step, and eight made it to the initial clusters but failed the alignment and trimming step prior to the gene tree reconstruction. Failure to recover these genes during the BLAST comparison, clustering and alignment steps is most likely due to a combination of frameshift errors and transcript fragmentation, and in certain cases, resulting in the taxon sampling threshold and cluster size criteria not being met.

Of the 11,140 ortholog clusters there were 546 clusters that contained sequences of at least 18 taxa and that had one ortholog cluster per homolog cluster. Of these, 171 were also contained in my 500 single-copy gene set. Hence, the Agalma pipeline identified 375 genes in addition to the 500 manually curated genes, which had optimum taxon sampling. The majority of these genes also represented the full CDS with 89% representing at least 80% of the length of the respective *L. gigantea* gene. I refer to this dataset as the “Agalma best” dataset and is 92% taxa complete and 85% character complete.

2.4.5 Phylogenetic analysis

I reconstructed phylogenies from three ortholog datasets for comparison: (1) the manually curated 500 single-copy gene set (Figure 2.4 a, d), (2) the Agalma equivalent dataset consisting of 635 orthologous clusters which corresponded to 457 of the 500 single-copy genes (Figure 2.4 b, e), and (3) the Agalma best dataset consisting of 546 orthologous cluster which had 18 or more taxa and were the only orthologous cluster from the respective homolog cluster (Figure 2.4 c, f). Of the manual curated dataset, 1.6% of the alignment was removed by Gblocks prior to phylogenetic analysis. The phylogenies for the 500 single-copy gene set and the Agalma best dataset had identical topologies, supporting all major clades with very high bootstrap support, namely Helicoidea, Limacoidea, Orthurethra, the Australian rhytidids and the Stylommatophora (Figure 2.4 a, c). In terms of phylogenetic relationships, the Rhytididae forms a sister relationship with the Limacoidea, and the Helicoidea occupies a basal position within Stylommatophora. In contrast, while also supporting the monophyly of all major clades, the phylogeny based on the ‘Agalma equivalent’ dataset places Orthurethra in a basal position within Stylommatophora, (Figure 2.4 b).

Of the Camaenidae exon capture dataset, 5% of the alignment was removed by Gblocks prior to phylogenetic analysis. The resulting phylogeny supported all major groups previously recognised by Hugall and Stanisic (2011). In terms of phylogenetic relationships, the two Chloritid groups formed a clade with the Hadroid group, with the Far-northern chloritids sister to the hadroids. There was poor resolution regarding the phylogenetic positions of the two remaining groups, the Eastern rainforests and the arid and monsoonal NW Australian clades (Figure 2.5).

2.5 DISCUSSION

The identification and qualification of orthology is a critical prerequisite for sound phylogenetic inference. My approach of orthology assessment involved an initial assessment and manual editing of homolog clusters, allowing us to correct for multiple isoforms and errors such as sequence fragmentation, frame-shifts and mis-indexing. Using this approach, I qualified the orthology and single-copy status of 500 genes across the eupulmonates, 130 of which were used in a previous phylogenomic study of the Mollusca (Kocot et al., 2011). The resulting 500 gene data matrix is the most complete produced for a major molluscan lineage to date, both in terms of taxon and character completeness. I further qualified orthology by capturing and sequencing 490 of the 500 genes from genomic DNA, revealing the presence of paralogs and/or pseudogenes otherwise not evident from the transcriptome data. Although the automated pipeline Agalma recovered the majority of the 500 genes as single copy and identified 375 additional putatively orthologous genes for the eupulmonates, it was hampered by transcript fragmentation within the assemblies. Furthermore, supported topologies for the 21 eupulmonate species were not entirely consistent between the manually curated and Agalma equivalent dataset, potentially a consequence of lower data matrix completeness in the latter. I discuss approaches to ortholog determination and implications for phylogenetic inference below.

2.5.1 Ortholog determination

To date, most transcriptome based phylogenomic studies have focused on resolving relatively deep evolutionary relationships (e.g. Kocot et al., 2011; Misof et al., 2014; O'Hara et al., 2014; Smith et al., 2011; Zapata et al., 2014), and a number have relied on annotated ortholog databases for the initial screening of suitable genes, such as OMA (Altenhoff et al., 2015), OrthoDB (Waterhouse et al., 2013), and the ortholog dataset associated with HaMStR (Ebersberger et al., 2009). Such databases are typically limited in the number of representatives per lineage (Altenhoff et al., 2015; Ranwez et al., 2007; e.g., Tatusov et al., 2003; Waterhouse et al., 2013). Nevertheless, it is a reasonable assumption that orthologous genes qualified as single-copy across many highly divergent taxa are more likely to maintain single-copy status with greater taxonomic sampling. I tested this idea at a preliminary stage of my work by first assessing genes used in a phylogenomic study of the Mollusca (Kocot et al., 2011). In that study, orthologous genes were identified using the program HaMStR, based on a 1,032 ortholog set resulting from the Inparanoid orthology database (Ostlund et al., 2010). I found that just under half of the genes used in Kocot *et al.* (Kocot et al., 2011) were paralogous within the eupulmonates. To some extent the high proportion of the Kocot *et al.* gene set being paralogous is due to the limited representation of eupulmonates in that study, and for these few taxa paralogs may have been absent. Alternatively, in such deep phylogenomic studies lineage-specific duplication may have manifested as in-paralogs and were dealt with by retaining one copy from the in-paralog set at random (Dunn et al., 2013; Kocot et al., 2011) or based on sequence similarity (Ebersberger et al., 2009). However, with an increase in taxonomic sampling, such paralogy may extend across multiple taxa and, unless conservation of function can be established (i.e. isorthology, Fitch, 2000), these genes would no longer be suitable for phylogenetic analysis.

When the 500 gene set was compared to the OMA database (Altenhoff et al., 2015), which at the time of this analysis only incorporated a single molluscan genome, namely *L. gigantea*, I found a similarly high proportion of eupulmonate specific paralogy. A more interesting result arising from this comparison, however, was that many genes classified as having putative paralogs in *L. gigantea* were single-copy across the eupulmonates. I cannot ascertain at this stage whether this is a consequence of duplication being derived within Patellogastropoda, the lineage containing *L. gigantea*, or the consequence of duplicate loss in the ancestral eupulmonate. Nevertheless, this result highlights that potentially suitable genes may be overlooked when restricted to ortholog database designations, especially when such databases have poor representation of the relevant lineage. Accordingly, although I used the *L. gigantea* gene set as a reference with which to identify and cluster homologous sequences, I did not rely on orthology database designations of the *L. gigantea* gene set to guide which genes to consider when assessing orthology across the eupulmonates examined here.

2.5.2 Automated vs manually curated aided pipelines

Pipelines that fully automate homology searches and clustering, orthology qualification, and final alignments are highly desirable for efficiency, consistency, and repeatability. Moreover, reference free methods, like that implemented in Agalma, are also highly desirable in cases where the study taxa are poorly represented in ortholog databases. There are characteristics of assembled transcriptome sequences, however, that can challenge fully automated methods, including transcript fragmentation, mis-indexing, frameshifts and contamination, and these aspects necessitate careful manual appraisal and editing (O’Hara et al., 2014; Philippe et al., 2011). Although recent phylogenomic studies have, to varying degrees, incorporated manual appraisal, such checks are typically conducted at the final proofing stage (e.g. Kocot et al., 2011; Simmons and Goloboff, 2014). In this study, I purposefully addressed the abovementioned issues at an early stage following the initial alignment of homologous sequences. The most important aspect of my manual curation was the creation of consensus sequences from fragmented transcripts (see also: O’Hara et al., 2014), which in turn ensured maximum retention of data (particularly for probe design) and placed subsequent orthology assessment on a sounder footing. Consequently, my final data matrix was highly complete (93% character complete whereas the 'Agalma best' dataset was 85% character complete).

The Agalma analysis confirmed the single-copy, orthology status for the majority of the 500 manually curated gene set, but it was hampered by transcript fragmentation within the transcriptome assemblies. In all cases where multiple ortholog clusters were derived using Agalma for any one of my 500 genes, this was due to transcript fragmentation, not missed paralogy. In essence, alignments of fragmented transcripts (whether or not they were partially overlapping) resulted in poorly reconstructed gene trees, which in turn misled subsequent tree pruning and ortholog clustering (e.g. Appendix 2.3). Consequently, for the 'Agalma equivalent' dataset, both taxon and character completeness was poor relative to the manually curated data matrix. To my knowledge, no fully automated phylogenomics pipeline currently implements the consensus of fragmented sequences, and studies that have made the effort to retain multiple fragments, as in this study, have decided which sequences to retain and merge manually (O'Hara et al., 2014; e.g., Rothfels et al., 2013). The issue of working with fragmented assemblies can be addressed, however, by incorporating an automated consensus making algorithm such as TGICL (Pertea et al., 2003) into the pipeline to address fragmentation at the homolog alignment stage. Doing so is particularly desirable, given that manual curation of homologous sequences requires considerable time investment.

A major strength of automated pipelines is that they enable a more comprehensive screening of putative orthologous genes. Manual curation requires considerable effort, and while more candidate genes were identified than were assessed, I ceased the manual assessment once my target of 500 genes had been attained. The Agalma analyses had no constraints, however, hence all possible orthologous clusters were considered. Consequently, I identified an additional 375 ortholog clusters which met a strict taxa completeness threshold (18 taxa or more) and represented the only ortholog cluster arising from original homolog clusters. These genes (i.e. the 'Agalma best' dataset) reconstructed a phylogeny that was very similar to the manually curated dataset. While beyond the scope of this study, there is potential for these genes to be included in future probe designs and further qualification of these additional genes using exon capture (see below) would be highly desirable.

2.5.3 Phylogenetic inference

The 500 gene set represents a significant contribution towards advancing molecular phylogenetics of the eupulmonates, providing the capacity to resolve both evolutionary relationships at shallow to moderate depths, and deep basal relationships. The phylogenetic reconstructions presented here are well resolved and support the *a priori* taxonomic hypotheses used as part of the orthology assessment. In terms of deeper relationships, reconstructions based on the two most complete datasets are consistent, namely the monophyly of Stylommatophora, within which Helicoidea is basal, and the sister relationship between the Rhytidoidea and the Limacoidea. For the less complete Agalma equivalent dataset, however, Orthurethra is basal within Stylommatophora, albeit with marginal support. Without greater taxonomic sampling of all the major lineages within the eupulmonates, however, a comprehensive phylogenetic assessment is beyond the scope of this study. Nevertheless, these phylogenomic datasets do afford greater resolution of deeper relationships than obtained in previous molecular studies (Wade et al., 2006, 2001). Secondly, convergence in supported topology between the two most complete and largely independent datasets (only 171 genes were in common), and the inconsistency between the manually curated and Agalma equivalent dataset (sharing 458 genes), suggests the possible importance of data matrix completeness in resolving short, basal internodes.

2.5.4 Exon capture

One of the overarching objectives of this study was to identify and qualify 500 genes suitable for exon capture work within the eupulmonates. Here I sequenced and analysed a small dataset for the family Camaenidae principally as a means to further qualify orthology. There are two principle outcomes from this exploration. First, for all reference sequences based on the concatenation of fragmented transcripts, there was no evidence that erroneous chimeric sequences were created. Second, as was the case with the increased sampling in the transcriptome work, the pervasiveness of lineage-specific duplication was also evident from the exon capture experiment. Despite qualification of single-copy orthology of the transcriptome dataset, increased taxonomic sampling within the family Camaenidae revealed lineage-specific duplication for potentially as high as one fifth of the targeted exons. In the great majority of cases, however, a very small proportion of taxa exhibited putative paralogy or pseudogenes, and removal of the affected exon per taxon only reduced the completeness of the final dataset by 3.7%. Similar results were achieved for the brittle stars with 1.5% of their target discarded due to putative paralogs or pseudogenes (Hugall et al., 2016). It is possible that these putative paralogs were only detected in the genomic sequencing because they were not expressed in the transcriptomes.

Within the Australian Camaenidae, uncorrected distances for the majority of the genes did not exceed 13%. This level of sequence variability is within the range of mismatch that is tolerated by in-solution exon capture protocols (Bi et al., 2012; Bragg et al., 2016; Hugall et al., 2016). This was qualified here given the high proportion of target recovery (>95%) across a broad representation of the camaenid diversity. As was the case for the Euplumonata phylogeny presented above, my preliminary phylogenomic dataset for camaenids provides considerable resolution, particularly among the chloritis and hadroid groups which to date have been difficult to resolve (Hugall and Stanisic, 2011).

Expanding the bait design to enrich across the Australasian camaenid radiation, indeed the family Helicoidea, would require the incorporation of multiple divergent reference taxa into the bait design. Recent "anchored enrichment" approaches to bait design (Faircloth et al., 2012; e.g. Lemmon et al., 2012) target highly conserved regions to allow capture across highly divergent taxa. By contrast, the approach taken here is to target both conserved and highly variable regions, and where possible the full coding region (Bi et al., 2012; Bragg et al., 2016; Hugall et al., 2016). Accordingly, this would require substantially greater reference diversity to be incorporated into the bait design relative to the anchored approach to capture across highly divergent lineages (e.g. across families). Recently, Hugall *et al.* (2016) used a similar approach to the one in the present study, but designed baits based on ancestral sequences, rather than representative tip taxa, to reduce the overall size of the reference set. Using this approach, Hugall *et al.* successfully enriched and sequenced both conserved and highly variable exons across the entire echinoderm class Ophiuroidea, spanning approximately 260 million years. Here I have presented a simple bait design targeting a specific family, but my transcriptome dataset could be used to produce a more diverse bait design to facilitate a more comprehensive study of Eupulmonata phylogenetics and systematics.

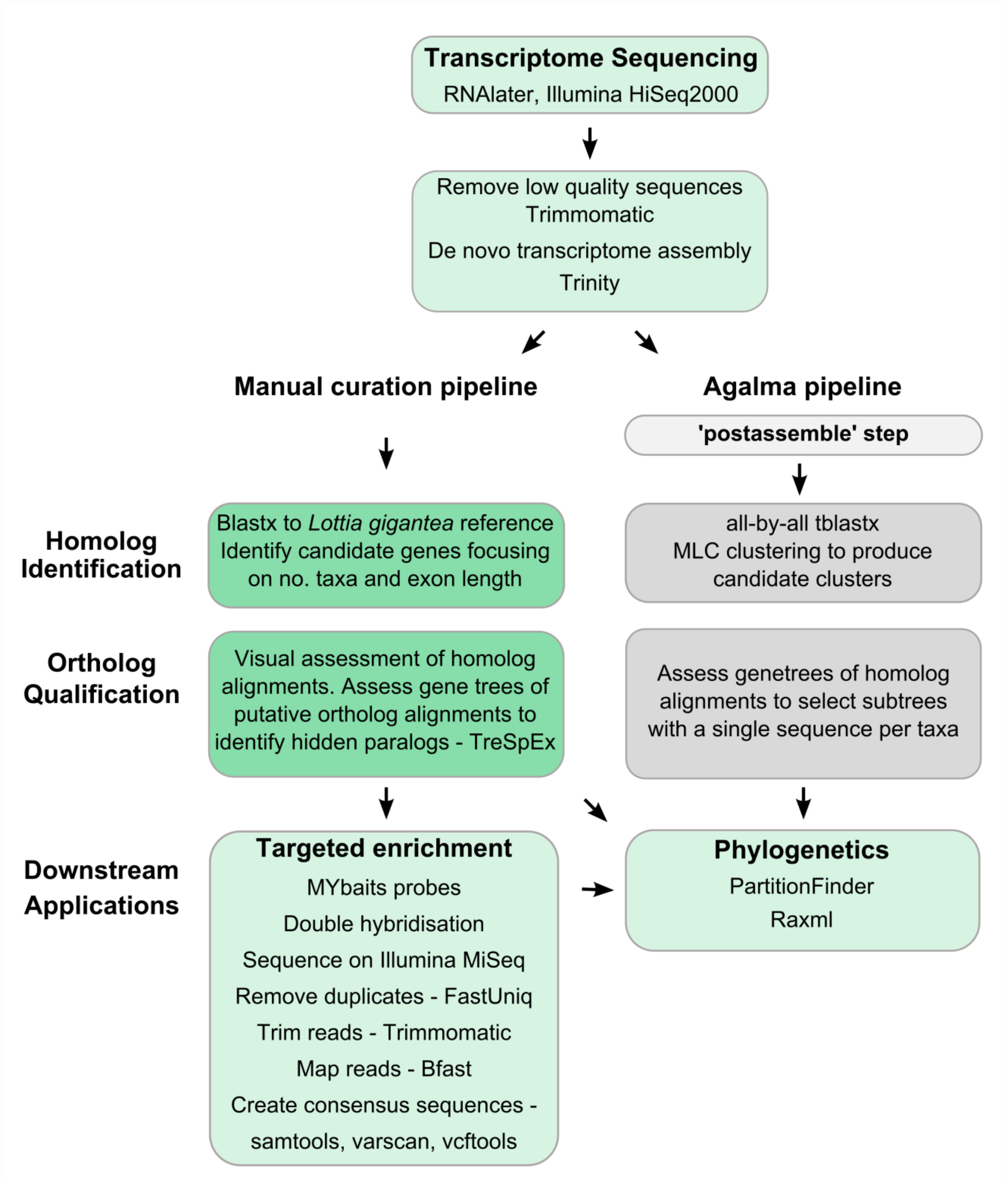

Transcriptome Sequencing
RNAlater, Illumina HiSeq2000
Remove low quality sequences
Trimmomatic
De novo transcriptome assembly
Trinity
Manual curation pipeline
Agalma pipeline
'postassemble' step
Homolog Identification
Blastx to *Lottia gigantea* reference Identify candidate genes focusing on no. taxa and exon length
all-by-all tblastx
MLC clustering to produce candidate clusters
Ortholog Qualification
Visual assessment of homolog alignments. Assess gene trees of putative ortholog alignments to identify hidden paralogs - TreSpEx
Assess genetrees of homolog alignments to select subtrees with a single sequence per taxa
Downstream Applications
Targeted enrichment
MYbaits probes
Double hybridisation
Sequence on Illumina MiSeq
Remove duplicates - FastUniq
Trim reads - Trimmomatic
Map reads - Bfast
Create consensus sequences - samtools, varscan, vcftools
Phylogenetics
PartitionFinder
Raxml

Figure 2.1. Outline of the two pipelines used to detect single-copy, orthologous genes from 21 eupulmonate transcriptomes.

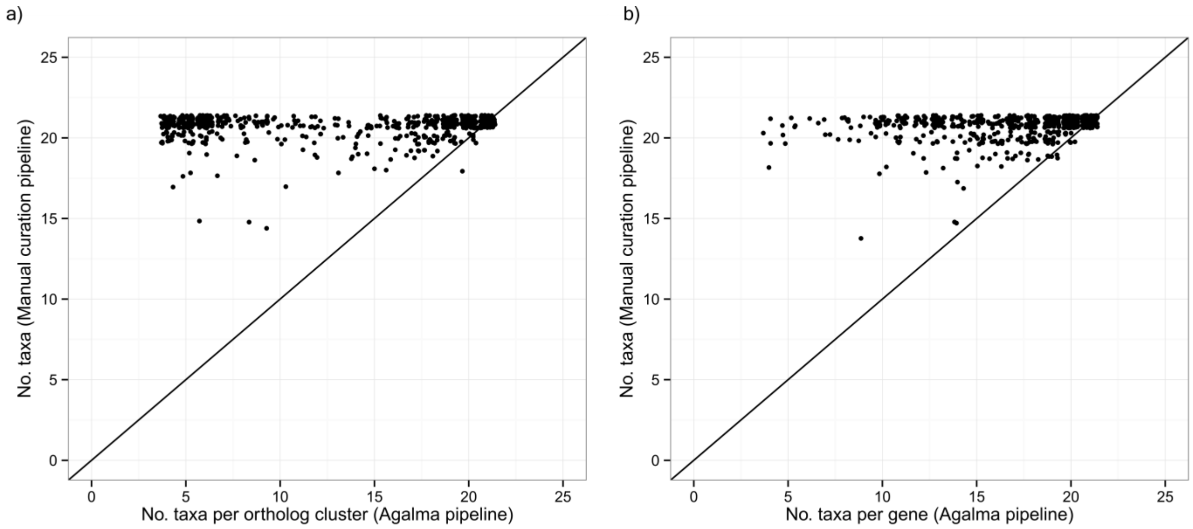


Figure. 2.2. A comparison between two orthology detection pipelines. (a) shows the relationship between the number of taxa per ortholog cluster for the ortholog clusters in common between the manual curation and Agalma pipelines. The manually curated alignments resulted in more taxa complete alignments than the corresponding Agalma alignments. (b) shows the same relationship, however, the number of taxa per gene for the Agalma pipeline were calculated across all ortholog clusters which matched the same *L. gigantea* gene. A comparison of the two plots demonstrates that Agalma tended to produce multiple independent alignments per *L. gigantea* gene, whereas a single alignment was produced through manual curation. Even when the number of taxa recovered across all Agalma alignments associated with a given gene are summed, taxa completeness of the Agalma dataset remained lower than that obtained through manual curation (see also Figure 2.4 e). These graphs are plotted using geom_jitter in ggplot2 to help visualise the large number of data points.

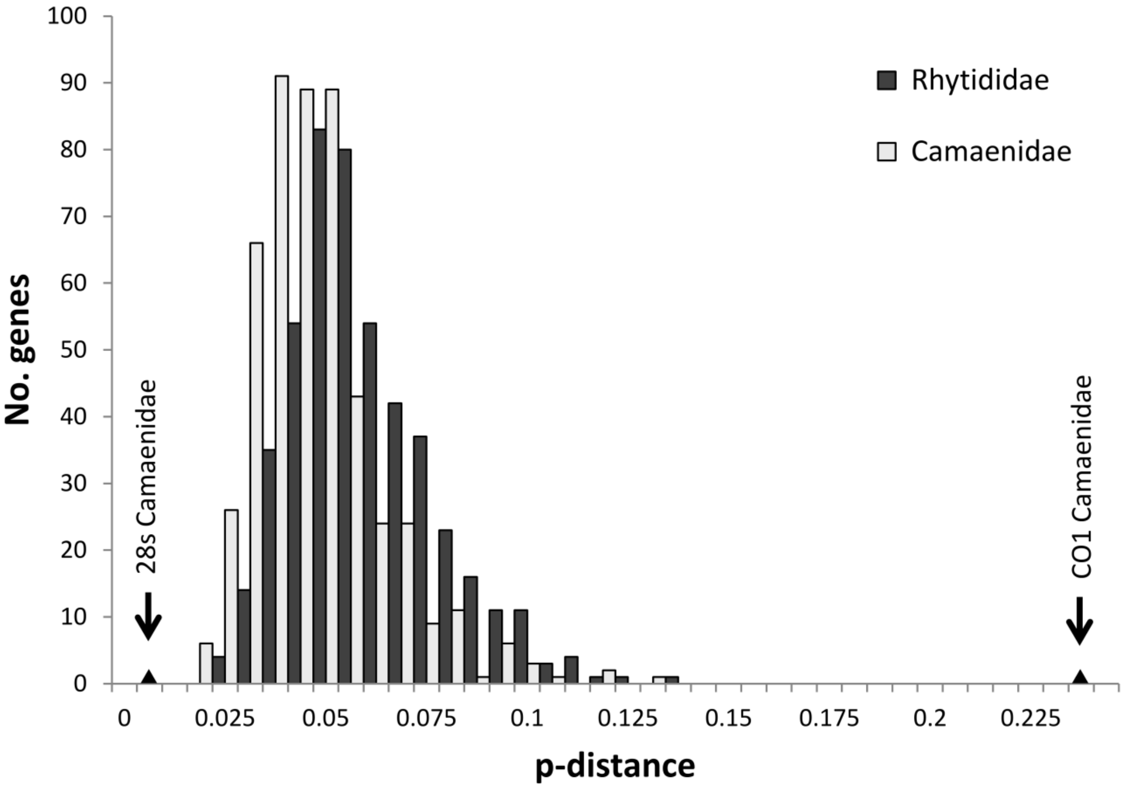


Figure 2.3. Distribution of the p-distance for 500 single-copy orthologous genes across two families. Uncorrected distances for both groups were calculated using alignments of *Terrycarlessia turbinata* and *Victaphanta atramentaria* (Rhytididae), and *Austrochloritis kosciuszkoensis* and *Sphaerospira fraseri* (Camaenidae). Triangles on the x-axis notate p-distances of two commonly used phylogenetic markers, CO1 and 28S, for the Camaenidae.

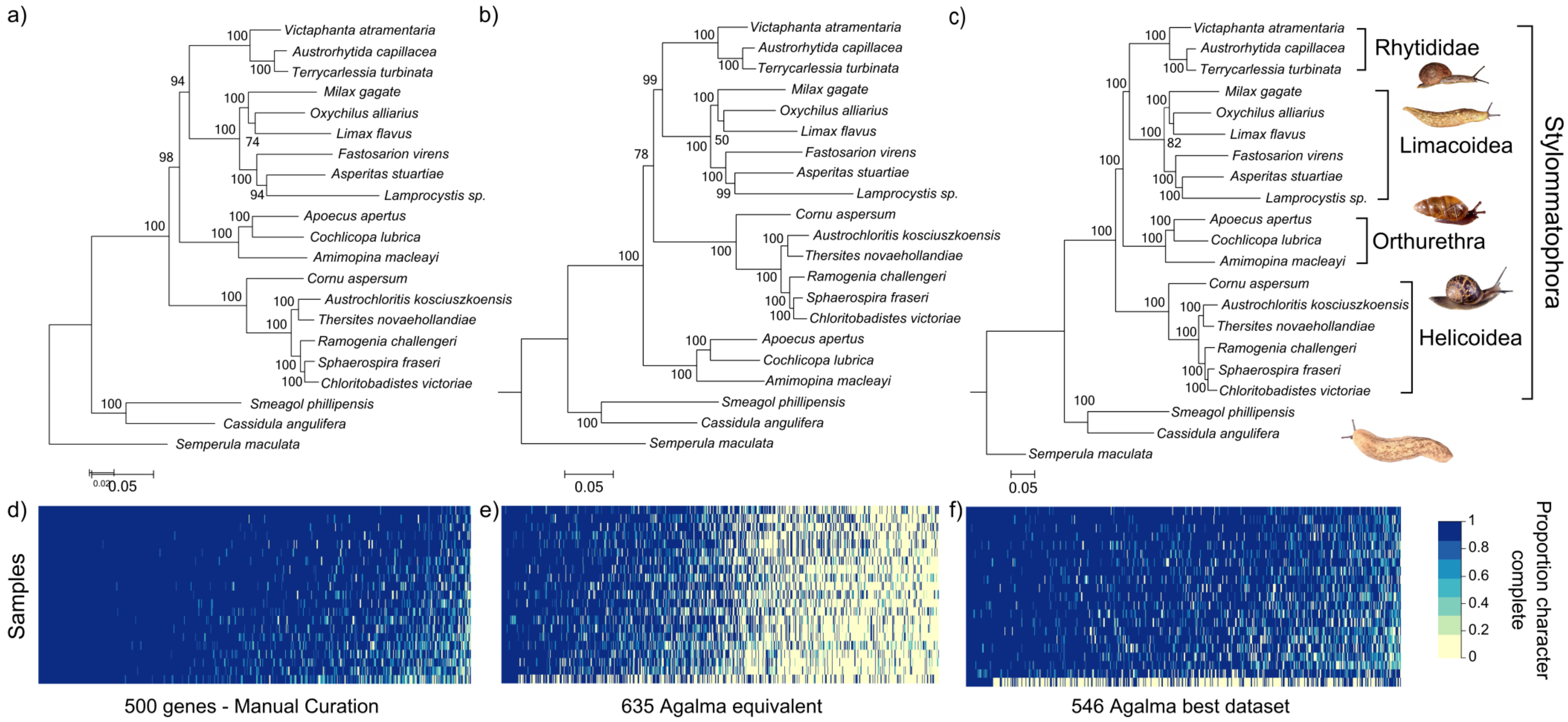


Figure 2.4. Maximum likelihood phylogenies for 21 eupulmonates based on three datasets. These datasets were (a) 500 nuclear single-copy, orthologous genes identified by manual curation, (b) 635 orthologous clusters identified by the automated pipeline Agalma, which correspond to the same 500 genes, and (c) 546 orthologous clusters identified by Agalma, where each orthologous cluster was the only one produced from the respective homolog cluster and had sequences for at least 18 taxa. Phylogenies are each based on analyses of amino acid sequences. Numbers on branches indicate bootstrap nodal support. Heat maps (d, e, f) indicate proportions of sequence obtained for each gene per sample for each dataset (sorted left to right by total proportion of data present per gene, top to bottom by total proportion of data present per sample). *Images: Dai Herbert*

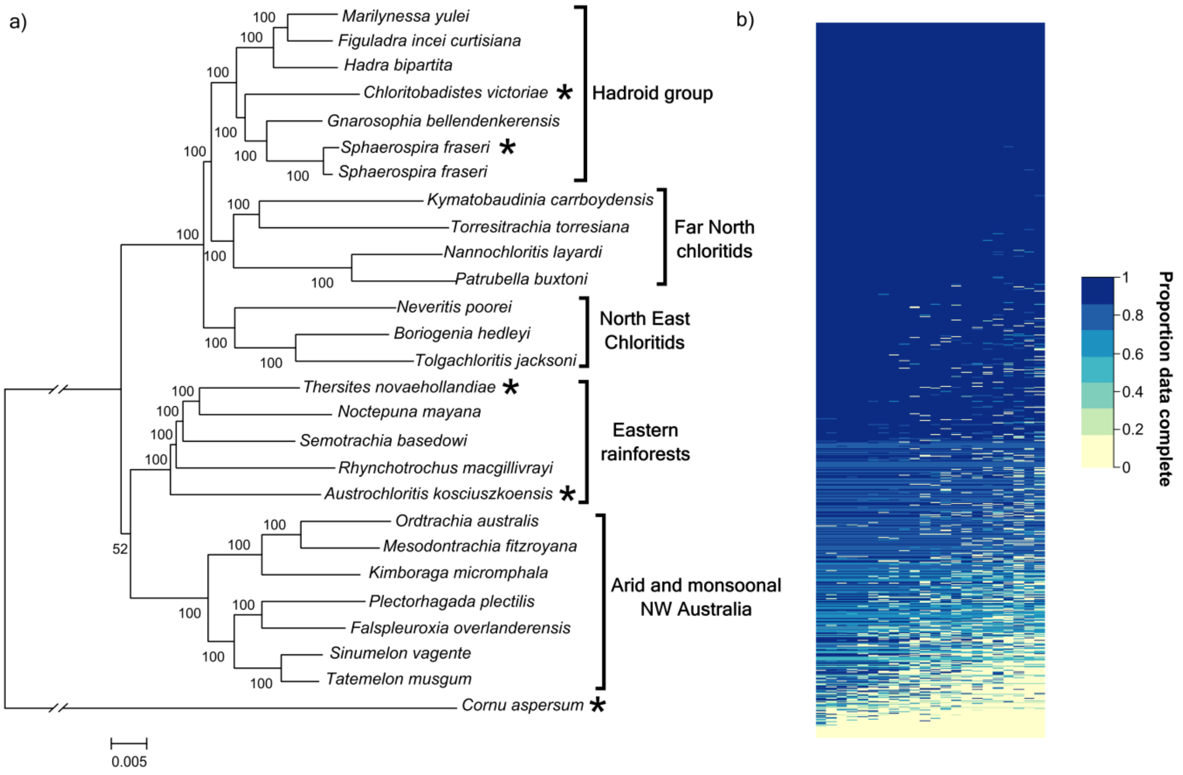


Figure 2.5. Maximum likelihood phylogeny of 26 Australian camaenid land snails. (a) Phylogenetic reconstruction based on nucleotides sequences from 2,648 exons obtained through exon capture. Sequences for the taxa marked with asterisks were derived from transcriptome datasets. Numbers on branches indicate bootstrap nodal support. (b) Heat map showing the proportion of available sequences for each sample per gene (sorted left to right by proportion of data present per sample; top to bottom by proportion of data present per exon).

Table 2.1. Taxon sampling: Transcriptome sequencing

Superfamilies or higher unranked classification	Family	Species	Voucher specimen	Collection locality*
Helicoidea	Camaenidae	*Austrochloritis kosciuszkoensis* Shea & Griffiths, 2010	NMV F193285	Sylvia Creek, VIC
Helicoidea	Camaenidae	*Chloritobadistes victoriae* (Cox, 1868)	NMV F193288	Crawford River, VIC
Helicoidea	Camaenidae	*Ramogenia challengeri* (Gude, 1906)	NMV F193287	Noosa, QLD
Helicoidea	Camaenidae	*Sphaerospira fraseri* (Griffith & Pidgeon, 1833)	NMV F193284	Noosa, QLD
Helicoidea	Camaenidae	*Thersites novaehollandiae* (Gray, 1834)	NMV F193248	Comboyne, NSW
Helicoidea	Helicidae	*Cornu aspersum* Müller, 1774	NMV F193280	Melbourne, VIC
Limacoidea	Dyakiidae	*Asperitas stuartiae* (Pfeiffer, 1845)	NMV F193286	North of Dili, Timor-Leste
Limacoidea	Helicarionidae	*Fastosarion cf virens* (Pfeiffer, 1849)	NMV F193282	Noosa, QLD
Limacoidea	Limacidae	*Limax flavus* Linnaeus, 1758	NMV F193283	Melbourne, VIC
Limacoidea	Microcystidae	*Lamprocystis sp.*	AM C.476947	Ramelau Mountains, Timor-Leste
Limacoidea	Milacidae	*Milax gagates* (Draparnaud, 1801)	NMV F226625	Melbourne, VIC
Limacoidea	Oxychilidae	*Oxychilus alliarius* (Miller, 1822)	NMV F226626	Melbourne, VIC
Orthurethra	Cerastidae	*Amimopina macleayi* (Brazier, 1876)	NMV F193290	Darwin, NT
Orthurethra	Cochlicopidae	*Cochlicopa lubrica* (Müller, 1774)	MV614	Blue Mountains, NSW
Orthurethra	Enidae	*Apoecus apertus* (Martens, 1863)	AM C.488753	Ramelau Mountains, Timor-Leste
Rhytidoidea	Rhytididae	*Austrorhytida capillacea* (Férussac, 1832)	NMV F193291	Blue Mountains, NSW
Rhytidoidea	Rhytididae	*Terrycarlessia turbinata* Stanisic, 2010	NMV F193292	Comboyne, NSW
Rhytidoidea	Rhytididae	*Victaphanta atramentaria* (Shuttleworth, 1852)	NMV F226627	Toolangi, VIC
Ellobioidea	Ellobiidae	*Cassidula angulifera* (Petit, 1841)	NMV F193289	Manatuto, Timor-Leste
Otinoidea	Smeagolidae	*Smeagol phillipensis* Tillier & Ponder, 1992	MVR13_138	Phillip Is., VIC
Veronicelloidea	Veronicellidae	*Semperula maculata* (Templeton, 1858)	AM C.476934	Manatuto, Timor-Leste

*All localities within Australia unless otherwise indicated

Table 2.2. Taxon sampling: Exon capture

Species	Voucher specimen	Collection locality*
Boriogenia hedleyi (Fulton, 1907)	MV1082	Cairns, QLD
Falspleuroxia overlanderensis Solem, 1997	WAM S70235	Shark Bay, WA
Figuladra incei curtisiana (Pfeiffer, 1864)	NMV F219323	Mt Archer, QLD
Gnarosophia bellendenkerensis (Brazier, 1875)	NMV F226513	Alligator creek, QLD
Hadra bipartita (Férussac, 1823)	AM C.476663	Green Island, QLD
Kimboraga micromphala (Gude, 1907)	AM C.463554	Windjana Gorge, WA
Kymatobaudinia carrboydensis Criscione & Köhler, 2013	WAM 49172	Carr Boyd Ranges, WA
Marilynessa yulei (Forbes, 1851)	MV1265	Brandy Creek, QLD
Mesodontrachia fitzroyana Solem, 1985	AM C.476985	Victoria River District, NT
Nannochloritis layardi (Gude, 1906)	AM C.477826	Somerset, QLD
Neveritis poorei (Gude, 1907)	MV1054	Mt Elliot, QLD
Noctepuna mayana (Hedley, 1899)	AM C.478270	Diwan, QLD
Ordtrachia australis Solem, 1984	AM C.462736	Victoria River District, NT
Patrubella buxtoni (Brazier, 1880)	AM C.478884	Moa Is., Torres Strait
Plectorhagada plectilis (Benson, 1853)	WAM S70240	Shark Bay, WA
Rhynchotrochus macgillivrayi (Forbes, 1851)	AM C.478271	Diwan, QLD
Semotrachia basedowi (Hedley, 1905)	AM C.476884	Musgrave Ranges, WA
Sinumelon vagente Iredale, 1939	WA 61253	Mt Gibson, WA
Sphaerospira fraseri (Griffith & Pidgeon, 1833)	MV1104	Benarkin State Forest, QLD
Tatemelon musgum (Iredale, 1937)	AM C.476881	Musgrave Ranges, WA
Tolgachloritis jacksoni (Hedley, 1912)	NMV F226521	Mt Garnet, QLD
Torresitrachia torresiana (Hombron & Jacquinot, 1841)	AM C.477860	Weipa, Cape York Peninsula, QLD

*All localities within Australia unless otherwise indicated

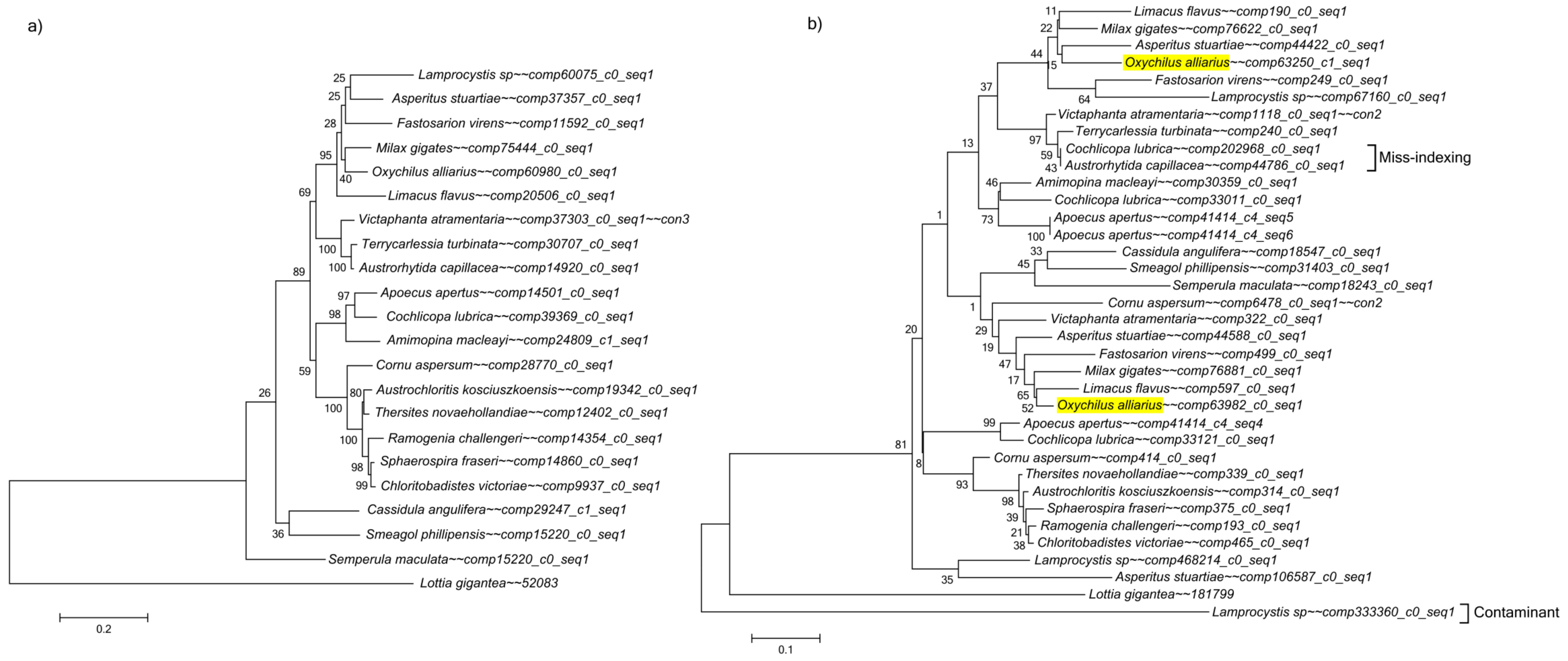
Table 2.3. Summary statistics for sequencing and *de novo* assembly of 21 eupulmonate transcriptomes

Species	**Pairs of raw reads**	**Proportion of reads after trimming**	**Trinity contigs**	**BLAST hits 1e-10 (*L. gigantea*)**	***L. gigantea* genes with hits**	**No. of the 500 single copy genes**
Ramogenia challengeri	11,726,377	0.84	103,471	14,665	7,011	488
Austrochloritis kosciuszkoensis	11,357,080	0.85	107,810	16,238	7,522	495
Sphaerospira fraseri	31,594,841	0.85	179,695	23,910	9,433	500
Thersites novaehollandiae	15,620,892	0.85	118,298	17,330	7,869	492
Chloritobadistes victoriae	26,433,009	0.85	148,817	20,453	8,792	498
Amimopina macleayi	7,874,195	0.97	93,250	17,258	8,091	494
Cochlicopa lubrica	8,074,560	0.97	111,396	21,675	9,086	497
Asperitas stuartiae	9,322,853	0.97	104,942	15,491	7,460	491
Cassidula angulifera	14,281,906	0.97	105,803	16,981	8,083	489
Apoecus apertus	9,362,182	0.97	119,711	21,275	9,095	497
Fastosarion cf *virens*	14,904,669	0.84	127,454	18,306	7,987	494
Cornu aspersum	21,273,910	0.86	160,490	23,114	9,254	498
Limax flavus	14,907,395	0.84	116,088	19,071	8,349	497
Lamprocystis sp.	22,539,699	0.97	128,611	23,797	9,679	499
Milax gagates	11,263,950	0.97	92,337	16,541	7,041	490
Oxychilus alliarius	12,925,111	0.97	136,044	21,183	8,940	499
Terrycarlessia turbinata	16,985,068	0.84	141,421	17,073	7,778	489
Victaphanta atramentaria	11,312,274	0.86	101,127	16,584	7,466	490
Austrorhytida capillacea	10,154,817	0.96	88,525	15,352	7,118	477
Smeagol phillipensis	6,393,571	0.96	95,429	23,067	9,699	497
Semperula maculata	12,461,924	0.97	76,847	21,851	9,276	492

Table 2.4. Sequencing and mapping summary statistics for the exon capture experiment.

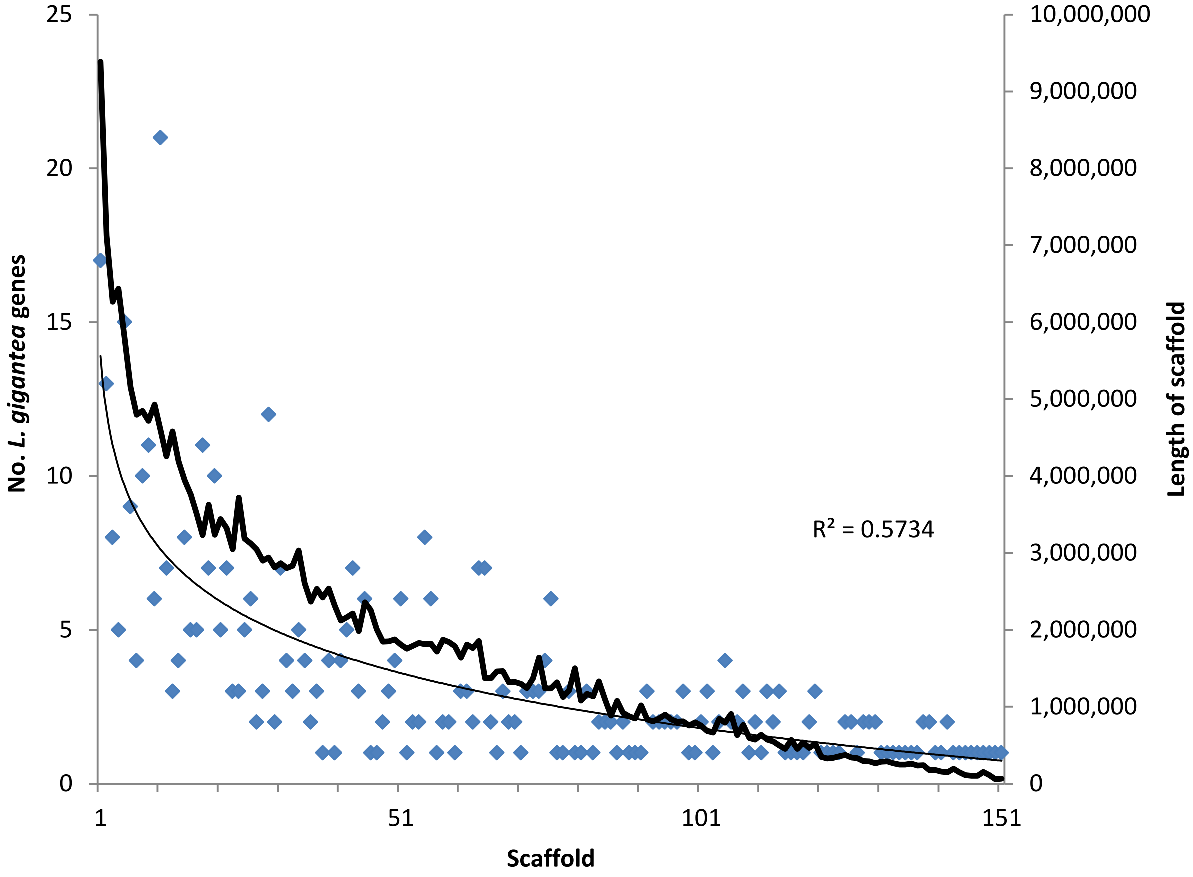
Species	No. raw paired end reads	Proportion of pairs of reads retained after duplicate removal	Proportion retained after Trimmomatic	Proportion of reads mapped to the final reference	Average coverage per exon	Proportion of exons captured (total 2648 exons)
Boriogenia hedleyi	836,437	0.60	0.97	0.64	145	0.96
Falspleuroxia overlanderensis	170,769	0.69	0.98	0.74	41	0.88
Figuladra incei curtisiana	1,117,954	0.57	0.96	0.6	167	0.97
Gnarosophia bellendenkerensis	1,490,686	0.57	0.98	0.63	235	0.98
Hadra bipartita	659,509	0.6	0.98	0.7	131	0.96
Kimboraga micromphala	186,942	0.86	0.99	0.73	55	0.90
Kymatobaudinia carrboydensis	666,965	0.78	0.98	0.63	145	0.94
Marilynessa yulei	865,712	0.56	0.97	0.62	139	0.97
Mesodontrachia fitzroyana	429,572	0.85	0.98	0.61	102	0.91
Nannochloritis layardi	179,432	0.86	0.97	0.72	50	0.90
Neveritis poorei	1,313,049	0.57	0.96	0.62	205	0.95
Noctepuna mayana	297,503	0.77	0.98	0.73	81	0.93
Ordtrachia australis	670,743	0.65	0.94	0.86	222	0.92
Patrubella buxtoni	492,474	0.82	0.97	0.7	125	0.92
Plectorhagada plectilis	220,636	0.81	0.98	0.76	65	0.90
Rhynchotrochus macgillivrayi	340,338	0.85	0.98	0.7	96	0.92
Semotrachia basedowi	290,966	0.92	0.88	0.83	119	0.92
Sinumelon vagente	282,838	0.86	0.97	0.75	86	0.92
Sphaerospira fraseri	796,591	0.56	0.98	0.66	130	0.98
Tatemelon musgum	242,614	0.87	0.99	0.7	66	0.91
Tolgachloritis jacksoni	1,207,039	0.38	0.97	0.65	139	0.95
Torresitrachia torresiana	192,031	0.87	0.98	0.74	61	0.90

2.6 APPENDICES

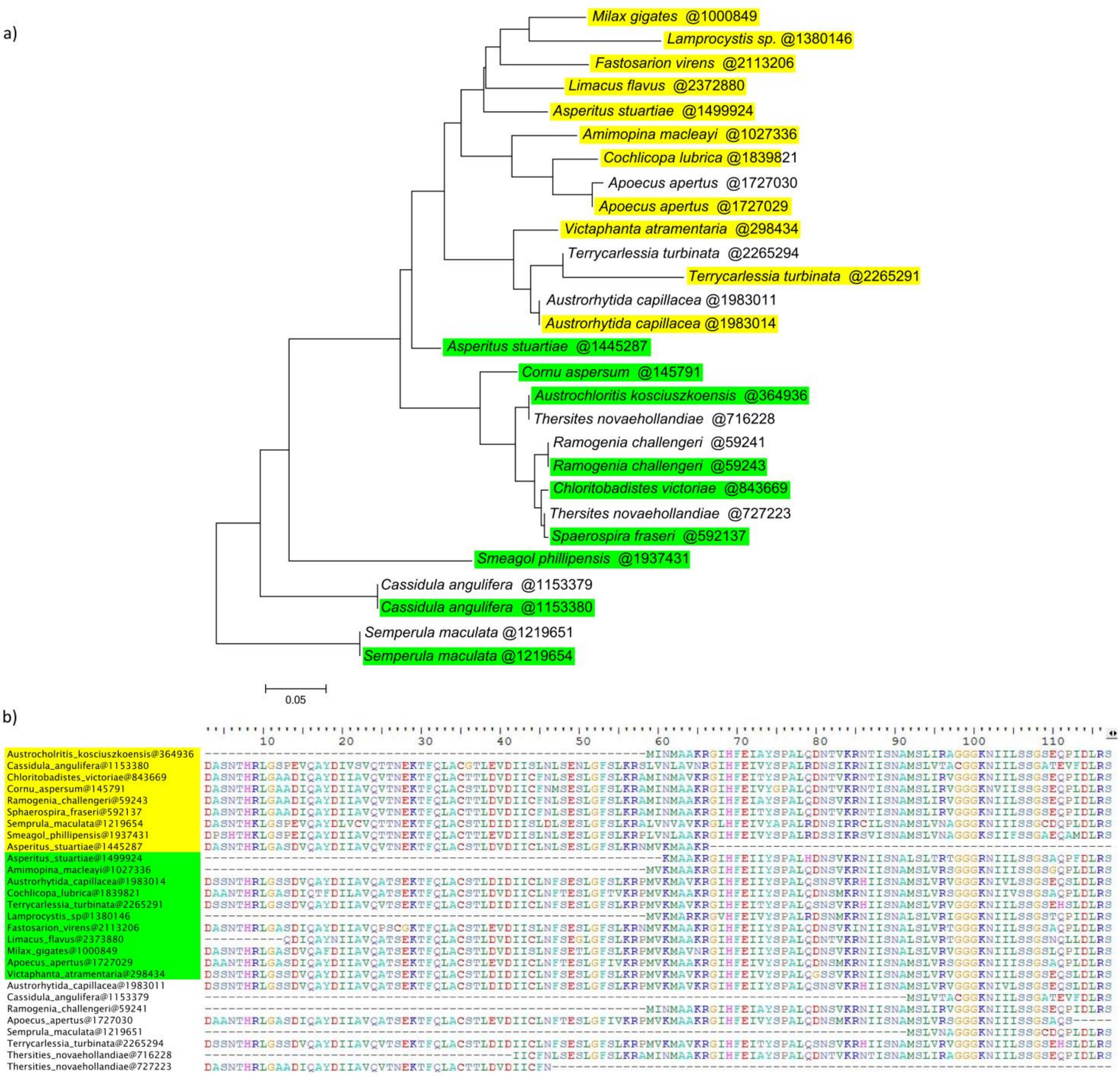


a)
Lamprocystis sp~~comp60075_c0_seq1
Asperitus stuartiae~~comp37357_c0_seq1
Fastosarion virens~~comp11592_c0_seq1
Milax gigates~~comp75444_c0_seq1
Oxychilus alliarius~~comp60980_c0_seq1
Limacus flavus~~comp20506_c0_seq1
Victaphanta atramentaria~~comp37303_c0_seq1~~con3
Terrycarlessia turbinata~~comp30707_c0_seq1
Austrorhytida capillacea~~comp14920_c0_seq1
Apoecus apertus~~comp14501_c0_seq1
Cochlicopa lubrica~~comp39369_c0_seq1
Amimopina macleayi~~comp24809_c1_seq1
Cornu aspersum~~comp28770_c0_seq1
Austrochloritis kosciuszkoensis~~comp19342_c0_seq1
Thersites novaehollandiae~~comp12402_c0_seq1
Ramogenia challengeri~~comp14354_c0_seq1
Sphaerospira fraseri~~comp14860_c0_seq1
Chloritobadistes victoriae~~comp9937_c0_seq1
Cassidula angulifera~~comp29247_c1_seq1
Smeagol phillipensis~~comp15220_c0_seq1
Semperula maculata~~comp15220_c0_seq1
Lottia gigantea~~52083
25
25
28
95
40
69
100
100
89
97
98
59
80
100
100
98
99
26
36
0.2
b)
Limacus flavus~~comp190_c0_seq1
Milax gigates~~comp76622_c0_seq1
Asperitus stuartiae~~comp44422_c0_seq1
Oxychilus alliarius~~comp63250_c1_seq1
Fastosarion virens~~comp249_c0_seq1
Lamprocystis sp~~comp67160_c0_seq1
Victaphanta atramentaria~~comp1118_c0_seq1~~con2
Terrycarlessia turbinata~~comp240_c0_seq1
Cochlicopa lubrica~~comp202968_c0_seq1
Austrorhytida capillacea~~comp44786_c0_seq1
Miss-indexing
Amimopina macleayi~~comp30359_c0_seq1
Cochlicopa lubrica~~comp33011_c0_seq1
Apoecus apertus~~comp41414_c4_seq5
Apoecus apertus~~comp41414_c4_seq6
Cassidula angulifera~~comp18547_c0_seq1
Smeagol phillipensis~~comp31403_c0_seq1
Semperula maculata~~comp18243_c0_seq1
Cornu aspersum~~comp6478_c0_seq1~~con2
Victaphanta atramentaria~~comp322_c0_seq1
Asperitus stuartiae~~comp44588_c0_seq1
Fastosarion virens~~comp499_c0_seq1
Milax gigates~~comp76881_c0_seq1
Limacus flavus~~comp597_c0_seq1
Oxychilus alliarius~~comp63982_c0_seq1
Apoecus apertus~~comp41414_c4_seq4
Cochlicopa lubrica~~comp33121_c0_seq1
Cornu aspersum~~comp414_c0_seq1
Thersites novaehollandiae~~comp339_c0_seq1
Austrochloritis kosciuszkoensis~~comp314_c0_seq1
Sphaerospira fraseri~~comp375_c0_seq1
Ramogenia challengeri~~comp193_c0_seq1
Chloritobadistes victoriae~~comp465_c0_seq1
Lamprocystis sp~~comp468214_c0_seq1
Asperitus stuartiae~~comp106587_c0_seq1
Lottia gigantea~~181799
Lamprocystis sp~~comp333360_c0_seq1
Contaminant
11
22
44
15
37
64
13
97
59
43
46
73
100
1
33
45
1
29
19
47
17
65
52
20
99
81
8
93
98
39
21
38
35
0.1

Appendix 2.1. *Example of paralogy, mis-indexing and contamination*. Gene trees demonstrating (a) a clean single copy orthologous gene with only one sequence per taxon and (b) a gene with evidence of paralogous sequences for multiple taxa. An example of paralogous sequences in *Oxychilus alliarus* is highlighted in yellow in gene tree (b). These sequences occur in different parts of the tree yet result from a duplication which has occurred within the Stylommatophora. Gene tree b) also shows evidence of mis-indexing and contamination. However, these attributes would not have led to rejection of this gene for phylogenetics. Mis-indexing occurs when a sequencing error in the read barcode leads to the read being assigned to the wrong sample. In this case *Austrorhytida capillacea* reads have been included in the *Cochlicopa lubrica* sample. The *Lamprocystis sp.* contig comp333360_c0_seq1 was identified as a nematode sequence when compared to Genbank. The gene trees were constructed using the Maximum Likelihood method based on the Tamura-Nei model and tested with 100 bootstraps in MEGA5.10.

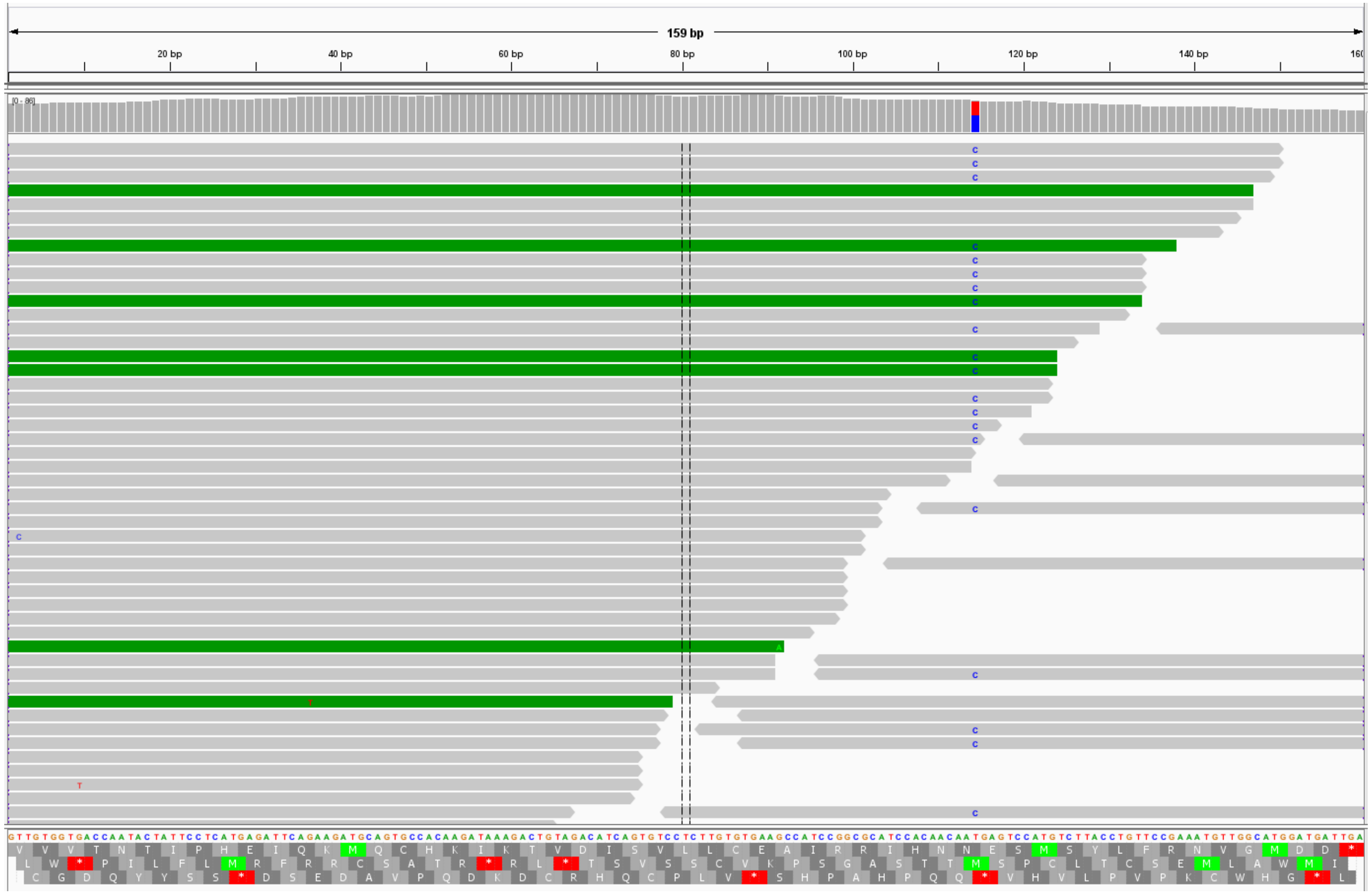


Appendix 2.2. The distribution of the 500 single-copy genes, identified through the manual curation pipeline, across the *Lottia gigantea* genome scaffolds. Each blue point represents the number of *L. gigantea* genes found on each of 151 genome scaffolds which had at least one of the 500 single copy orthologous genes. The thick black line represents the length of each scaffold and the thin black line shows the logarithmic regression of the number of 500 single-copy genes on each scaffold. This graph demonstrates that the 500 genes are essentially randomly distributed across the *L. gigantea* scaffolds with the number of my single copy genes on each *L. gigantea* scaffold correlated with its length ($R^2 = 0.57$).

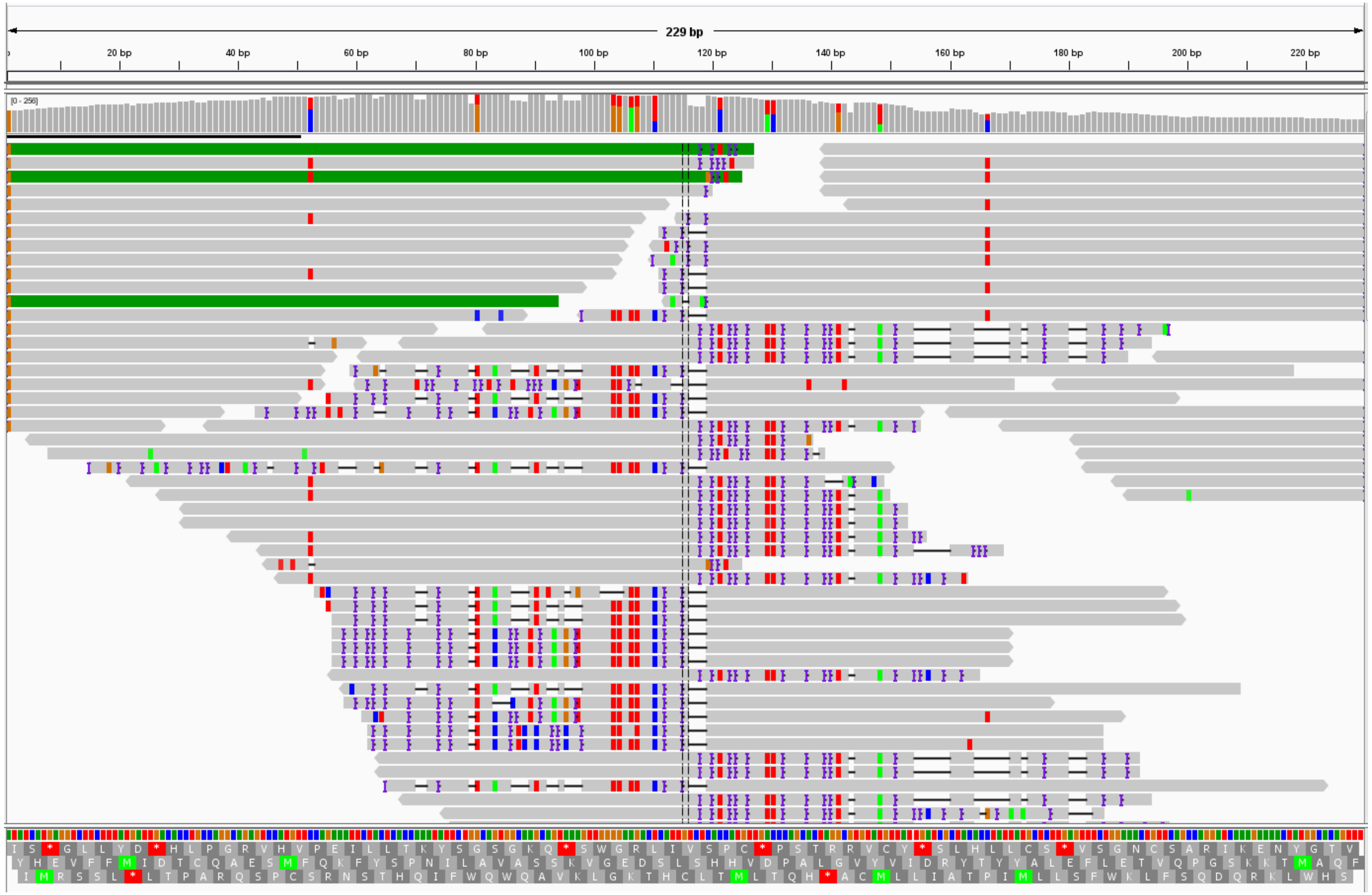

a)
Milax gigates @1000849
Lamprocystis sp. @1380146
Fastosarion virens @2113206
Limacus flavus @2372880
Asperitus stuartiae @1499924
Amimopina macleayi @1027336
Cochlicopa lubrica @1839821
Apoecus apertus @1727030
Apoecus apertus @1727029
Victaphanta atramentaria @298434
Terrycarlessia turbinata @2265294
Terrycarlessia turbinata @2265291
Austrorhytida capillacea @1983011
Austrorhytida capillacea @1983014
Asperitus stuartiae @1445287
Cornu aspersum @145791
Austrochloritis kosciuszkoensis @364936
Thersites novaehollandiae @716228
Ramogenia challengeri @59241
Ramogenia challengeri @59243
Chloritobadistes victoriae @843669
Thersites novaehollandiae @727223
Spaerospira fraseri @592137
Smeagol phillipensis @1937431
Cassidula angulifera @1153379
Cassidula angulifera @1153380
Semperula maculata @1219651
Semperula maculata @1219654
0.05
b)

```
                                                 10        20        30        40        50        60        70        80        90       100       110
Austrocholritis_kosciuszkoensis@364936  --------------------------------------------------------MINMAAKRGIHFEIAYSPALQDNTVKRNTISNAMSLIRAGGGKNIILSSGSEQPIDLRS
Cassidula_angulifera@1153380            DASNTHRLGSPEVQAYDIVSVQTTNEKTFQLACGTLEVDIISLNLSENLGFSLKRSLVNLAVNRGIHFEIVYSPALRDNSIKRNIISNAMSLVTACGGKNIILSSGATEVFDLRS
Chloritobadistes_victoriae@843669       DASNTHRLGAADIQAYDIIAVQVTNEKTFQLACTTLDVDIICFNLSESLGFSLKRAMINMAVKRGIHFEITYSPALQDNTVKRNTISNAMSLIRVGGGKNIILSSGSEQPIDLRS
Cornu_aspersum@145791                   DASNTHRLGAADIQAYDIIAVQVTNEKTFQLACSTLDVDIICFNMSESLGFSLKRAMINMAAKRGIHFEIVYGPALQDNTVKRNTISNAMSLIRVGGGKNVIISSGSEQPLDLRS
Ramogenia_challengeri@59243             DASNTHRLGAADIQAYDIIAVQVTNEKTFQLACTTLDVDIICFNLSESLGFSLKRAMINMAAKRGIHFEIAYSPALQDNSVKRNIISNAMSLIRVGGGKNIILSSGSEQPIDLRS
Sphaerospira_fraseri@592137             DASNTHRLGAADIQAYDIIAVQVTNEKTFQLACTTLDVDIICFNLSESLGFSLKRAMINMAAKRGIHFEITYSPALQDNTVKRNTISNAMSLIRVGGGKNIILSSGSEQPIDLRS
Semprula_maculata@1219654               DASNTHKLGSPEVQAYDLVCVQTTNEKTFQLACSTLDIDIISLDLSESLGFSLKRALVNVAVKRGLHFEIVYAPALRDNSIRRCILSNAMSLVNAGGGKNIIISSGCDQPLDLRS
Smeagol_phillipensis@1937431            DPSHTHKLGSPEIQAYDIIAVQTTNEKTFQLACTTLEVDIISLNLSESLGFSLKRPLVNLAAKRGIHFEIVYSPALRDSSIKRSVISNAMSLVNAGGGKSIIFSSGAEQAMDLRS
Asperitus_stuartiae@1445287             DASNTHRLGASDVQAYDIIAVQVTNEKTFQLACSTLDVDIICLNLSESLGFSLKRNMVKMAAKR---------------------------------------------------
Asperitus_stuartiae@1499924             ----------------------------------------------------------KMAAKRGIHFEIIYSPALHDNSVKRNIISNALSLTRTGGGRNIILSSGSAQPFDLRS
Amimopina_macleayi@1027336              --------------------------------------------------------MVKMAAKRGIHFEIIYSPALQDNSVKRNIISNAMSLVRSGGGKNIIISSGSGQPLDLRS
Austrorhytida_capillacea@1983014        DSSNTHRLGSSDVQAYDIIAVQATSEKTFQLACSTLDIDIICLNFSESLGFSLKRPMVKMAVKRGIHFEIIYSPALQSNSVKRHIISNAMSLVRVGGGKNIVLSSGSEQSLDLRS
Cochlicopa_lubrica@1839821              DAANTHRLGASDIQTFDILAVQATSEKTFQLACSTLDVDIICLNFTESLGFTVKRPMVKMAAKRGIHFEIIYSPALQDNSMKRNIISNAMSLVRSGGGKNIIVSSGSAQPLDLRS
Terrycarlessia_turbinata@2265291        DSSNTHRLGSSDVQAYDIIAVQVTSEKTFQLACSTLDIDIICLNFSESLGFSLKRPMVKMAVKRGIHFEITYSPALQSNSVKRHIISNAMSLVRVGGGKNIILSSGSEHSLDLRS
Lamprocystis_sp@1380146                 --------------------------------------------------------MVKMARKRGVHFEIVYSPALRDSNMKRNIISNALSLVRIGGGKNIILSSGSTQPIDLRS
Fastosarion_virens@2113206              DASNTHRLGASDIQAYDIIAVQPSCGKTFQLACSTLEVDIISLNFSESLGFSLKRNMVKMAAKRGIHFEIIYSPALQDNSVKINIISNALSLVRTGGGKNIILSSGSAQPLDLRS
Limacus_flavus@2373880                  ----------QDIQAYNIIAVQATSEKTFQLACSTLDVDIICLNFSEGLGFSLKRPMVKMAAKRGIHFEIIYSPALQDNTVKRNIISNALSLVRTGGGKNIILSSGSNQLLDLRS
Milax_gigates@1000849                   DASNTHRLGASDVQAFDIIAVQATSEKTFQLACSTLDVDIISLNFSETLGFSLKRPMVKMAVNRGIHFEIIYSPALQDNSVKRNIISNALSLVRVGGGKNIILSSGSLQPIDLRS
Apoecus_apertus@1727029                 DAANTHRLGASDIQAYDIIAVQATSEKTFQLACSTLEVDIICLNFTESLGFIVKRPMVKMAAKRGIHFEIVYSPALQDNSMKRNIISNAMSLVRSGGGKNIIISSGSAQPLDLRS
Victaphanta_atramentaria@298434         DSSNTHRLGSSDVQAYDIIAVQATSEKTFQLACSTLDIDIICLNFSESLGFSLKRPMVKMAVKRGIHFEIVYSPALQGSSVKRNIISNAMSLVRVGGGKNIILSSGSEQSLDLRS
Austrorhytida_capillacea@1983011        DSSNTHRLGSSDVQAYDIIAVQATSEKTFQLACSTLDIDIICLNFSESLGFSLKRPMVKMAVKRGIHFEIIYSPALQSNSVKRHIISNAMSLVRVGGGKNIVLSSGSEQSLDLRS
Cassidula_angulifera@1153379            -----------------------------------------------------------------------------------------MSLVTACGGKNIILSSGATEVFDLRS
Ramogenia_challengeri@59241             --------------------------------------------------------MINMAAKRGIHFEIAYSPALQDNSVKRNIISNAMSLIRVGGGKNIILSSGSEQPIDLRS
Apoecus_apertus@1727030                 DAANTHRLGASDIQAYDIIAVQATSEKTFQLACSTLEVDIICLNFTESLGFIVKRPMVKMAAKRGIHFEIVYSPALQDNSMKRNIISNAMSLVRSGGGKNIIISSGSAQS-----
Semprula_maculata@1219651               -----------------------------------------------------------------------------------------MSLVNAGGGKNIIISSGCDQPLDLRS
Terrycarlessia_turbinata@2265294        DSSNTHRLGSSDVQAYDIIAVQVTSEKTFQLACSTLDIDIICLNFSESLGFSLKRPMVKMAVKRGIHFEITYSPALQSNSVKRHIISNAMSLVRVGGGKNIILSSGSEHSLDLRS
Thersities_novaehollandiae@716228       ---------------------------------------IICFNLSESLGFSLKRAMINMAAKRGIHFEIAYSPALQDNTVKRNTISNAMSLIRAGGGKNIILSSGSEQPIDLRS
Thersities_novaehollandiae@727223       DASNTHRLGAADIQAYDIIAVQVTNEKTFQLACTTLDVDIICFN-----------------------------------------------------------------------
```

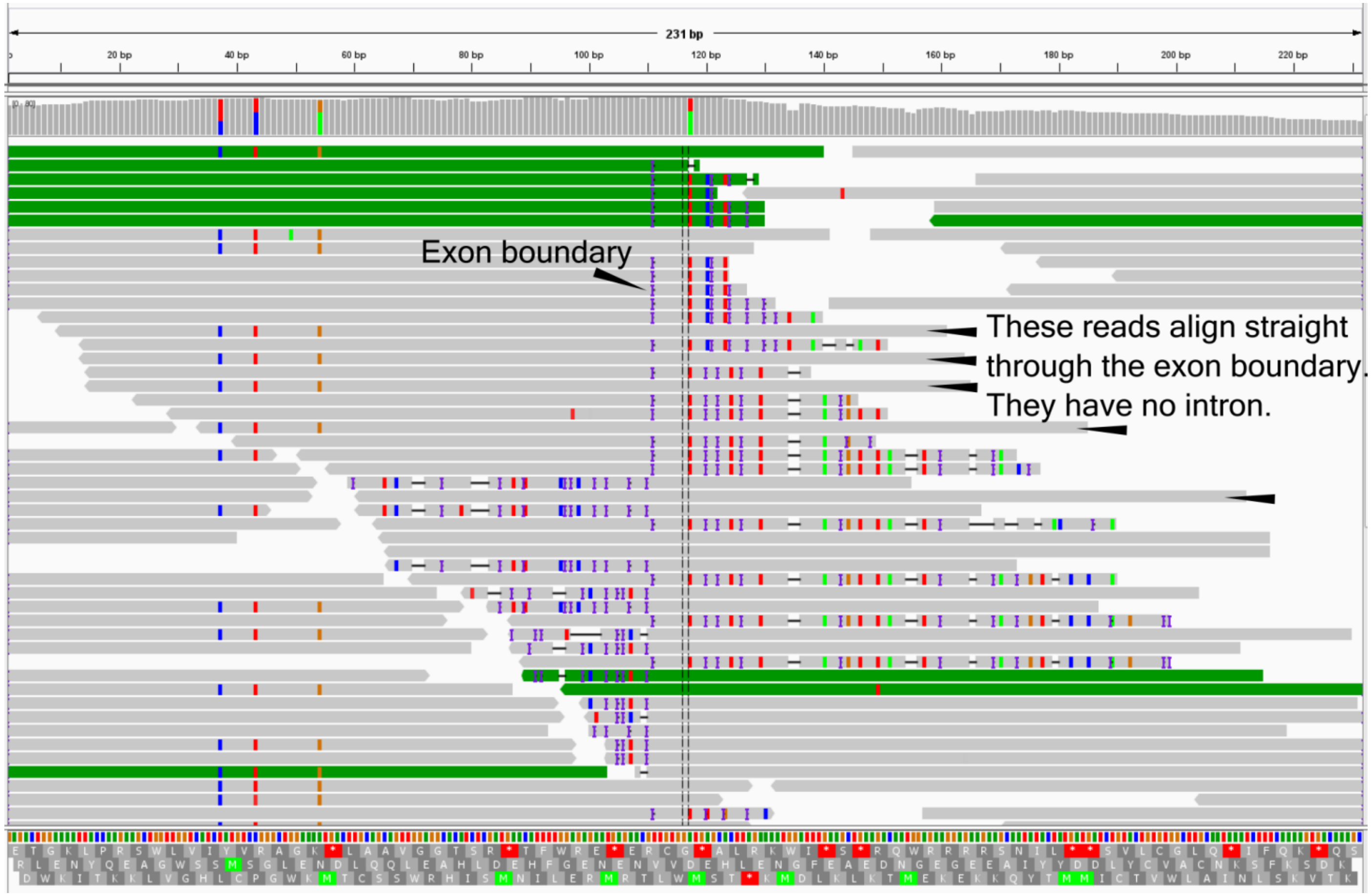
Appendix 2.3. The impact of fragmentation. An example where the automated pipeline Agalma split a cluster of homologous sequences into multiple ortholog clusters due to a fragmented transcript. (a) The gene tree produced from the Agalma homolog alignment that corresponded to the *L. gigantea* gene 197656. (b) A subset of the homolog alignment from which the Agalma gene tree was produced. This alignment has been broken up into two orthologous clusters due to the slightly overlapping sequences for *Asperitus stuartiae*. The colour blocks in (b) represent the two ortholog clusters resulting from the initial homolog cluster. The unhighlighted sequences represent clusters which did not pass the minimum taxa criteria of four or were not placed in an ortholog cluster. A subset of the sequences were removed by Agalma because they are identical to sequences for the same taxa which have been placed in one of the ortholog clusters.



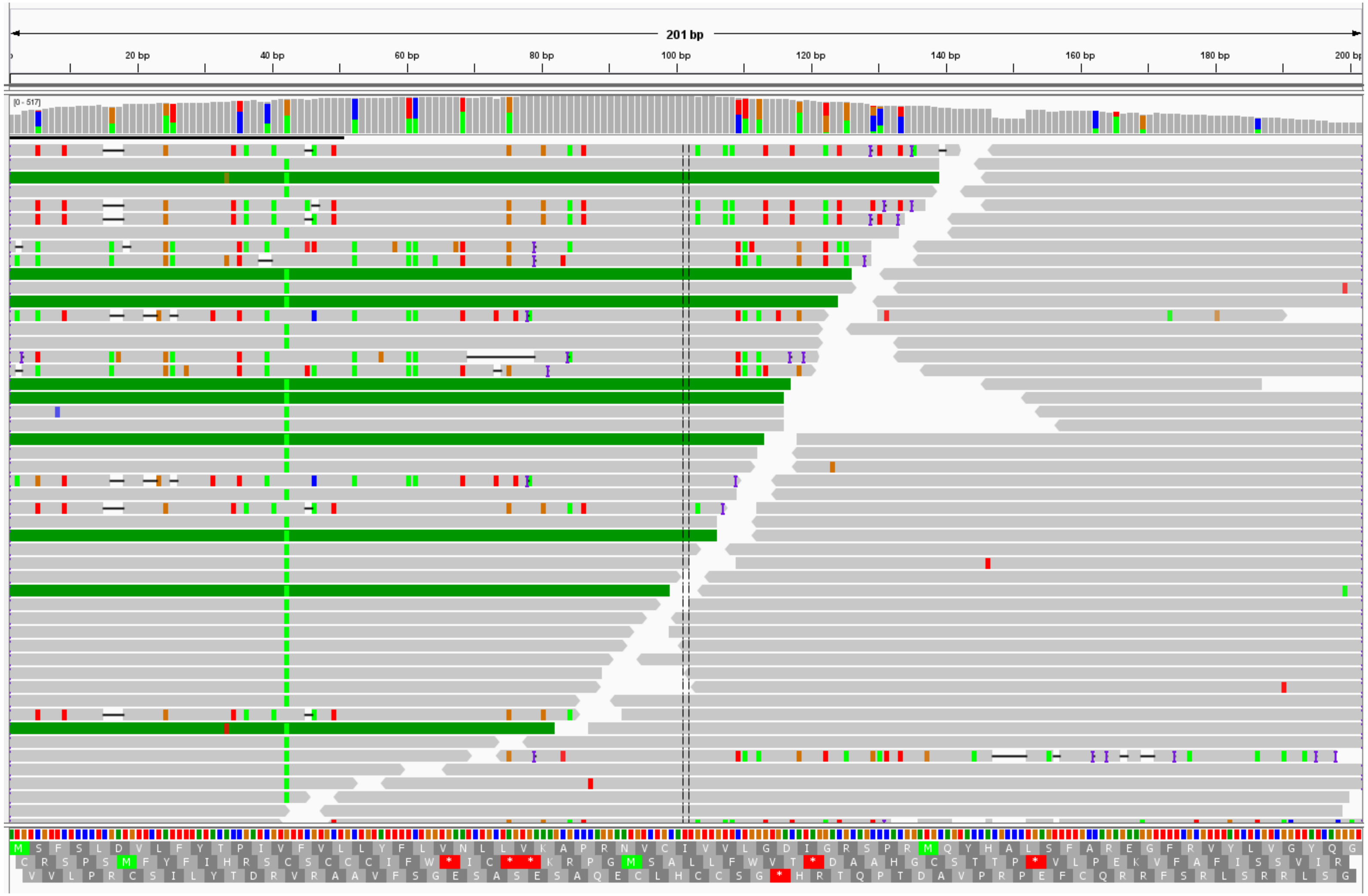
Appendix 2.4. An alignment of short reads resulting from exon capture sequencing to the reference sequence used to design the exon capture probes. The reads and the reference are from different individuals of the same species (*Sphareospira fraseri*). A heterozygous site is evident at 114bp.



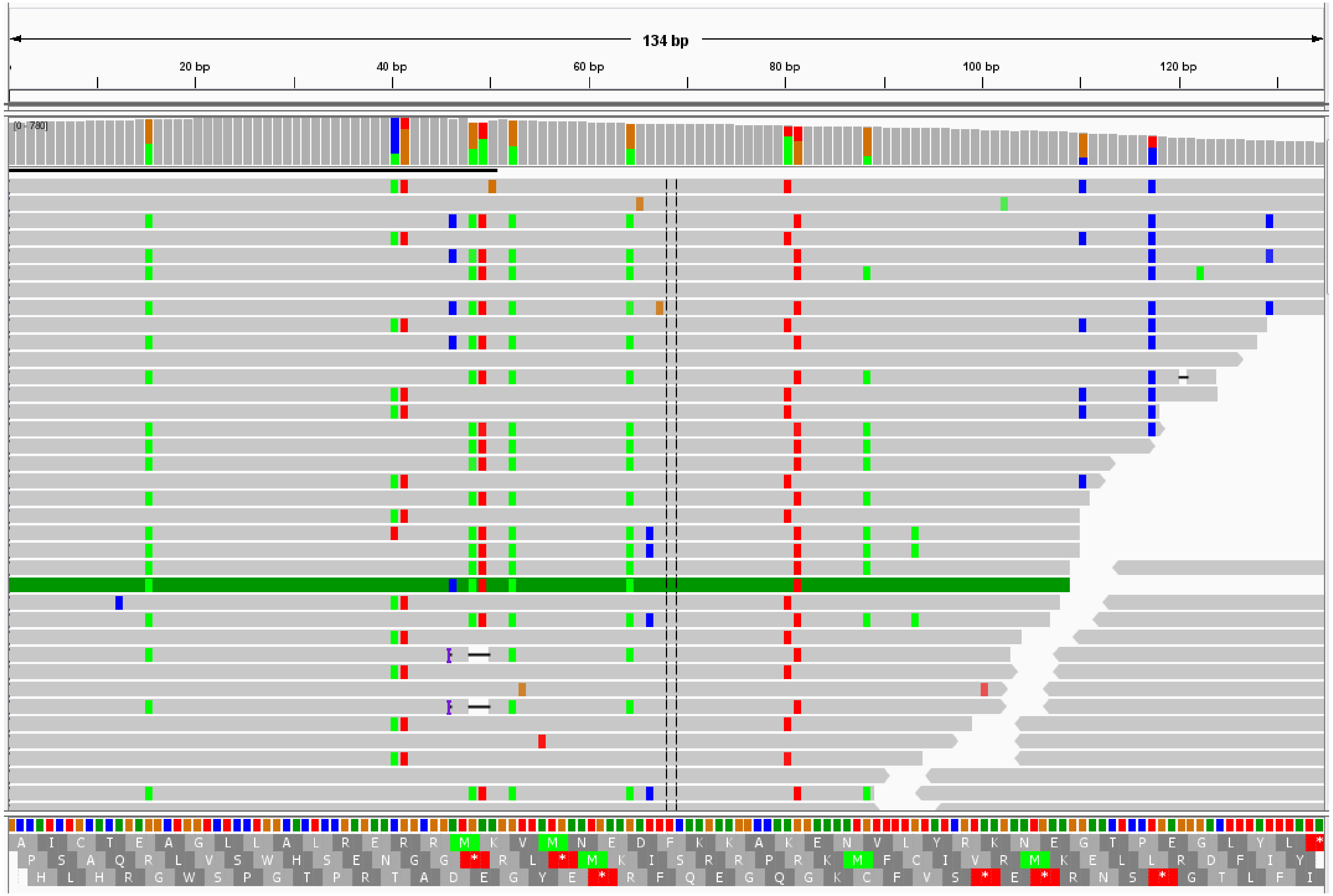
Appendix 2.5. A short read alignment showing a novel exon boundary. The probes were designed using Camaenidae transcriptomes but using the *Lottia gigantea* exon boundaries. Half the reads align to the first half of the transcriptome sequence but stop aligning at 119 bp. The other half starts aligning at 120bp.



Appendix 2.6. A read alignment which represents an apparent processed pseudogene. Most of the aligned reads stop aligning at the exon boundary at 111 bp. However, a number of reads do not contain any intronic sequence. This suggests they are processed pseudogenes where processed RNA molecules are reinserted into the genome. Therefore there are two copies of the gene from two different loci in the genome, one with introns and one without. The copy without intronic sequence is responsible for the three heterozygous sites in the first 60bp of the alignment.


201 bp
20 bp
40 bp
60 bp
80 bp
100 bp
120 bp
140 bp
160 bp
180 bp
200 bp
[0 - 517]
M S F S L D V L F Y T P I V F V L L Y F L V N L L V K A P R N V C I V V L G D I G R S P R M Q Y H A L S F A R E G F R V Y L V G Y Q G
C R S P S M F Y F I H R S C S C C C I F W * I C * * K R P G M S A L L F W V T * D A A H G C S T T P * V L P E K V F A F I S S V I R
V V L P R C S I L Y T D R V R A A V F S G E S A S E S A Q E C L H C C S G * H R T Q P T D A V P R P E F C Q R R F S R L S R R L S G

Appendix 2.7. A read alignment that represents an apparent pseudogene. The pseudogene sequences are characterised by a high proportion of sequence mismatches and indels which are out of frame.


134 bp
20 bp
40 bp
60 bp
80 bp
100 bp
120 bp
[0 - 780]
A I C T E A G L L A L R E R R M K V M N E D F K K A K E N V L Y R K N E G T P E G L Y L *
P S A Q R L V S W H S E N G G * R L * M K I S R R P R K M F C I V R M K E L L R D F I Y
H L H R G W S P G T P R T A D E G Y E * R F Q E G Q G K C F V S * E * R N S * G T L F I

Appendix 2.8. A read alignment which represents an apparent paralog. The paralogous sequences are characterised by a high proportion of sequence mismatches but the sequence still translates. However, it is also possible that this is a variable allele.

Appendix 2.9. 500 single copy, orthologous genes determined from 21 eupulmonate transcriptomes. The ‘all-by-all *L. gigantea* blast’ and ‘OMA’ columns notate whether the cluster containing the respective *L. gigantea* contains only a single *L. gigantea* sequence (1) or multiple (>1). In the case of the OMA results, zeros mean the *L. gigantea* sequence was not present in the OMA database at the time of download (3rd of September, 2014). The Kocot et al. column notates genes which are also present (1) in a molluscan phylogenomic dataset (Kocot et al. 2011).

***L. gigantea* gene ID**	**NCBI Gene ID**	***L. gigantea* family description**	**Length of alignment (bp)**	**No. taxa**	**Proportion data complete**	**No. exons (*L. gigantea*)**	**G/C composition**	**Kog ID**	**All-by-all *L. gigantea* blast**	**OMA**	**Kocot et al.**
52083	20251210	Pancreatic carboxypeptidases	1794	21	0.98	8	0.45	-	>1	1	0
56892	20251374	Glutathione S-transferase (GST), C-terminal domain	759	20	0.9	1	0.49	-	>1	>1	0
57629	20251402	DTD-like	531	21	0.92	1	0.4	-	1	1	0
58444	20251445	-	471	21	0.96	3	0.44	-	1	1	0
59279	20251499	Ribosomal protein L21p	777	20	0.92	5	0.43	KOG1686	1	1	0
62731	20251659	Canonical RBD	1725	21	0.92	7	0.43	KOG1548	1	1	0
63496	20251672	WD40-repeat	1869	20	0.82	1	0.45	KOG4547	1	1	0
64187	20251697	Cyclophilin (peptidylprolyl isomerase)	1950	21	0.99	3	0.41	KOG0415	1	1	0
65091	20251725	-	1110	21	0.91	8	0.43	KOG2989	>1	1	1
66926	20251798	RWD domain	744	21	0.96	7	0.41	-	1	1	0
68595	20251874	TBP-associated factors, TAFs	684	19	0.88	1	0.47	KOG0871	>1	1	0
75888	20252205	-	1346	20	0.9	1	0.41	KOG4260	1	1	0
76584	20252232	WWE domain	1089	21	0.76	1	0.47	KOG0824	1	1	0
82054	20252524	-	2301	21	0.83	13	0.44	-	1	>1	0
82870	20252548	Cation efflux protein transmembrane domain-like	1290	21	0.96	9	0.47	-	>1	1	0
87019	20252715	RNase P subunit p30	1020	21	0.87	6	0.4	KOG2363	1	1	0
87302	20252729	-	702	20	0.81	1	0.44	KOG4515	1	1	0
89339	20252843	-	477	20	0.94	1	0.4	-	1	1	0
93548	20252975	Chaperone J-domain	1893	21	0.96	1	0.41	-	1	1	1
93698	20252980	RING finger domain, C3HC4	2037	21	0.9	10	0.45	-	1	1	0
94819	20253018	-	1098	21	0.77	4	0.39	KOG3941	1	1	0
95075	20253032	Tyrosine-dependent oxidoreductases	1245	21	0.99	1	0.45	KOG2865	1	1	0
95085	20253035	Nucleotide and nucleoside kinases	1269	21	0.98	1	0.41	KOG3877	1	1	0
96262	20253092	-	825	20	0.89	2	0.43	KOG3331	1	1	0
96885	20253127	-	930	20	0.95	1	0.46	-	>1	>1	0
96984	20253132	Ankyrin repeat	810	21	0.84	1	0.41	-	1	1	0
97481	20253162	Ribosomal L11/L12e N-terminal domain	594	21	0.96	3	0.43	KOG3257	1	1	1
98069	20253204	Canonical RBD	759	19	0.76	2	0.37	KOG3152	1	1	1
98370	20253233	TBP-associated factors, TAFs	1083	20	0.86	5	0.41	KOG1659	>1	1	0
98700	20253253	-	1563	21	0.78	2	0.43	KOG4461	1	1	0
100771	20229525	-	315	21	0.97	1	0.42	-	1	1	0
101578	20229558	-	429	21	0.99	3	0.42	-	1	1	0

L. gigantea gene ID	NCBI Gene ID	*L. gigantea* family description	Length of alignment (bp)	No. taxa	Proportion data complete	No. exons (*L. gigantea*)	G/C composition	Kog ID	All-by-all *L. gigantea* blast	OMA	Kocot et al.
101919	20229566	Ribosomal protein S4	549	21	0.99	4	0.41	KOG4655	1	1	1
102151	20229573	-	1632	19	0.64	2	0.44	KOG3748	1	1	0
102242	20229587	FKBP immunophilin/proline isomerase	1341	21	0.98	7	0.42	-	>1	1	0
102889	20229670	-	780	21	0.95	6	0.43	KOG2678	1	1	0
103111	20229701	Nicotinic receptor ligand binding domain-like	1833	20	0.78	6	0.43	-	1	>1	0
103333	20229737	Ankyrin repeat	1569	21	0.58	3	0.48	-	1	1	0
103560	20229760	G proteins	1104	21	0.96	9	0.45	KOG1487	>1	>1	0
103711	20229775	WD40-repeat	1551	21	0.97	15	0.48	KOG0289	>1	1	0
104548	20229880	Ngr ectodomain-like	936	21	0.88	4	0.42	KOG0473	1	1	0
105900	20230042	-	834	19	0.89	1	0.4	-	1	1	0
105979	20230049	Thiolase-related	1254	21	0.99	11	0.44	-	>1	>1	0
106290	20230085	-	1242	18	0.79	2	0.51	KOG2625	1	1	0
106520	20230123	-	636	21	0.99	5	0.45	KOG4040	1	1	0
107347	20230207	Glutathione S-transferase (GST), C-terminal domain	1038	21	0.89	7	0.44	KOG2903	1	1	0
108143	20230309	-	882	19	0.83	2	0.41	KOG2873	1	1	0
108249	20230324	Cold shock DNA-binding domain-like	2478	21	0.99	15	0.47	-	1	1	0
108291	20230329	-	1668	21	0.93	12	0.44	KOG4506	1	1	0
108695	20230384	Cytochrome P450	1512	20	0.87	1	0.44	-	>1	>1	0
108787	20230391	Rieske iron-sulfur protein (ISP)	840	20	0.94	3	0.47	KOG1671	1	1	0
108795	20230393	Ribosomal protein L3	1167	21	0.97	8	0.45	KOG3141	1	1	0
108932	20230411	Armadillo repeat	1686	21	0.98	15	0.44	KOG2734	1	1	0
108975	20230415	-	1350	20	0.89	11	0.42	KOG2552	1	1	0
109203	20230439	Type I phosphomannose isomerase	1296	21	0.91	7	0.47	KOG2757	1	1	0
109608	20230490	RNA polymerase subunit RBP8	450	19	0.89	4	0.49	KOG3400	1	1	1
109671	20230494	Rhomboid-like	1050	21	0.84	6	0.45	KOG4463	1	1	0
109924	20230524	Ubiquitin carboxyl-terminal hydrolase, UCH	6261	21	0.63	11	0.46	KOG1887	1	1	0
109978	20230534	N-terminal, heterodimerisation domain of RBP7 (RpoE)	657	18	0.75	7	0.43	KOG3297	1	1	1
110039	20230540	-	1671	21	0.99	13	0.43	-	1	1	0
110063	20230545	Canonical RBD	1407	21	0.96	7	0.5	KOG0153	1	1	0
110501	20230592	Phosphoribosylpyrophosphate synthetase-like	1086	21	0.9	9	0.44	KOG1503	>1	>1	0
110519	20230594	Nuclear movement domain	1191	20	0.91	7	0.42	-	1	1	0
111616	20230753	WD40-repeat	1374	21	0.86	6	0.47	KOG0302	>1	1	1
111926	20230794	Papain-like	1653	21	0.98	8	0.45	-	1	1	0
111940	20230795	Fe-only hydrogenase	1449	21	0.82	9	0.44	KOG2439	1	1	0
112115	20230818	-	774	21	0.9	1	0.42	KOG4380	1	1	0
113043	20230913	-	1257	21	0.94	9	0.45	KOG0972	1	1	0
113682	20230996	Calmodulin-like	1311	21	0.84	6	0.4	KOG4251	>1	1	0
113844	20231013	Ribosomal protein S2	972	21	0.93	1	0.44	KOG0832	1	1	0
114038	20231039	ABC transporter ATPase domain-like	2235	21	0.97	15	0.44	KOG0066	>1	>1	0
114242	20231068	-	1083	21	0.98	7	0.45	KOG1349	>1	1	0
114414	20231088	5' to 3' exonuclease catalytic domain	1158	20	0.88	10	0.42	-	>1	>1	1
114503	20231101	-	564	20	0.89	4	0.42	KOG4093	1	1	1

L. gigantea gene ID	NCBI Gene ID	*L. gigantea* family description	Length of alignment (bp)	No. taxa	Proportion data complete	No. exons (*L. gigantea*)	G/C composition	Kog ID	All-by-all *L. gigantea* blast	OMA	Kocot et al.
114637	20231119	Coiled-coil domain of nucleotide exchange factor GrpE	723	21	0.97	1	0.4	KOG3003	1	1	1
114808	20231139	WD40-repeat	1398	21	0.93	7	0.44	KOG2055	1	1	0
114823	20231142	L-arabinose binding protein-like	3174	18	0.77	2	0.44	-	1	0	0
115165	20231187	Nitrogenase iron protein-like	1227	21	0.97	7	0.43	KOG1532	>1	>1	0
115322	20231205	-	1074	19	0.72	3	0.44	-	1	1	0
115372	20231216	Integrin A (or I) domain	1200	21	0.99	10	0.46	KOG2884	1	1	1
115671	20231255	-	1125	18	0.63	4	0.46	KOG2612	1	1	0
115736	20231261	FHA domain	2256	21	0.97	8	0.42	KOG1881	1	1	0
115833	20231271	-	1803	21	0.97	12	0.44	-	1	1	0
116525	20231360	-	642	19	0.8	1	0.46	-	1	1	0
116743	20231389	Ubiquitin activating enzymes (UBA)	1338	21	0.84	15	0.45	KOG2015	>1	1	0
117320	20231480	Pseudouridine synthase I TruA	1800	20	0.8	13	0.42	KOG2553	>1	1	0
117522	20231508	2'-5'-oligoadenylate synthetase 1, OAS1, N-terminal domain	1230	21	0.97	11	0.48	KOG3793	>1	1	0
117888	20231551	FHA domain	954	20	0.78	2	0.44	KOG1882	>1	1	0
118510	20231623	Calmodulin-like	558	20	0.88	6	0.42	-	1	1	0
118545	20231629	Mannose 6-phosphate receptor domain	1506	21	0.92	12	0.44	-	>1	1	0
118615	20231636	-	774	18	0.74	1	0.45	-	1	1	0
118654	20231641	-	384	21	0.99	2	0.41	KOG3450	1	1	0
118845	20231666	Class II aminoacyl-tRNA synthetase (aaRS)-like, catalytic domain	1581	21	0.95	14	0.46	KOG0556	>1	>1	0
119041	20231691	-	1283	21	0.95	10	0.43	KOG3973	1	1	0
119246	20231718	-	516	20	0.87	1	0.39	KOG4253	1	1	0
119784	20231776	FHA domain	2610	21	0.76	7	0.44	-	1	1	0
119936	20231797	Dihydrodipicolinate reductase-like	858	19	0.78	1	0.46	-	1	1	0
120210	20231836	Hypothetical esterase YJL068C	867	21	0.99	6	0.47	KOG3101	1	1	1
120815	20231900	SH3-domain	1350	20	0.79	3	0.44	KOG3875	1	1	0
121553	20231981	UbiE/COQ5-like	1050	21	0.96	10	0.45	KOG2940	1	1	0
121872	20232038	ApaG-like	1077	21	0.99	9	0.48	-	1	1	0
122655	20232140	mRNA cap (Guanine N-7) methyltransferase	1215	21	0.99	8	0.42	KOG1975	1	1	0
123204	20232207	-	1383	21	0.96	1	0.46	KOG2703	1	1	0
123335	20232229	-	1659	21	0.85	1	0.42	KOG2459	1	1	0
123420	20232240	WD40-repeat	1737	21	0.92	9	0.45	-	1	1	0
123440	20232243	Plant O-methyltransferase, C-terminal domain	651	19	0.83	1	0.42	-	1	1	0
123644	20232275	-	1383	20	0.89	2	0.43	-	1	1	0
124007	20232317	Class I aminoacyl-tRNA synthetases (RS), catalytic domain	1248	20	0.88	8	0.43	KOG2145	1	1	1
124038	20232320	Activator of Hsp90 ATPase, Aha1	1101	21	1	8	0.42	KOG2936	1	1	0
126181	20232564	TRAPP components	570	21	0.99	7	0.41	KOG3315	1	1	1
126234	20232570	Transcriptional regulator IclR, N-terminal domain	1320	21	1	12	0.46	KOG2758	1	1	0
126388	20232590	-	1914	21	1	18	0.42	-	>1	>1	0
126569	20232609	RIO1-like kinases	1635	21	0.87	10	0.43	KOG2270	>1	1	1
127279	20232693	Inositol monophosphatase/fructose-1,6-bisphosphatase-like	1102	21	0.92	7	0.49	-	>1	>1	0
127623	20232733	TIM44-like	1413	21	0.99	13	0.39	KOG2580	1	1	0

L. gigantea gene ID	NCBI Gene ID	*L. gigantea* family description	Length of alignment (bp)	No. taxa	Proportion data complete	No. exons (*L. gigantea*)	G/C composition	Kog ID	All-by-all *L. gigantea* blast	OMA	Kocot et al.
127678	20232741	-	582	21	0.98	2	0.41	KOG3241	1	1	0
129046	20232908	-	945	18	0.71	7	0.39	KOG1297	1	1	0
129614	20232977	-	690	19	0.83	3	0.41	-	1	1	0
129894	20233006	IF2B-like	2289	21	0.91	9	0.44	KOG1467	>1	>1	0
131897	20233233	HkH motif-containing C2H2 finger	1029	20	0.91	7	0.48	-	1	1	0
131935	20233237	PDI-like	1305	21	0.94	11	0.45	-	>1	1	0
132027	20233253	La domain	1347	21	0.98	9	0.43	KOG4213	>1	1	1
132099	20233268	Transcriptional factor domain	384	21	0.99	5	0.45	KOG2691	>1	>1	1
132169	20233278	Extended AAA-ATPase domain	1197	21	0.99	11	0.47	KOG0651	>1	>1	1
132288	20233297	Tetratricopeptide repeat (TPR)	1353	19	0.84	3	0.44	-	>1	1	0
132351	20233306	Functional domain of the splicing factor Prp18	1065	21	0.98	10	0.44	KOG2808	1	1	0
132697	20233333	Tetratricopeptide repeat (TPR)	1200	21	0.87	1	0.42	-	1	1	0
132718	20233338	-	1041	21	0.93	8	0.45	-	1	1	0
133686	20233467	MIF4G domain-like	3069	21	0.99	20	0.47	-	>1	1	0
133988	20233505	FtsH protease domain-like	2490	21	0.93	16	0.45	-	>1	>1	0
134046	20233511	Phosphatidylethanolamine binding protein	1245	21	0.94	6	0.43	-	>1	1	0
134665	20233586	Tyrosine-dependent oxidoreductases	972	20	0.86	8	0.45	-	>1	>1	1
136567	20233821	PP2C-like	1197	20	0.89	4	0.44	-	1	1	0
136673	20233828	PCI domain (PINT motif)	2832	21	0.98	22	0.44	KOG1076	1	1	0
136847	20233848	Ribosomal protein L36	618	21	0.83	1	0.44	KOG4122	1	0	0
137312	20233894	-	759	21	1	7	0.46	-	>1	>1	0
137823	20233954	Elongation factor Ts (EF-Ts), dimerisation domain	1110	20	0.9	5	0.44	KOG1071	1	1	0
137826	20233955	Aquaporin-like	917	21	0.98	2	0.49	-	>1	>1	0
137913	20233969	-	1005	21	0.85	7	0.46	KOG0917	1	1	0
138363	20234044	-	2016	21	0.89	13	0.43	KOG2491	1	1	0
138864	20234117	Chaperone J-domain	1074	21	0.97	9	0.43	-	>1	>1	0
139266	20234166	Pseudouridine synthase II TruB	1602	21	0.98	9	0.45	-	1	1	0
140883	20234370	-	3051	21	0.92	16	0.41	-	1	1	0
141157	20234398	-	1338	21	0.92	10	0.43	KOG3871	1	1	0
141173	20234400	Tetratricopeptide repeat (TPR)	2796	21	0.94	25	0.45	KOG0495	1	1	0
141873	20234501	-	603	21	0.91	4	0.45	KOG4067	1	1	0
141904	20234504	t-snare proteins	657	21	0.91	1	0.41	KOG1666	1	1	0
142111	20234538	Dimeric isocitrate & isopropylmalate dehydrogenases	1251	21	0.99	10	0.42	-	>1	>1	0
142233	20234548	ABC transporter ATPase domain-like	1836	21	0.99	9	0.44	KOG0063	1	1	0
142681	20234597	Class I aldolase	990	21	0.99	9	0.43	KOG2772	1	1	1
144016	20234776	Extended AAA-ATPase domain	1230	21	0.97	10	0.4	KOG3928	1	1	0
144966	20234897	Nop domain	1776	21	1	13	0.45	-	>1	1	1
150024	20235497	Group II chaperonin (CCT, TRIC), ATPase domain	1632	21	0.97	10	0.47	KOG0357	>1	>1	1
150117	20235504	Ribosomal protein L16p	774	21	0.97	4	0.4	KOG3422	1	1	0
150160	20235507	Hydroxyisobutyrate and 6-phosphogluconate dehydrogenase domain	1461	21	0.94	13	0.46	KOG2653	1	1	0
150592	20235518	Mitochondrial ribosomal protein L51/S25/CI-B8 domain	576	20	0.94	3	0.41	KOG3445	1	1	0

***L. gigantea* gene ID**	**NCBI Gene ID**	***L. gigantea* family description**	**Length of alignment (bp)**	**No. taxa**	**Proportion data complete**	**No. exons (*L. gigantea*)**	**G/C composition**	**Kog ID**	**All-by-all *L. gigantea* blast**	**OMA**	**Kocot et al.**
150746	20235522	-	513	21	0.97	3	0.45	KOG4697	1	1	0
152785	20235782	-	735	21	0.95	5	0.44	KOG2659	1	1	0
152961	20235823	Extended AAA-ATPase domain	1005	21	0.94	11	0.44	KOG0990	>1	>1	0
153055	20235853	-	1353	20	0.89	11	0.43	-	>1	>1	1
154620	20236349	Ferritin	594	21	0.99	5	0.43	KOG4061	1	1	1
154698	20236366	-	1203	21	0.95	6	0.45	-	1	1	0
155494	20236729	TRAPP components	543	20	0.91	5	0.42	KOG3330	1	1	1
155607	20236759	GS domain	1038	21	0.99	9	0.47	KOG1667	1	1	0
156082	20236922	Canonical RBD	1296	21	0.96	4	0.41	KOG0126	1	1	1
156500	20237018	Ribosomal protein L7/12, C-terminal domain	612	21	0.99	5	0.4	KOG1715	1	1	1
156505	20237020	Tetratricopeptide repeat (TPR)	1290	15	0.64	6	0.49	KOG1941	1	1	0
156651	20237053	Canonical RBD	1254	21	0.96	5	0.47	-	1	0	0
156673	20237059	MPP-like	1464	21	1	14	0.48	KOG2583	>1	>1	0
156843	20237128	WD40-repeat	1626	20	0.82	11	0.46	-	>1	>1	0
156936	20237148	Clathrin coat assembly domain	579	21	0.99	7	0.41	KOG3343	1	1	1
157797	20237491	Thiolase-related	1449	21	0.96	13	0.46	KOG1392	1	>1	0
157909	20237550	Cytochrome c oxidase subunit E	534	21	0.99	5	0.44	KOG4077	1	0	1
157925	20237555	-	576	21	0.96	8	0.41	-	1	1	0
159336	20238033	-	966	21	0.95	12	0.45	KOG2487	1	1	0
159839	20238167	-	417	21	0.92	4	0.4	-	1	1	0
160105	20238295	Rhomboid-like	1089	21	0.9	9	0.44	KOG2980	1	1	0
160203	20238332	Fumarate reductase/Succinate dehydogenase iron-sulfur protein, C-terminal domain	906	21	0.99	8	0.46	KOG3049	1	1	1
160712	20238480	beta-Lactamase/D-ala carboxypeptidase	1791	21	0.93	5	0.46	-	1	>1	0
162156	20238930	Mitochondrial carrier	1302	21	0.91	10	0.46	KOG2954	1	1	0
164315	20239708	-	2136	21	1	15	0.46	-	>1	>1	0
165034	20239895	-	1389	21	0.91	9	0.48	-	1	0	0
165279	20239977	Tandem AAA-ATPase domain	5949	21	0.86	18	0.47	-	>1	1	0
165337	20239988	Protein prenyltransferases	993	20	0.91	9	0.47	KOG0366	>1	1	0
165683	20240157	-	666	20	0.83	5	0.43	KOG3339	1	1	1
166223	20240287	Proteasome subunits	843	21	1	8	0.47	KOG0175	>1	>1	0
166906	20240477	Quinoprotein alcohol dehydrogenase-like	1470	21	0.95	11	0.43	KOG0646	1	1	0
167298	20240580	ABC transporter ATPase domain-like	1917	21	1	12	0.44	KOG0927	>1	>1	0
167341	20240595	Sm motif of small nuclear ribonucleoproteins, SNRNP	276	21	0.99	5	0.41	KOG1775	1	1	0
167800	20240698	WD40-repeat	1206	21	1	14	0.46	KOG1523	1	1	1
168344	20240850	Chaperone J-domain	426	21	0.96	1	0.47	KOG0723	1	1	0
168577	20240926	MPP-like	3090	21	0.79	11	0.45	-	1	1	0
168884	20240994	Rna1p (RanGAP1), N-terminal domain	1743	21	0.95	17	0.43	KOG1909	1	1	0
170554	20241450	-	831	17	0.73	6	0.44	-	1	1	0
171554	20241812	YjjX-like	1098	21	0.88	5	0.44	-	1	1	0
171717	20241869	Predicted hydrolases Cof	771	20	0.9	7	0.44	KOG3189	1	1	1
172374	20242076	Ribonuclease PH domain 1-like	1275	21	0.95	10	0.43	KOG1614	>1	>1	1
172563	20242128	Tandem AAA-ATPase domain	1683	21	0.86	14	0.44	KOG0344	>1	>1	0

L. gigantea gene ID	NCBI Gene ID	*L. gigantea* family description	Length of alignment (bp)	No. taxa	Proportion data complete	No. exons (*L. gigantea*)	G/C composition	Kog ID	All-by-all *L. gigantea* blast	OMA	Kocot et al.
173015	20242262	Divalent ion tolerance proteins CutA (CutA1)	495	20	0.89	5	0.42	KOG3338	1	1	0
173162	20242302	Tandem AAA-ATPase domain	1710	21	0.99	17	0.45	KOG0346	1	>1	0
173766	20242519	Glyoxalase II (hydroxyacylglutathione hydrolase)	783	21	0.99	7	0.45	-	>1	>1	0
174207	20242655	Arrestin/Vps26-like	1089	21	0.86	1	0.47	-	>1	>1	0
174413	20242724	Ferrochelatase	1290	21	0.87	9	0.45	KOG1321	1	1	0
175212	20243078	TBP-associated factors, TAFs	1887	21	0.89	3	0.43	-	>1	1	0
177207	20244193	Tetratricopeptide repeat (TPR)	3609	21	0.94	20	0.44	KOG2002	>1	1	0
177471	20244211	Cytidylyltransferase	1191	21	0.94	13	0.44	KOG2803	>1	1	1
178117	20244260	Ubiquitin activating enzymes (UBA)	1917	21	0.92	17	0.46	KOG2013	>1	1	0
178463	20244279	-	1374	21	0.99	10	0.46	KOG2330	1	1	0
178649	20244291	Histone deacetylase, HDAC	1662	21	0.98	13	0.44	-	>1	>1	0
179268	20244331	Canonical RBD	717	19	0.74	1	0.4	-	1	1	1
179278	20244333	Chaperone J-domain	1809	21	0.94	12	0.42	KOG0717	>1	1	0
180428	20244403	FKBP immunophilin/proline isomerase	465	21	0.99	3	0.45	KOG3259	1	1	0
181127	20244423	-	996	21	0.98	6	0.44	KOG1560	1	1	0
181293	20244435	STM3548-like	1341	21	0.99	13	0.45	KOG3861	1	1	0
181363	20244439	PCI domain (PINT motif)	1332	21	0.99	13	0.4	KOG1464	>1	1	0
181759	20244457	YfcE-like	549	21	1	4	0.45	KOG3325	1	1	1
182042	20244469	UbiE/COQ5-like	852	21	0.99	11	0.46	-	1	1	1
182182	20244479	Extended AAA-ATPase domain	1068	21	0.91	8	0.47	-	>1	>1	1
182203	20244482	GABA-aminotransferase-like	1509	21	0.93	1	0.44	KOG1358	1	1	0
182398	20244492	MED7 hinge region	681	20	0.82	2	0.42	KOG0570	1	1	1
182505	20244496	-	1341	21	0.92	8	0.41	KOG2927	1	1	0
182822	20244515	WD40-repeat	1044	21	0.99	7	0.42	KOG0278	>1	1	0
183100	20244525	-	468	21	0.99	3	0.48	KOG3391	1	1	0
183373	20244541	Translational machinery components	1239	21	1	13	0.43	KOG1697	1	1	0
183804	20244566	-	507	20	0.91	2	0.45	-	1	1	0
184295	20244592	G proteins	1107	21	0.95	14	0.42	KOG1486	>1	>1	1
184303	20244593	Alcohol dehydrogenase-like, N-terminal domain	1044	21	0.9	5	0.43	-	>1	>1	0
184615	20244611	Leucine aminopeptidase, C-terminal domain	1572	21	0.97	11	0.5	-	>1	>1	0
185379	20244649	-	933	21	0.96	1	0.43	KOG1563	1	1	1
185419	20244653	-	495	21	0.99	3	0.48	-	1	1	0
185481	20244655	Sedlin (SEDL)	429	21	0.99	1	0.37	KOG3487	1	1	1
185700	20244662	spliceosomal protein U5-15Kd	429	19	0.89	1	0.42	-	>1	1	0
185777	20244668	RING finger domain, C3HC4	1017	21	0.97	10	0.42	KOG1813	1	1	1
186175	20244690	PP2C-like	1644	21	0.88	7	0.46	-	1	1	0
186348	20244698	-	1614	21	0.99	1	0.49	KOG3786	1	1	0
186799	20244718	Sm motif of small nuclear ribonucleoproteins, SNRNP	291	17	0.76	4	0.46	KOG1784	>1	1	0
186812	20244719	-	2052	21	0.87	2	0.44	-	1	1	0
187118	20244734	Ribosome recycling factor, RRF	888	21	0.95	1	0.41	KOG4759	1	1	0
187172	20244736	ZZ domain	1389	20	0.91	8	0.44	-	1	1	0
187174	20244737	Polypeptide N-	1023	21	0.79	1	0.4	-	>1	>1	0

***L. gigantea* gene ID**	**NCBI Gene ID**	***L. gigantea* family description**	**Length of alignment (bp)**	**No. taxa**	**Proportion data complete**	**No. exons (*L. gigantea*)**	**G/C composition**	**Kog ID**	**All-by-all *L. gigantea* blast**	**OMA**	**Kocot et al.**
		acetylgalactosaminyltransferase 1, N-terminal domain									
187336	20244747	PaaD-like	477	21	0.97	5	0.44	-	>1	>1	1
187615	20244752	PBS lyase HEAT-like repeat	2748	21	0.99	20	0.47	KOG2005	>1	1	0
187625	20244753	IF2B-like	918	21	0.96	9	0.43	KOG1466	>1	1	1
188753	20244808	Leucine rich effector protein YopM	1689	21	0.93	5	0.42	-	>1	1	0
189149	20244834	PBS lyase HEAT-like repeat	963	21	0.96	4	0.45	KOG0567	1	1	1
189380	20244842	Proteasome subunits	708	21	0.99	7	0.47	-	>1	>1	1
189619	20244849	-	1464	21	0.97	8	0.45	-	1	1	0
189635	20244850	Cap-Gly domain	1635	21	0.99	14	0.44	-	>1	>1	0
189885	20244863	Armadillo repeat	1413	20	0.91	1	0.42	KOG4199	1	1	0
190068	20244872	Inorganic pyrophosphatase	888	21	0.97	11	0.43	KOG1626	1	1	1
190227	20244881	RNA polymerase II subunit RBP4 (RpoF)	441	21	1	2	0.41	KOG2351	1	1	0
190525	20244900	-	603	21	0.99	1	0.41	-	1	1	0
190618	20244908	Calmodulin-like	612	21	0.99	1	0.47	-	>1	>1	0
191449	20244933	Quinoprotein alcohol dehydrogenase-like	1830	21	0.97	16	0.44	KOG0318	1	>1	0
191571	20244941	-	1053	21	1	6	0.42	KOG1556	>1	1	1
191967	20244959	Clathrin adaptor core protein	2640	21	1	25	0.48	KOG1078	1	1	0
192242	20244969	Supernatant protein factor (SPF), C-terminal domain	723	20	0.89	1	0.45	-	>1	>1	0
192266	20244972	-	528	21	0.95	3	0.45	-	>1	1	0
192289	20244973	TatD Mg-dependent DNase-like	885	21	0.9	12	0.42	-	>1	>1	0
192615	20244990	GABA-aminotransferase-like	1146	21	1	10	0.43	KOG2790	1	1	1
193380	20245026	Thioltransferase	675	20	0.93	1	0.42	-	>1	>1	0
193840	20245043	RPB6	408	21	0.95	1	0.45	-	>1	1	0
194459	20245079	HIT (HINT, histidine triad) family of protein kinase-interacting proteins	546	20	0.86	1	0.44	KOG4359	1	1	0
195151	20245115	N-acetyl transferase, NAT	1392	21	0.94	2	0.41	-	>1	>1	0
195390	20245127	-	534	20	0.91	1	0.43	KOG3269	1	1	0
195467	20245130	JAB1/MPN domain	942	21	0.99	10	0.49	-	>1	1	0
195680	20245143	-	1521	21	0.99	16	0.45	KOG2636	>1	1	1
196086	20245167	dUTPase-like	447	19	0.85	6	0.47	KOG3370	1	1	1
196232	20245179	NagB-like	741	20	0.89	1	0.47	KOG3147	>1	1	0
196504	20245195	Arp2/3 complex 21 kDa subunit ARPC3	534	21	1	7	0.45	KOG3155	1	1	1
196653	20245204	Insert subdomain of RNA polymerase alpha subunit	1065	21	0.95	10	0.42	KOG1521	>1	1	1
196769	20245210	U-box	891	21	0.96	1	0.43	KOG3039	1	1	0
196960	20245219	EDF1-like	459	21	1	5	0.46	KOG3398	1	1	1
197181	20245234	Calmodulin-like	456	21	0.99	4	0.46	-	1	>1	0
197242	20245237	Extended AAA-ATPase domain	1326	21	0.99	12	0.45	KOG0726	>1	>1	1
197656	20245261	Insert subdomain of RNA polymerase alpha subunit	837	21	0.98	8	0.44	KOG1522	>1	1	1
198443	20245302	WD40-repeat	1542	21	1	12	0.46	-	>1	1	0
198678	20245314	Glycosyl transferases group 1	1497	20	0.85	3	0.44	KOG1387	>1	1	0
199122	20245333	N-acetyl transferase, NAT	522	21	1	1	0.44	KOG3234	>1	>1	1
199820	20245367	Group II chaperonin (CCT, TRIC), ATPase domain	1653	21	1	14	0.44	KOG0358	>1	>1	1

L. gigantea gene ID	NCBI Gene ID	*L. gigantea* family description	Length of alignment (bp)	No. taxa	Proportion data complete	No. exons (*L. gigantea*)	G/C composition	Kog ID	All-by-all *L. gigantea* blast	OMA	Kocot et al.
200332	20245402	WD40-repeat	984	21	1	9	0.46	KOG0643	>1	1	1
200807	20245426	Ribonuclease H	918	20	0.87	8	0.41	KOG2299	1	1	1
200903	20245434	PCI domain (PINT motif)	1173	21	0.97	8	0.43	KOG0687	>1	1	1
201479	20245465	Protein prenylyltransferase	993	21	0.87	8	0.44	KOG0530	>1	1	1
201543	20245471	Extended AAA-ATPase domain	1371	21	0.99	1	0.44	KOG1942	>1	>1	1
201648	20245479	Hypothetical protein PH1602	1515	21	0.99	1	0.48	KOG3833	1	1	0
201812	20245485	Branched-chain alpha-keto acid dehydrogenase PP module	1167	21	1	9	0.47	KOG0225	>1	>1	0
201859	20245490	Brix domain	882	20	0.87	1	0.43	KOG2781	>1	1	1
201929	20245495	Canonical RBD	1518	21	0.99	9	0.48	KOG0131	>1	1	0
201946	20245496	Citrate synthase	3321	21	1	27	0.44	KOG1254	>1	1	0
202318	20245522	WD40-repeat	1047	21	1	9	0.45	KOG0265	>1	>1	0
202320	20245523	-	930	21	0.85	4	0.46	-	1	1	0
202405	20245527	-	519	21	0.92	3	0.42	KOG4808	1	1	0
202488	20245534	DNA-repair protein XRCC1	1917	21	0.97	16	0.44	KOG2481	1	1	1
202604	20245541	VPS36 N-terminal domain-like	1173	21	0.96	13	0.43	KOG2760	1	1	0
202995	20245567	Prefoldin	564	21	0.97	1	0.42	KOG3313	1	1	1
203282	20245592	UbiE/COQ5-like	945	21	0.98	6	0.46	KOG4020	1	1	0
203414	20245602	PCI domain (PINT motif)	1227	21	0.99	11	0.42	KOG1497	1	1	0
203563	20245618	Ribosomal protein L24e	498	20	0.94	4	0.43	KOG1723	1	1	1
203656	20245622	Extended AAA-ATPase domain	1308	21	1	12	0.49	KOG0729	>1	>1	1
203677	20245625	-	474	21	0.95	5	0.4	-	>1	>1	0
203791	20245643	-	552	21	1	1	0.46	KOG4835	1	1	0
203870	20245652	-	1209	19	0.89	6	0.46	-	>1	>1	0
203990	-	gamma-carbonic anhydrase-like	558	21	0.95	6	0.45	KOG3121	1	1	0
204047	20245678	NHL repeat	1203	21	0.96	9	0.47	-	>1	1	0
204318	20245703	Group II chaperonin (CCT, TRIC), ATPase domain	1671	21	0.99	16	0.44	KOG0364	>1	>1	1
204460	20245715	YdeN-like	582	21	0.95	1	0.43	-	1	1	0
204590	20245731	-	486	20	0.91	6	0.41	KOG4502	1	1	0
205140	20245805	Ribosomal protein L14	459	21	0.99	2	0.42	KOG3441	1	1	0
205412	20245829	Pym (Within the bgcn gene intron protein, WIBG), N-terminal domain	693	21	0.95	3	0.39	KOG4325	1	1	0
205768	20245865	Tyrosine-dependent oxidoreductases	963	21	1	8	0.42	KOG1431	1	1	0
205824	20245870	Eukaryotic type KH-domain (KH-domain type I)	741	21	0.99	7	0.4	KOG3273	1	1	1
205831	20245872	AtpF-like	372	21	0.99	5	0.44	KOG3432	1	1	1
206094	20245892	-	921	19	0.76	3	0.42	-	1	1	0
206277	20245904	-	1530	21	0.97	14	0.46	KOG2613	1	1	0
206284	20245905	Hypothetical protein AF0491, N-terminal domain	765	21	0.97	6	0.4	KOG2917	1	1	1
206392	20245911	Ribosome anti-association factor eIF6 (aIF6)	738	21	1	6	0.45	KOG3185	1	1	1
206542	20245922	Exportin HEAT-like repeat	2439	21	0.98	1	0.47	KOG1107	1	1	0
206945	20245950	ISY1 N-terminal domain-like	897	20	0.74	1	0.43	KOG3068	1	1	1
207015	20245955	Exocyst complex component	2079	21	0.95	1	0.45	KOG2215	1	1	0
207043	20245958	C-terminal fragment of elongation factor SelB	1542	21	1	11	0.46	-	>1	1	1

L. gigantea gene ID	NCBI Gene ID	*L. gigantea* family description	Length of alignment (bp)	No. taxa	Proportion data complete	No. exons (*L. gigantea*)	G/C composition	Kog ID	All-by-all *L. gigantea* blast	OMA	Kocot et al.
207139	20245972	Regulatory subunit H of the V-type ATPase	1470	21	0.99	14	0.47	KOG2759	1	1	0
207505	20245998	WD40-repeat	963	21	1	10	0.46	KOG1332	>1	1	1
207546	20246000	SSRP1-like	2469	21	1	17	0.46	-	1	1	1
208313	20246055	PIN domain	753	21	0.86	3	0.42	KOG3164	1	1	0
208666	20246079	-	228	21	0.99	4	0.37	-	1	1	0
209108	20246119	Eukaryotic type KH-domain (KH-domain type I)	885	21	0.95	2	0.46	KOG3013	1	1	1
209321	20246133	V-type ATPase subunit E	693	21	1	1	0.46	KOG1664	1	1	1
209758	20246167	-	630	21	0.99	8	0.42	KOG3215	1	1	0
210048	20246187	-	333	21	0.98	4	0.41	KOG1705	1	1	1
210079	20246190	Spore coat polysaccharide biosynthesis protein SpsA	726	20	0.95	1	0.42	KOG2978	>1	>1	1
210320	20246201	-	1047	21	1	7	0.5	KOG1630	>1	1	0
210400	20246211	-	315	20	0.95	5	0.44	-	1	1	0
210506	20246221	Glycosyl transferases group 1	1506	21	0.89	13	0.42	KOG2941	1	1	1
210645	20246232	AraD-like aldolase/epimerase	747	20	0.84	1	0.45	KOG2631	1	1	0
210692	20246238	Calponin-homology domain, CH-domain	1119	21	0.99	13	0.45	-	1	1	0
211338	20246286	CRAL/TRIO domain	1140	20	0.79	4	0.51	KOG1470	1	1	0
211535	20246299	NOB1 zinc finger-like	1305	21	0.96	8	0.45	KOG2463	1	1	0
211712	20246310	Armadillo repeat	1521	20	0.79	8	0.44	KOG4413	1	1	0
211912	20246328	Hypothetical protein SAV1430	915	21	0.98	8	0.42	KOG2358	1	1	1
211937	20246329	-	999	21	0.92	11	0.4	KOG2962	1	1	0
212047	20246335	PCI domain (PINT motif)	1353	21	0.99	11	0.43	KOG1498	1	1	1
212070	20246337	-	618	20	0.94	1	0.38	-	1	1	0
212296	20246348	-	693	21	0.96	10	0.46	-	>1	>1	0
212332	20246352	Extended AAA-ATPase domain	1266	21	1	9	0.46	KOG0727	>1	>1	1
212487	20246368	Proteasome subunits	765	21	1	9	0.44	KOG0184	1	>1	1
212711	20246378	N-acetyl transferase, NAT	546	21	1	5	0.41	-	1	1	0
213277	20246419	FAD/NAD-linked reductases, N-terminal and central domains	2052	21	0.88	15	0.43	KOG1346	>1	>1	0
213398	20246424	C-terminal domain of ribosomal protein L2	918	20	0.93	4	0.46	KOG0438	1	>1	0
213693	20246438	Tetrapyrrole methylase	864	20	0.9	1	0.43	KOG3123	1	1	1
213741	20246444	-	450	21	1	4	0.45	KOG3356	1	1	0
213954	20246457	Group II chaperonin (CCT, TRIC), ATPase domain	1638	21	0.99	12	0.47	KOG0361	>1	>1	1
214293	20246485	-	1878	21	0.99	2	0.43	KOG2498	1	>1	0
214378	20246494	-	714	19	0.82	6	0.42	KOG3229	>1	1	1
214460	20246504	Transferrin	2532	21	1	15	0.43	-	>1	>1	0
214609	20246516	PCI domain (PINT motif)	1146	21	0.99	1	0.44	KOG2908	1	1	1
214610	20246517	-	370	21	0.97	3	0.44	-	1	1	0
215258	20246562	-	630	21	0.88	3	0.43	KOG3337	1	1	0
215790	20246595	Cyclin	1767	20	0.86	6	0.4	-	>1	>1	0
216100	20246615	-	585	21	0.86	2	0.47	KOG4054	1	1	0
216644	20246649	Nitrogenase iron protein-like	903	20	0.9	3	0.41	KOG1533	>1	>1	1
216779	20246659	-	630	21	0.97	9	0.41	KOG3272	1	1	1
216798	20246662	WD40-repeat	1125	21	0.96	10	0.46	KOG0647	>1	1	1

L. gigantea gene ID	NCBI Gene ID	*L. gigantea* family description	Length of alignment (bp)	No. taxa	Proportion data complete	No. exons (*L. gigantea*)	G/C composition	Kog ID	All-by-all *L. gigantea* blast	OMA	Kocot et al.
217054	20246680	Class I aminoacyl-tRNA synthetases (RS), catalytic domain	1632	21	0.97	12	0.45	KOG2144	>1	1	0
217090	20246682	Class II aminoacyl-tRNA synthetase (aaRS)-like, catalytic domain	1524	21	0.85	15	0.43	-	>1	>1	0
217484	20246714	WD40-repeat	930	21	0.99	11	0.46	-	>1	>1	0
217535	20246720	Phosphoglucose isomerase, PGI	1677	21	1	16	0.44	KOG2446	1	1	1
217988	20246746	SNARE fusion complex	351	20	0.89	1	0.39	-	>1	>1	0
218343	20246779	NadC C-terminal domain-like	1449	21	0.88	9	0.44	-	1	1	0
218353	20246780	Eukaryotic proteases	3072	21	0.7	2	0.46	-	1	>1	0
218660	20246797	SPRY domain	1602	21	0.87	1	0.46	KOG2626	1	1	0
219043	20246835	FAD/NAD-linked reductases, N-terminal and central domains	1530	21	1	14	0.47	KOG1335	1	>1	1
219117	20246840	Biotin synthase	1179	21	0.96	1	0.45	-	>1	1	0
219222	20246850	PUA domain	549	21	1	6	0.43	KOG2523	1	1	1
219898	20246902	NSFL1 (p97 ATPase) cofactor p47, SEP domain	1179	21	0.71	1	0.46	-	1	1	0
219929	20246904	-	1002	21	0.96	1	0.42	KOG3113	1	1	0
219936	20246905	beta-CASP RNA-metabolising hydrolases	2277	21	0.68	9	0.46	-	>1	1	0
220505	20246933	Ribosomal protein S16	483	21	0.99	2	0.41	KOG3419	1	1	0
220774	20246957	Arp2/3 complex subunits	903	21	0.96	10	0.4	KOG2826	1	1	1
220929	20246968	Transhydrogenase domain III (dIII)	3237	21	1	16	0.5	-	1	1	0
220939	20246970	-	1905	21	1	16	0.45	KOG2447	1	1	0
221202	20246979	RNA methyltransferase FtsJ	1065	21	0.99	1	0.44	KOG1099	>1	>1	1
221243	20246983	-	642	21	0.94	1	0.43	-	1	1	0
221335	20246988	Succinate dehydrogenase/fumarate reductase flavoprotein N-terminal domain	2013	21	1	15	0.49	KOG2403	1	1	1
222302	20247046	Pseudouridine synthase II TruB	1164	21	0.94	8	0.44	KOG2559	1	1	0
222316	20247048	MPP-like	1608	21	0.95	12	0.46	KOG2067	>1	>1	1
222603	20247062	-	1167	20	0.91	5	0.43	KOG2894	1	1	0
223434	20247124	Purine and uridine phosphorylases	903	21	0.88	6	0.46	-	>1	>1	0
223691	20247147	Cell cycle arrest protein BUB3	984	21	0.97	6	0.43	KOG1036	>1	1	0
223843	20247160	Hsp90 co-chaperone CDC37	1161	21	0.99	1	0.44	KOG2260	1	1	1
224000	20247176	L domain	915	21	0.93	6	0.45	-	>1	>1	0
224093	20247185	Bacterial dinuclear zinc exopeptidases	1419	21	1	14	0.45	KOG2276	1	1	0
224100	20247186	ETFP subunits	1008	21	1	1	0.45	KOG3954	1	1	1
224176	20247191	-	981	21	0.99	11	0.42	KOG3117	1	1	0
224262	20247203	JAB1/MPN domain	1047	21	0.99	8	0.45	KOG1554	>1	1	1
224404	20247212	G proteins	1212	21	0.96	8	0.45	KOG3887	1	1	0
224434	20247213	-	615	19	0.88	1	0.42	-	1	1	0
224543	20247221	PTPA-like	1131	21	0.92	8	0.45	KOG2867	1	1	0
225027	20247250	FHA domain	1749	20	0.83	9	0.47	KOG2293	1	1	0
225029	20247251	-	1683	20	0.79	8	0.42	-	1	1	0
225039	20247254	DNA-binding protein AlbA	639	21	0.91	1	0.4	-	1	1	0
225095	20247258	Cytidylytransferase	2301	21	0.96	16	0.44	KOG1461	>1	1	0
225373	20247279	-	663	21	0.98	1	0.42	KOG3096	1	1	0
225615	20247296	AD-003 protein-like	750	19	0.89	2	0.45	-	1	1	1
225644	20247298	-	1038	21	0.85	10	0.47	-	1	1	0

L. gigantea gene ID	NCBI Gene ID	*L. gigantea* family description	Length of alignment (bp)	No. taxa	Proportion data complete	No. exons (*L. gigantea*)	G/C composition	Kog ID	All-by-all *L. gigantea* blast	OMA	Kocot et al.
225722	20247303	Glyceraldehyde-3-phosphate dehydrogenase-like, N-terminal domain	1710	20	0.89	3	0.43	KOG0693	1	1	0
225963	20247320	SPT5 KOW domain-like	3357	21	0.96	2	0.48	-	1	1	0
226261	20247339	Armadillo repeat	1203	20	0.86	3	0.41	KOG2973	1	1	0
226791	20247364	Proteasome subunits	618	21	1	6	0.45	KOG0180	>1	1	1
227129	20247384	CCCH zinc finger	1323	21	0.99	11	0.42	KOG1763	1	1	1
227166	20247387	YeaZ-like	1011	21	0.96	1	0.44	KOG2708	>1	>1	0
227355	20247406	28-residue LRR	1074	21	1	9	0.47	KOG3735	1	1	0
227386	20247410	Threonyl-tRNA synthetase (ThrRS), second 'additional' domain	1038	21	0.99	9	0.45	-	>1	1	0
227502	20247416	Bacterial dinuclear zinc exopeptidases	1467	17	0.7	1	0.44	-	1	1	0
227818	20247439	-	1368	21	0.97	12	0.43	KOG4508	1	1	0
227857	20247455	Ubiquitin carboxyl-terminal hydrolase, UCH	1557	21	0.92	17	0.44	-	1	>1	1
228040	20247512	VPS28 N-terminal domain	660	21	0.98	7	0.41	KOG3284	1	1	1
228043	20247514	-	1875	21	0.97	10	0.39	-	1	1	0
228100	20247539	-	788	21	0.98	9	0.42	KOG4813	1	1	0
228134	20247553	Ubiquitin-related	1074	19	0.8	10	0.45	-	1	1	0
228336	20247626	-	579	21	0.99	3	0.46	-	1	1	0
228377	20247640	-	468	21	0.99	3	0.45	-	1	1	0
228391	20247646	RING finger domain, C3HC4	2706	21	0.94	16	0.44	-	1	1	0
228440	20247666	Extended AAA-ATPase domain	1086	21	0.97	9	0.41	KOG0991	>1	>1	0
228476	20247675	Classic zinc finger, C2H2	615	21	0.93	6	0.45	KOG4727	1	1	1
228627	20247720	-	1332	21	0.99	8	0.46	-	1	1	0
228741	20247767	G proteins	795	21	0.98	7	0.47	KOG0090	1	1	0
228809	20247794	-	1602	21	0.99	5	0.36	KOG4049	1	1	0
228997	20247863	Leucine zipper domain	1359	21	0.98	2	0.43	KOG4571	1	0	0
229010	20247869	Tandem AAA-ATPase domain	1365	21	0.96	13	0.42	-	>1	>1	1
229034	20247874	Ribosomal protein S6	450	21	0.98	3	0.42	KOG4708	1	1	0
229272	20247954	Ribosomal protein L1	1116	21	0.97	9	0.44	KOG1569	1	1	0
229368	20247981	Spermadhesin, CUB domain	1050	21	1	6	0.43	-	1	0	0
229474	20248014	-	309	21	0.97	6	0.42	KOG3476	1	1	0
229754	20248095	-	966	21	0.94	1	0.42	KOG3031	1	1	1
229846	20248126	-	873	21	1	8	0.41	-	1	1	0
230243	20248244	-	336	21	0.99	1	0.43	KOG4104	1	1	0
230249	20248246	DPP6 N-terminal domain-like	2112	21	0.98	18	0.47	KOG2314	>1	1	0
230289	20248256	-	471	21	0.98	4	0.47	KOG4092	1	1	0
230810	20248394	Proteasome subunits	780	21	0.97	1	0.45	KOG0185	1	1	1
230880	20248413	Translational machinery components	1338	21	0.97	9	0.47	KOG2646	1	1	0
230887	20248418	-	988	21	0.98	1	0.41	-	>1	0	0
231007	20248436	-	351	21	0.99	5	0.47	KOG4455	1	1	0
231140	20248490	Proteasome subunits	831	20	0.95	8	0.43	KOG0173	>1	>1	1
231346	20248556	-	465	21	0.99	1	0.46	KOG4559	1	1	0
231565	20248636	Group II chaperonin (CCT, TRIC), ATPase domain	1609	21	0.92	14	0.42	KOG0359	1	>1	1
231752	20248701	Histone lysine methyltransferases	1209	21	0.98	11	0.45	-	1	1	0

L. gigantea gene ID	NCBI Gene ID	*L. gigantea* family description	Length of alignment (bp)	No. taxa	Proportion data complete	No. exons (*L. gigantea*)	G/C composition	Kog ID	All-by-all *L. gigantea* blast	OMA	Kocot et al.
231795	20248716	G proteins	1947	21	1	17	0.43	KOG1490	1	1	1
232029	20248787	Proteasome subunits	726	21	0.99	1	0.41	KOG0176	>1	>1	1
232051	20248797	-	987	21	0.96	1	0.42	-	1	0	0
232137	20248818	Prefoldin	1593	21	0.95	9	0.43	-	1	1	0
232335	20248884	-	927	21	0.98	12	0.44	KOG1639	1	1	1
232340	20248885	-	624	21	0.98	4	0.47	KOG4633	1	1	0
232467	20248932	-	504	21	0.94	2	0.42	-	1	0	0
232538	20248951	N-acetyl transferase, NAT	1239	21	0.97	11	0.43	KOG2696	1	1	0
232701	20249015	WD40-repeat	1239	21	0.99	11	0.43	-	1	1	0
233189	20249175	Ribosomal proteins L15p and L18e	903	21	0.94	4	0.47	KOG0846	1	1	0
233882	20249389	WD40-repeat	1014	21	0.99	3	0.43	KOG0645	1	1	0
233963	20249406	-	759	21	1	7	0.49	KOG1647	1	1	1
234125	20249454	-	771	21	0.96	3	0.42	-	1	0	0
234160	20249460	Vacuolar sorting protein domain	756	21	0.98	1	0.45	KOG3341	1	1	1
234486	20249553	DPP6 N-terminal domain-like	1794	20	0.82	17	0.41	KOG2315	>1	1	0
234493	20249555	mRNA decapping enzyme DcpS C-terminal domain	1053	21	1	1	0.5	KOG3969	1	1	0
234495	20249556	WD40-repeat	1401	20	0.9	13	0.44	KOG0285	>1	>1	1
234571	20249585	WD40-repeat	1530	20	0.86	9	0.41	KOG0294	1	1	0
234912	20249681	-	636	20	0.93	2	0.38	-	1	1	0
234962	20249702	-	1110	21	0.95	9	0.44	KOG3190	1	1	0
235396	20249837	-	1029	21	0.95	7	0.43	KOG4681	1	1	0
235411	20249843	Smg-4/UPF3	2010	21	0.89	7	0.44	KOG1295	1	>1	0
235540	20249879	-	2091	21	1	9	0.49	KOG3756	1	>1	0
235667	20249922	Group II chaperonin (CCT, TRIC), ATPase domain	1674	21	0.99	12	0.45	-	>1	>1	0
235720	20249944	Staphylococcal nuclease	2727	20	0.94	22	0.4	-	1	1	0
235732	20249948	Ribosomal protein S15	1143	21	1	7	0.43	KOG2815	1	1	0
235879	20249993	-	936	21	0.98	7	0.42	KOG3188	1	1	1
235937	20250006	LexA-related	549	21	0.97	6	0.47	KOG3342	1	1	1
235960	20250013	-	1395	21	0.99	10	0.44	-	1	1	0
236022	20250029	N-acetyl transferase, NAT	573	21	0.99	2	0.42	KOG3235	>1	>1	1
236049	20250041	-	1158	21	0.98	9	0.43	KOG3933	1	1	0
236225	20250086	-	1899	21	0.88	22	0.41	KOG2701	1	1	0
236282	20250100	-	582	21	0.95	5	0.44	KOG2424	1	1	1
236402	20250137	-	945	21	0.98	9	0.44	KOG3050	>1	1	0
236479	20250165	Tandem AAA-ATPase domain	1572	21	0.96	10	0.48	KOG0332	1	>1	0
236612	20250199	-	972	20	0.91	2	0.45	-	1	1	0
236999	20250315	VHL	486	21	0.95	2	0.44	KOG4710	1	1	0
237076	20250338	Ribonuclease PH domain 1-like	750	21	0.98	3	0.46	-	>1	>1	1
237410	20250452	Aldo-keto reductases (NADP)	1014	21	0.96	7	0.4	-	>1	>1	0
237436	20250459	Tetratricopeptide repeat (TPR)	1173	20	0.85	11	0.45	-	1	1	0
237593	20250498	-	2094	21	0.96	3	0.43	-	>1	>1	0
237715	20250541	Single strand DNA-binding domain, SSB	813	20	0.93	2	0.44	-	1	1	0
237836	20250572	Canonical RBD	1653	21	0.94	10	0.43	KOG0533	1	1	0

L. gigantea gene ID	NCBI Gene ID	*L. gigantea* family description	Length of alignment (bp)	No. taxa	Proportion data complete	No. exons (*L. gigantea*)	G/C composition	Kog ID	All-by-all *L. gigantea* blast	OMA	Kocot et al.
238255	20250701	WD40-repeat	1338	21	1	12	0.46	KOG0313	1	1	0
238326	20250725	Triosephosphate isomerase (TIM)	756	14	0.59	6	0.44	KOG1643	1	1	1
238486	20250773	-	1026	19	0.88	1	0.43	-	1	1	0
238978	-	Inositol monophosphatase/fructose-1,6-bisphosphatase-like	1101	21	1	7	0.42	KOG1458	1	1	1
239065	20250944	RNase III catalytic domain-like	1134	21	0.94	10	0.43	KOG3769	1	1	0
239070	20250946	WD40-repeat	1371	21	1	12	0.49	KOG2096	1	1	0
239110	20250960	Phosphonoacetaldehyde hydrolase-like	855	21	1	1	0.42	-	1	1	0
239114	20250963	Ribosomal protein L10-like	831	21	0.85	5	0.42	KOG4241	1	1	0
239373	20251041	-	2619	21	0.92	2	0.44	KOG2673	1	1	0
239443	20251062	WW domain	948	23	0.93	9		KOG0150	1	1	0

Appendix 2.10. 208 multiple copy genes, with evidence of paralogous sequences, determined from 21 eupulmonate transcriptomes. The ‘all-by-all *L. gigantea* blast’ and ‘OMA’ columns notate whether the cluster containing the respective *L. gigantea* contains only a single *L. gigantea* sequence (1) or multiple (>1). In the case of the OMA results, zeros mean the *L. gigantea* sequence was not present in the OMA database at the time of download (3rd of September, 2014). The Kocot et al. column notates genes which are also present (1) in a molluscan phylogenomic dataset (Kocot et al. 2011).

***L. gigantea* gene ID**	**NCBI Gene ID**	**No. of species with evidence of paralogy**	***L. gigantea* family description**	**KOG ID**	**Length (bp)**	**All-by-all *L. gigantea* blast**	**OMA**	**Kocot et al.**
54396	20251282	multiple	Ubiquitin-related	KOG0011	1053	1	1	1
66846	20251791	multiple	-	KOG1646	672	1	1	1
70408	20251967	single	Ribosomal proteins L15p and L18e	-	462	>1	>1	1
74451	20252131	multiple	-	-	990	>1	>1	0
80461	20252462	multiple	Ubiquitin-related	-	204	>1	0	0
102125	20229571	multiple	-	-	1041	>1	1	1
103152	20229711	single	ETX/MTX2	-	633	>1	>1	0
103941	20229803	multiple	NAP-like	-	1008	>1	1	1
105568	20230003	multiple	MPP-like	KOG0960	1395	>1	>1	1
108713	20230386	multiple	-	KOG3318	510	1	1	0
108853	20230401	multiple	V-type ATP synthase subunit C	KOG2957	1047	1	1	1
109561	20230483	single	Ribosomal protein S7	-	606	1	1	1
110000	20230537	multiple	Tandem AAA-ATPase domain	-	1500	>1	>1	1
110080	20230547	multiple	Ribosomal protein S24e	KOG3424	447	1	1	1
114287	20231074	multiple	Extended AAA-ATPase domain	-	969	>1	>1	1
119755	20231770	multiple	RecA protein-like (ATPase-domain)	KOG1351	1536	>1	>1	1
128584	20232856	multiple	Ribosomal protein L14	KOG0901	411	1	1	1
132223	20233286	multiple	Enolase-phosphatase E1	KOG2630	924	1	1	1
132224	20233287	multiple	MIF4G domain-like	KOG2767	1239	1	1	1
134010	20233507	multiple	Nucleolar RNA-binding protein Nop10-like	KOG3503	195	1	1	1
138721	20234102	multiple	Extended AAA-ATPase domain	KOG0728	1194	>1	>1	1
149167	20235476	multiple	-	KOG1725	1850	>1	1	1
149778	20235492	multiple	Ribosomal protein S19	-	503	>1	1	1
150191	20235509	multiple	L30e/L7ae ribosomal proteins	KOG3406	614	1	1	1
150772	20235523	multiple	Cold shock DNA-binding domain-like	KOG3502	301	1	1	1
153858	20236129	multiple	Proteasome subunits	KOG0174	624	>1	>1	1
156627	20237051	multiple	Nucleosome core histones	KOG1757	309	1	1	1
158030	20237602	multiple	Epsilon subunit of F1F0-ATP synthase N-terminal domain	KOG1758	393	1	1	1
161608	20238734	multiple	Hsp90 middle domain	-	2175	>1	>1	1
164153	20239659	multiple	Phosphoserine phosphatase	KOG1615	684	1	1	1
166689	20240419	multiple	RNA polymerase subunit RPB10	KOG3497	901	1	1	1
167500	20240631	multiple	Ribosomal protein L5	KOG0397	504	1	1	1
170380	20241389	multiple	Citrate synthase	KOG2617	2517	1	1	1
171636	20241850	multiple	IPP isomerase-like	KOG0142	1756	1	1	1

L. gigantea gene ID	NCBI Gene ID	No. of species with evidence of paralogy	*L. gigantea* family description	KOG ID	Length (bp)	All-by-all *L. gigantea* blast	OMA	Kocot et al.
173095	20242286	multiple	Pepsin-like	-	2524	>1	1	1
177209	20244194	multiple	Arginine methyltransferase	-	987	>1	>1	1
177527	20244218	single	CAT-like	-	573	>1	>1	1
178160	20244264	multiple	G proteins	-	1242	>1	>1	1
178960	20244308	multiple	-	KOG3434	387	1	1	1
181421	20244441	multiple	-	KOG2240	1410	1	1	0
182626	20244504	multiple	GABARAP-like	-	1453	>1	>1	1
183079	20244523	single	N-terminal domain of the delta subunit of the F1F0-ATP synthase	KOG1662	991	1	1	1
184120	20244582	multiple	WD40-repeat	-	1106	>1	>1	1
184255	20244589	multiple	PDI-like	-	2301	>1	>1	1
184532	20244606	multiple	S-adenosylhomocystein hydrolase	-	1649	>1	>1	1
184997	20244637	single	Glutamine synthetase catalytic domain	-	2006	>1	>1	0
185272	20244647	multiple	DNA polymerase processivity factor	KOG1636	1620	1	1	1
186221	20244692	multiple	Ribosomal protein S8	KOG1754	512	1	1	1
186985	20244726	multiple	C-terminal domain of ribosomal protein L2	KOG2309	817	1	>1	1
190501	20244897	multiple	Protein kinases, catalytic subunit	-	2597	>1	>1	1
190601	20244905	multiple	Class II aminoacyl-tRNA synthetase (aaRS)-like, catalytic domain	-	2681	>1	>1	1
190901	20244914	multiple	-	KOG3320	813	1	1	1
192827	20245000	multiple	-	KOG3452	455	1	1	1
193923	20245051	single	Actin/HSP70	KOG0678	2151	1	>1	1
194715	20245090	multiple	Preprotein translocase SecY subunit	KOG1373	3098	1	1	1
195818	20245152	multiple	Ribosomal protein L19 (L19e)	KOG1696	764	1	1	1
196203	20245177	multiple	Ribosomal protein L3	KOG0746	1291	1	1	1
196510	20245196	multiple	Protein serine/threonine phosphatase	KOG0372	3005	>1	>1	0
196756	20245209	single	-	KOG2291	2338	1	1	1
196809	20245212	multiple	Capz alpha-1 subunit	KOG0836	1616	1	1	1
197575	20245256	multiple	Acireductone dioxygenase	KOG2107	808	1	1	1
197780	20245268	multiple	-	KOG2239	1268	1	1	1
197848	20245273	single	-	KOG2754	1698	1	1	1
199050	20245331	multiple	G proteins	-	3143	>1	>1	1
199626	20245359	multiple	Pyruvate kinase	-	3589	>1	>1	1
200562	20245411	multiple	Succinyl-CoA synthetase, beta-chain, N-terminal domain	KOG2799	1424	>1	>1	1
200623	20245414	multiple	-	KOG3998	2276	1	1	1
200884	20245432	single	Class II aminoacyl-tRNA synthetase (aaRS)-like, catalytic domain	KOG0555	3202	>1	>1	1
201223	20245455	multiple	-	KOG1790	421	1	1	1
201878	20245491	multiple	RecA protein-like (ATPase-domain)	KOG1350	1804	>1	>1	1
202003	20245503	multiple	L30e/L7ae ribosomal proteins	KOG3166	930	1	1	1
202410	20245529	multiple	Cold shock DNA-binding domain-like	KOG1749	467	>1	1	1
202499	20245535	multiple	-	KOG1628	1196	1	1	1
202957	20245564	multiple	Cold shock DNA-binding domain-like	KOG1728	504	1	1	1
203071	20245575	single	Extended AAA-ATPase domain	KOG2680	1988	>1	>1	1
203722	20245634	multiple	-	KOG1656	2774	>1	1	1

L. gigantea gene ID	NCBI Gene ID	No. of species with evidence of paralogy	*L. gigantea* family description	KOG ID	Length (bp)	All-by-all *L. gigantea* blast	OMA	Kocot et al.
203788	20245642	multiple	ETFP subunits	KOG3180	1046	1	1	1
203874	20245654	multiple	Supernatant protein factor (SPF), C-terminal domain	KOG1692	2730	>1	>1	1
203916	20245662	multiple	Nitrogenase iron protein-like	KOG2825	1022	1	1	1
204040	20245676	multiple	Rps17e-like	KOG0187	616	1	1	1
204936	20245768	single	-	KOG3404	958	1	1	1
205544	20245843	single	Second domain of Mu2 adaptin subunit (ap50) of ap2 adaptor	-	1785	>1	>1	1
205560	20245845	multiple	G proteins	-	826	>1	>1	1
205662	20245855	multiple	Nucleoside diphosphate kinase, NDK	-	715	1	>1	1
205749	20245862	multiple	Protein kinases, catalytic subunit	-	1120	>1	>1	1
205756	20245863	multiple	WD40-repeat	-	2061	>1	1	1
206537	20245921	single	Synatpobrevin N-terminal domain	KOG0861	815	1	1	1
206617	20245929	multiple	RecA protein-like (ATPase-domain)	KOG1353	1918	>1	>1	1
207066	20245962	multiple	Extended AAA-ATPase domain	KOG0652	1371	1	>1	1
207101	20245967	multiple	Tandem AAA-ATPase domain	KOG0327	2120	>1	>1	1
207121	20245970	multiple	Nucleotide and nucleoside kinases	-	811	>1	>1	1
207423	20245988	multiple	RplX-like	KOG0829	712	1	1	1
207552	20246002	multiple	Ribosomal protein L14e	KOG1694	758	1	1	1
207717	20246017	multiple	Ribosomal protein L22	KOG3353	902	1	1	1
207726	20246021	single	Sm motif of small nuclear ribonucleoproteins, SNRNP	KOG3460	1204	1	1	1
209986	20246182	multiple	Ribosomal protein L18 and S11	KOG0407	587	1	1	1
210271	20246197	multiple	-	KOG0378	923	1	1	1
210661	20246233	multiple	Fe,Mn superoxide dismutase (SOD), C-terminal domain	KOG0876	1475	1	1	1
211297	20246281	multiple	L30e/L7ae ribosomal proteins	KOG3167	1743	1	1	1
212293	20246347	multiple	-	KOG1655	1652	>1	1	1
212802	20246386	multiple	-	KOG3486	322	1	1	1
214125	20246473	multiple	Ferredoxin reductase FAD-binding domain-like	-	3024	>1	1	1
215959	20246606	multiple	Ubiquitin-related	KOG0009	836	1	1	1
216416	20246633	multiple	Actin/HSP70	-	2926	>1	>1	1
216998	20246676	multiple	Purine and uridine phosphorylases	KOG3985	1349	>1	>1	1
217766	20246736	multiple	Ribosomal protein L6	KOG3255	663	1	1	1
218864	20246816	multiple	Prefoldin	KOG4098	1253	1	1	1
219170	20246845	multiple	HEAT repeat	-	2905	1	1	1
219397	20246863	multiple	Ribosomal protein S10	KOG0900	921	1	1	1
219464	20246869	multiple	-	-	835	>1	>1	0
219559	20246878	single	Ribosomal protein L4	KOG1475	1383	1	1	1
219589	20246879	multiple	monodomain cytochrome c	KOG3453	730	1	1	1
220690	20246949	multiple	-	KOG1772	946	1	1	1
222194	20247043	multiple	Mitochondrial carrier	KOG0767	1952	1	1	1
222708	20247071	single	Nop domain	-	2081	>1	1	1
223715	20247149	single	F1F0 ATP synthase subunit C	KOG0233	2059	>1	>1	1
223907	20247169	multiple	UBC-related	KOG0418	1410	1	1	1
223917	20247170	multiple	Prokaryotic type KH domain (KH-domain type II)	KOG3181	1792	1	1	1
224562	20247222	multiple	Ribosomal protein L18 and S11	KOG0875	1055	1	1	1

L. gigantea gene ID	NCBI Gene ID	No. of species with evidence of paralogy	*L. gigantea* family description	KOG ID	Length (bp)	All-by-all *L. gigantea* blast	OMA	Kocot et al.
225011	20247249	multiple	Ribosomal protein L37ae	KOG0402	347	1	1	1
225558	20247291	multiple	Lactate & malate dehydrogenases, C-terminal domain	KOG1494	2507	1	>1	1
225601	20247295	multiple	Glutathione peroxidase-like	KOG0854	1076	1	>1	1
225993	20247323	multiple	Casein kinase II beta subunit	KOG3092	1402	1	1	1
226017	20247325	single	WD40-repeat	KOG0310	1606	1	1	0
226411	20247345	multiple	Ribosomal proteins L24p and L21e	KOG1732	561	1	1	1
226825	20247367	multiple	Branched-chain alpha-keto acid dehydrogenase Pyr module	-	2978	>1	>1	1
227198	20247392	multiple	G proteins	-	1011	>1	1	1
227257	20247397	multiple	-	KOG3296	1611	1	1	0
229207	20247924	multiple	PX domain	-	1599	>1	1	1
229535	20248035	multiple	Ribosomal protein L10e	KOG0857	1282	1	1	1
229894	20248138	single	Ubiquitin-related	KOG3493	578	1	1	1
230007	20248174	multiple	G proteins	-	803	1	>1	1
230263	20248250	multiple	Ribosomal protein L1	-	718	>1	1	1
231806	20248720	multiple	PP2C-like	KOG1379	1666	1	1	0
232303	20248877	multiple	Ribosomal protein L37e	KOG3475	331	1	1	1
232500	20248943	multiple	Peptide methionine sulfoxide reductase	KOG1635	770	1	1	1
233303	20249217	single	-	KOG3407	1007	1	1	0
233453	20249272	multiple	-	KOG1816	1138	1	1	1
233682	20249330	multiple	-	-	1547	>1	1	1
233776	20249346	multiple	G proteins	-	1190	>1	>1	0
233830	20249367	multiple	Translation initiation factor 2 beta, aIF2beta, N-terminal domain	KOG2768	2210	1	1	1
234041	20249430	multiple	Capz beta-1 subunit	KOG3174	3194	1	1	1
234281	20249481	multiple	Calmodulin-like	-	1320	>1	>1	0
234443	20249538	multiple	Mitochondrial carrier	KOG0758	1380	>1	1	1
235900	20249996	single	ATP synthase (F1-ATPase), gamma subunit	KOG1531	1443	1	1	1
235941	20250007	multiple	Rps19E-like	KOG3411	2625	1	1	1
236064	20250047	multiple	Heme-dependent catalases	KOG0047	2292	1	1	1
236339	20250118	multiple	UBC-related	-	957	>1	1	1
236462	20250157	multiple	WD40-repeat	-	4522	>1	>1	1
236815	20250264	single	L30e/L7ae ribosomal proteins	-	432	1	1	1
237408	20250451	multiple	GroEL chaperone, ATPase domain	KOG0356	3851	1	1	1
237412	#N/A	multiple	PCI domain (PINT motif)	KOG1463	1354	>1	1	1
237446	20250464	multiple	Band 7/SPFH domain	KOG3083	1736	>1	>1	1
237709	20250538	multiple	Ribosomal protein L10-like	KOG0815	1158	1	1	1
239089	20250957	multiple	Thioltransferase	KOG2603	1975	1	1	1
239290	20251019	multiple	-	KOG3283	724	1	1	1
68260	20251857	multiple	Bcl-2 inhibitors of programmed cell death	-	522	>1	>1	0
173993	20242579	multiple	-	-	1371	1	1	0
110384	20230583	multiple	Histidine acid phosphatase	KOG1382	849	1	1	0
112701	20230878	multiple	Pleckstrin-homology domain (PH domain)	-	597	1	1	0
139793	20234232	multiple	CAC2371-like	-	417	1	1	0

L. gigantea gene ID	NCBI Gene ID	No. of species with evidence of paralogy	*L. gigantea* family description	KOG ID	Length (bp)	All-by-all *L. gigantea* blast	OMA	Kocot et al.
184680	20244615	multiple	Thioesterase domain of polypeptide, polyketide and fatty acid synthases	-	2109	1	1	0
205447	20245834	multiple	Roadblock/LC7 domain	KOG4107	1811	1	1	0
230724	20248387	multiple	-	-	2027	1	1	0
231402	20248574	multiple	-	-	2490	1	1	0
231783	20248713	multiple	-	-	2619	1	0	0
237013	20250320	multiple	-	-	2803	>1	1	0
239113	20250962	multiple	Dimerization-anchoring domain of cAMP-dependent PK regulatory subunit	-	2006	1	0	0
133654	20233463	multiple	MutT-like	KOG1689	675	1	1	0
210667	20246234	multiple	Prefoldin	KOG3478	1359	1	1	0
214465	20246505	single	-	-	897	1	1	0
214570	20246513	multiple	-	-	944	1	1	0
219736	20246887	multiple	Tyrosine-dependent oxidoreductases	KOG3019	3078	1	1	0
236766	20250242	multiple	-	-	1771	1	1	0
237702	20250534	single	Elongation factor TFIIS domain 2	-	948	1	1	0
239042	20250939	multiple	HMG-box	-	2220	1	1	0
238461	20250766	multiple	Canonical RBD	KOG0122	1736	1	1	0
234917	20249685	multiple	-	-	1794	>1	>1	0
233363	20249243	multiple	-	-	1318	1	1	0
232897	20249081	single	-	-	1377	1	0	0
232875	20249073	multiple	GS domain	KOG3260	1438	1	1	0
230812	20248395	multiple	UBC-related	-	3418	>1	1	0
230231	20248239	multiple	Ribosomal protein S4	KOG3301	652	1	1	0
229997	20248171	multiple	-	-	3532	1	0	0
227005	20247374	multiple	Sm motif of small nuclear ribonucleoproteins, SNRNP	-	882	1	1	0
220498	20246932	multiple	YbiA-like	-	745	1	>1	0
181799	20244459	multiple	Ribosomal protein S15	KOG0400	519	>1	1	0
203638	20245619	multiple	Rhamnogalacturonase B, RhgB, middle domain	KOG3306	1522	>1	1	0
236386	20250131	multiple	Pleckstrin-homology domain (PH domain)	-	750	1	1	0
56896	20251375	multiple	-	-	738	1	>1	0
68346	20251861	multiple	-	-	498	1	>1	0
120581	20231880	multiple	-	-	312	>1	0	0
164038	20239605	multiple	L-arabinose binding protein-like	-	2211	>1	0	0
198249	20245294	multiple	Bcl-2 inhibitors of programmed cell death	-	4802	1	1	0
205086	20245791	multiple	-	-	1059	>1	>1	0
214161	20246475	multiple	ERP29 C domain-like	-	1457	>1	1	0
233061	20249137	multiple	Ubiquitin-related	-	2960	>1	1	0
234069	20249437	multiple	Phosphotyrosine-binding domain (PTB)	-	1521	1	0	0
72260	20252044	single	-	-	1761	1	>1	0
208067	20246041	multiple	Glyceraldehyde-3-phosphate dehydrogenase-like, N-terminal domain	-	1271	1	1	0
211795	20246317	multiple	-	KOG4737	1103	>1	1	0
231033	20248442	multiple	WD40-repeat	KOG1446	959	1	1	0
233248	20249200	multiple	-	KOG2441	2174	1	1	0

***L. gigantea* gene ID**	**NCBI Gene ID**	**No. of species with evidence of paralogy**	***L. gigantea* family description**	**KOG ID**	**Length (bp)**	**All-by-all *L. gigantea* blast**	**OMA**	**Kocot et al.**
234402	20249526	multiple	Formate/glycerate dehydrogenases, NAD-domain	-	1226	1	>1	0
115714	20231259	multiple	-	KOG1348	449	1	1	1
165837	20240193	multiple	DnaQ-like 3'-5' exonuclease	KOG3242	187	1	1	1
205433	20245831	multiple	Brix domain	KOG2971	306	1	1	1
206255	20245901	multiple	Vacuolar ATP synthase subunit C	KOG2909	384	1	1	1

Appendix 2.11. 375 genes identified from the Agalma analysis which contained sequences for ≥ 18 taxa and were the only orthologous cluster resulting from the respective homolog cluster.

Agalma orthologous clusters	*L. gigantea* gene ID	*L. gigantea* family description	No. of Taxa	Length (aa)
homologs_49_0	224478	Nitrogenase iron protein-like	19	359
homologs_49_10003	235566	PDI-like	18	254
homologs_49_10008	197143	Nip7p homolog, N-terminal domain	19	180
homologs_49_1002	229002	Glutathione peroxidase-like	19	145
homologs_49_10040	239150	Ribosomal protein S10	18	170
homologs_49_10073	-	-	20	419
homologs_49_1008	193477	WD40-repeat	18	302
homologs_49_10152	179707	-	19	217
homologs_49_10153	197001	Translationally controlled tumor protein TCTP (histamine-releasing factor)	19	177
homologs_49_10175	231485	-	19	161
homologs_49_10204	-	-	20	241
homologs_49_10208	186083	Ribosomal protein L13	19	198
homologs_49_10224	215045	PF0523-like	18	181
homologs_49_10352	-	-	18	132
homologs_49_10366	183391	-	18	212
homologs_49_10370	116530	WW domain	20	328
homologs_49_10386	162091	Chaperone J-domain	20	360
homologs_49_10392	125037	-	21	448
homologs_49_10412	110772	Mitochondrial import receptor subunit Tom20	19	165
homologs_49_10455	235650	Tudor domain	20	257
homologs_49_10522	229048	-	19	198
homologs_49_10533	99569	-	18	302
homologs_49_10545	233722	Signal recognition particle alu RNA binding heterodimer, SRP9/14	18	109
homologs_49_10575	201223	-	19	123
homologs_49_10580	231111	WD40-repeat	20	510
homologs_49_10644	231867	-	18	361
homologs_49_10676	97242	-	18	222
homologs_49_10711	234305	-	19	273
homologs_49_10736	210667	Prefoldin	19	126
homologs_49_10755	124295	MutT-like	19	252
homologs_49_10770	154623	-	20	526
homologs_49_10827	-	-	19	233
homologs_49_10856	231601	Tetraspanin	18	260
homologs_49_1088	96853	Chaperone J-domain	20	275
homologs_49_10890	236637	Mitochondrial ATP synthase coupling factor 6	20	142
homologs_49_10891	187630	-	18	218
homologs_49_10894	198435	RNase P subunit p29-like	19	235
homologs_49_10910	86689	Elafin-like	20	261
homologs_49_10943	192237	FKBP immunophilin/proline isomerase	18	143
homologs_49_10973	205433	Brix domain	19	298
homologs_49_10992	80486	-	20	323
homologs_49_11051	136530	-	20	352

Agalma orthologous clusters	***L. gigantea* gene ID**	***L. gigantea* family description**	**No. of Taxa**	**Length (aa)**
homologs_49_1107	205112	Aldo-keto reductases (NADP)	19	286
homologs_49_11074	232861	-	19	211
homologs_49_11088	179564	TBP-associated factors, TAFs	18	127
homologs_49_11099	228275	-	20	238
homologs_49_11136	126582	DnaQ-like 3'-5' exonuclease	18	209
homologs_49_11143	217500	Hypothetical protein AT3g04780/F7O18 27	20	287
homologs_49_11146	223821	DJ-1/PfpI	20	195
homologs_49_1117	159581	-	18	286
homologs_49_11174	181139	Proteasome subunits	19	201
homologs_49_11214	185986	RecA protein-like (ATPase-domain)	21	622
homologs_49_11352	214285	Txnl5-like	18	140
homologs_49_11372	177712	JAB1/MPN domain	18	265
homologs_49_11387	192880	Proteasome subunits	19	266
homologs_49_1143	117396	ADP-ribosylglycohydrolase	18	359
homologs_49_11437	203482	ATP synthase D chain-like	20	171
homologs_49_11439	149249	Peptidyl-tRNA hydrolase II	19	184
homologs_49_1146	75658	Canonical RBD	20	183
homologs_49_11498	162789	NlpC/P60	19	186
homologs_49_11508	194136	Eukaryotic translation initiation factor 3 subunit 12, eIF3k, N-terminal domain	20	216
homologs_49_11648	110056	-	19	343
homologs_49_11735	187188	FKBP immunophilin/proline isomerase	18	138
homologs_49_11749	157663	-	18	340
homologs_49_11786	203449	Toll/Interleukin receptor TIR domain	21	415
homologs_49_11813	128093	Cold shock DNA-binding domain-like	19	375
homologs_49_11818	141139	HIT zinc finger	20	291
homologs_49_11834	-	-	20	218
homologs_49_11909	203942	-	19	160
homologs_49_11911	191474	FolH catalytic domain-like	18	308
homologs_49_11950	186317	Thioesterases	19	288
homologs_49_12010	232677	Proteasome subunits	19	285
homologs_49_12016	220342	Bacterial dinuclear zinc exopeptidases	19	480
homologs_49_12049	86941	Ribosomal protein S10	19	193
homologs_49_12072	172070	WD40-repeat	20	383
homologs_49_1215	206537	Synatpobrevin N-terminal domain	18	199
homologs_49_12203	221428	Phosducin	20	309
homologs_49_12227	139895	Sulfatase-modifying factor-like	20	368
homologs_49_12236	102351	-	18	188
homologs_49_12361	232409	Ribosomal protein L32p	18	184
homologs_49_12411	218017	-	20	229
homologs_49_12448	106249	-	20	179
homologs_49_12449	238948	Cu,Zn superoxide dismutase-like	18	1035
homologs_49_12451	110935	-	20	296
homologs_49_12488	239238	SAP domain	19	273
homologs_49_12572	232038	Single strand DNA-binding domain, SSB	18	170

Agalma orthologous clusters	*L. gigantea* gene ID	*L. gigantea* family description	No. of Taxa	Length (aa)
homologs_49_12575	-	-	18	284
homologs_49_12586	204243	-	19	249
homologs_49_12766	159596	-	19	202
homologs_49_12793	238770	UBX domain	18	249
homologs_49_12842	224221	Ribosomal protein L28	19	323
homologs_49_1287	234533	-	20	194
homologs_49_12892	122247	Cytochrome bc1 domain	21	308
homologs_49_12902	233001	Ribosomal protein L18 and S11	20	206
homologs_49_12910	237427	-	19	507
homologs_49_12929	-	-	18	293
homologs_49_13027	204570	Canonical RBD	20	123
homologs_49_13039	156565	-	19	159
homologs_49_1309	234076	CHY zinc finger	21	600
homologs_49_13196	232655	-	20	269
homologs_49_13222	236786	Ribonuclease PH domain 1-like	19	292
homologs_49_13274	214033	Eukaryotic proteases	21	305
homologs_49_13280	209731	-	20	135
homologs_49_1329	203444	-	19	433
homologs_49_13290	211681	Translation initiation factor eIF4e	18	231
homologs_49_13335	111655	Retrovirus zinc finger-like domains	20	211
homologs_49_13348	128725	-	21	331
homologs_49_1338	192905	MAL13P1.257-like	19	160
homologs_49_13408	238723	C-type lectin domain	18	509
homologs_49_13441	229773	-	19	278
homologs_49_13560	233896	Glutathione peroxidase-like	19	191
homologs_49_13594	233865	-	18	228
homologs_49_13636	216116	VPS36 N-terminal domain-like	19	276
homologs_49_13667	-	-	18	158
homologs_49_13772	128222	Tyrosine-dependent oxidoreductases	18	258
homologs_49_13794	199614	Calmodulin-like	19	151
homologs_49_13830	119809	-	18	196
homologs_49_13883	184158	-	19	238
homologs_49_13886	157968	-	19	506
homologs_49_1392	231362	-	19	252
homologs_49_13969	235789	G proteins	20	209
homologs_49_13988	189800	eEF1-gamma domain	21	453
homologs_49_14092	202251	eIF1-like	20	113
homologs_49_14193	77324	-	19	572
homologs_49_14288	151060	Tetratricopeptide repeat (TPR)	21	502
homologs_49_1566	204895	LIM domain	19	198
homologs_49_1644	221428	Phosducin	19	239
homologs_49_173	196202	-	18	162
homologs_49_1744	182228	Creatinase/aminopeptidase	18	261
homologs_49_1763	-	-	18	142

Agalma orthologous clusters	*L. gigantea* gene ID	*L. gigantea* family description	No. of Taxa	Length (aa)
homologs_49_1799	195007	Cold shock DNA-binding domain-like	18	202
homologs_49_1813	234596	-	20	265
homologs_49_185	123544	-	21	311
homologs_49_1858	215949	Pancreatic lipase, N-terminal domain	20	346
homologs_49_189	56489	N-acetyl transferase, NAT	18	244
homologs_49_1899	196190	UBC-related	18	608
homologs_49_1936	210477	Translation initiation factor eIF4e	20	230
homologs_49_1964	115810	-	20	167
homologs_49_2028	138224	-	20	565
homologs_49_2040	203293	Cofilin-like	18	142
homologs_49_2049	98317	-	18	203
homologs_49_2101	171882	Ribosomal protein L9 N-domain	20	232
homologs_49_2138	204660	-	19	127
homologs_49_2147	211297	L30e/L7ae ribosomal proteins	20	129
homologs_49_2204	107145	-	18	171
homologs_49_2237	204839	Dimerization-anchoring domain of cAMP-dependent PK regulatory subunit	19	224
homologs_49_2277	160698	Ferredoxin domains from multidomain proteins	20	211
homologs_49_2361	208067	Glyceraldehyde-3-phosphate dehydrogenase-like, N-terminal domain	21	337
homologs_49_2363	96129	I set domains	20	369
homologs_49_2438	119072	Amylase, catalytic domain	21	706
homologs_49_2455	235322	-	19	313
homologs_49_2516	200724	-	19	200
homologs_49_2570	238259	-	18	196
homologs_49_2582	204146	-	18	191
homologs_49_2583	235240	PDI-like	18	221
homologs_49_2649	238741	Transcription factor IIA (TFIIA), beta-barrel domain	18	109
homologs_49_2650	228266	-	20	540
homologs_49_2663	147513	-	18	218
homologs_49_2672	238738	ZZ domain	19	811
homologs_49_2717	207719	Protein kinases, catalytic subunit	21	580
homologs_49_2724	157138	-	21	549
homologs_49_2729	138166	TNF-like	20	351
homologs_49_2776	218243	Kelch motif	18	429
homologs_49_2796	187941	-	18	145
homologs_49_2822	168166	Ribosomal protein L44e	18	112
homologs_49_288	210633	ATP synthase B chain-like	20	281
homologs_49_2890	120941	-	18	378
homologs_49_2915	218281	RNase Z-like	18	360
homologs_49_292	164694	-	20	202
homologs_49_2966	195736	-	18	790
homologs_49_3011	223519	-	19	375
homologs_49_3037	203798	Universal stress protein-like	19	148
homologs_49_3129	205046	eEF-1beta-like	20	270
homologs_49_315	234377	-	20	354

Agalma orthologous clusters	*L. gigantea* gene ID	*L. gigantea* family description	No. of Taxa	Length (aa)
homologs_49_3152	219822	Prefoldin	20	133
homologs_49_3212	185007	Cyclophilin (peptidylprolyl isomerase)	18	165
homologs_49_3213	183079	N-terminal domain of the delta subunit of the F1F0-ATP synthase	19	210
homologs_49_3268	213091	UFC1-like	20	173
homologs_49_3297	84508	STAR domain	18	304
homologs_49_3339	234815	-	18	206
homologs_49_3353	233342	DJ-1/PfpI	20	229
homologs_49_3370	210091	-	19	217
homologs_49_3376	175527	Fatty acid binding protein-like	18	139
homologs_49_3402	153781	-	20	198
homologs_49_3403	139948	Cyclophilin (peptidylprolyl isomerase)	19	175
homologs_49_3420	154157	BolA-like	19	144
homologs_49_3424	167413	tRNA(1-methyladenosine) methyltransferase-like	19	348
homologs_49_3486	59725	Canonical RBD	19	337
homologs_49_3488	229335	DnaQ-like 3'-5' exonuclease	18	188
homologs_49_3560	125208	Ribosomal protein L22	19	261
homologs_49_3564	125047	Ribosomal proteins L24p and L21e	20	242
homologs_49_3581	233176	Integrin A (or I) domain	20	428
homologs_49_3596	117918	-	20	165
homologs_49_3613	183671	-	19	207
homologs_49_3662	137449	Canonical RBD	21	494
homologs_49_3726	135559	-	20	205
homologs_49_3755	182759	-	18	133
homologs_49_3769	207726	Sm motif of small nuclear ribonucleoproteins, SNRNP	19	103
homologs_49_377	225752	-	20	189
homologs_49_3820	231994	-	18	195
homologs_49_3828	76938	Ribosomal protein L18 and S11	18	203
homologs_49_3837	201494	RbsD-like	19	150
homologs_49_3936	190107	FAH	20	218
homologs_49_3984	237171	G proteins	19	203
homologs_49_4001	110717	Ribosomal protein L4	21	336
homologs_49_4040	164490	Hydroxyisobutyrate and 6-phosphogluconate dehydrogenase domain	18	332
homologs_49_406	112447	Dcp2 box A domain	20	373
homologs_49_4199	227973	Cytochrome c oxidase Subunit F	18	146
homologs_49_4205	226441	Sm motif of small nuclear ribonucleoproteins, SNRNP	20	131
homologs_49_425	138684	-	19	186
homologs_49_4316	204915	Calmodulin-like	19	142
homologs_49_4369	237068	-	19	196
homologs_49_4439	104450	Phosducin	19	207
homologs_49_4458	205331	SNARE fusion complex	20	247
homologs_49_4476	213024	-	18	230
homologs_49_454	232327	BTB/POZ domain	20	120
homologs_49_464	218604	-	19	199
homologs_49_4679	229333	Canonical RBD	19	275

Agalma orthologous clusters	*L. gigantea* gene ID	*L. gigantea* family description	No. of Taxa	Length (aa)
homologs_49_4681	-	-	18	228
homologs_49_4683	198478	Isochorismatase-like hydrolases	18	193
homologs_49_4745	203703	PDI-like	18	358
homologs_49_4794	-	-	19	313
homologs_49_4806	228614	-	20	227
homologs_49_4844	237822	-	19	250
homologs_49_4872	97879	Prokaryotic ribosomal protein L17	19	190
homologs_49_4874	237623	Calcium ATPase, transmembrane domain M	20	659
homologs_49_4879	216676	Brix domain	19	301
homologs_49_4983	193902	eEF-1beta-like	19	245
homologs_49_5035	215214	SRP19	18	168
homologs_49_5038	115668	-	18	373
homologs_49_5079	149685	Glutathione S-transferase (GST), C-terminal domain	20	167
homologs_49_5080	183370	-	20	249
homologs_49_5162	238085	variant C2H2 finger	20	384
homologs_49_5222	179055	Creatinase/aminopeptidase	21	501
homologs_49_5229	202237	-	18	161
homologs_49_5237	133598	Histone H3 K4-specific methyltransferase SET7/9 N-terminal domain	18	323
homologs_49_5259	236781	tRNA-intron endonuclease catalytic domain-like	18	363
homologs_49_5271	196090	HMG-box	21	349
homologs_49_5301	197848	-	21	441
homologs_49_535	204770	-	19	175
homologs_49_5353	162533	Clp protease, ClpP subunit	18	232
homologs_49_5379	149249	Peptidyl-tRNA hydrolase II	18	213
homologs_49_5396	198754	Proteasome activator	20	258
homologs_49_541	133658	TBP-associated factors, TAFs	18	252
homologs_49_5437	187846	Translin	20	380
homologs_49_544	157712	Tetratricopeptide repeat (TPR)	20	301
homologs_49_5464	157959	F-box domain	19	344
homologs_49_5504	216578	Sm motif of small nuclear ribonucleoproteins, SNRNP	19	144
homologs_49_5697	238502	YgfY-like	20	156
homologs_49_572	122169	-	20	346
homologs_49_5730	113739	Tyrosine-dependent oxidoreductases	19	216
homologs_49_5758	231775	RING finger domain, C3HC4	20	451
homologs_49_5811	110395	ATP12-like	21	291
homologs_49_5815	140358	-	21	289
homologs_49_5851	206565	WD40-repeat	19	383
homologs_49_5902	144791	-	20	478
homologs_49_5935	145862	Ribosomal protein L10-like	20	280
homologs_49_5959	56299	-	18	228
homologs_49_5968	109061	Ankyrin repeat	19	223
homologs_49_5976	109474	Ribosomal protein S18	19	159
homologs_49_6005	66003	-	18	368
homologs_49_6017	217219	Prefoldin	20	160

Agalma orthologous clusters	*L. gigantea* gene ID	*L. gigantea* family description	No. of Taxa	Length (aa)
homologs_49_6071	184506	-	18	324
homologs_49_6130	228149	-	18	216
homologs_49_6141	230028	Rhomboid-like	21	431
homologs_49_6148	123653	Cyclin A/CDK2-associated p19, Skp2	20	523
homologs_49_6164	117398	-	19	434
homologs_49_6197	151243	Biotinyl/lipoyl-carrier proteins and domains	19	165
homologs_49_6204	218024	-	19	308
homologs_49_6242	199612	Myosin rod fragments	18	227
homologs_49_6246	120973	Ribonuclease H	20	265
homologs_49_6255	219544	-	19	260
homologs_49_6306	153375	Nuclear movement domain	18	152
homologs_49_634	119560	L23p	19	169
homologs_49_6347	201172	Thioltransferase	18	149
homologs_49_6454	204046	PX domain	19	171
homologs_49_65	234719	-	18	310
homologs_49_6603	217206	-	18	274
homologs_49_6611	233682	-	20	265
homologs_49_6614	192388	MTH938-like	18	222
homologs_49_667	205410	-	18	154
homologs_49_67	238013	Acylamino-acid-releasing enzyme, C-terminal donain	18	656
homologs_49_6707	230213	-	20	400
homologs_49_6753	201339	Frizzled cysteine-rich domain	18	306
homologs_49_6796	69719	-	19	181
homologs_49_6870	138644	-	19	373
homologs_49_6889	76788	Calponin-homology domain, CH-domain	18	235
homologs_49_6945	226808	Mannose 6-phosphate receptor domain	21	554
homologs_49_6955	94382	Type II chitinase	19	342
homologs_49_6964	175061	C-type lectin domain	19	340
homologs_49_7020	-	-	18	225
homologs_49_706	137721	Sedlin (SEDL)	18	147
homologs_49_7102	116635	PDZ domain	19	207
homologs_49_7136	191932	Mitochondrial ribosomal protein L51/S25/CI-B8 domain	20	171
homologs_49_7145	182189	WD40-repeat	21	336
homologs_49_7156	238643	TBP-associated factors, TAFs	19	194
homologs_49_7181	186221	Ribosomal protein S8	19	137
homologs_49_7249	120927	U2A'-like	18	265
homologs_49_7262	236234	Ubiquitin carboxyl-terminal hydrolase UCH-L	21	331
homologs_49_7309	139457	-	19	209
homologs_49_7327	237618	Transglutaminase core	18	228
homologs_49_7329	228218	PDZ domain	19	122
homologs_49_7354	195719	N-acetyl transferase, NAT	19	200
homologs_49_7383	110271	Ribosomal protein S7	20	253
homologs_49_7388	235760	Gar1-like SnoRNP	20	195
homologs_49_7404	219825	HesB-like domain	18	160

Agalma orthologous clusters	*L. gigantea* gene ID	*L. gigantea* family description	No. of Taxa	Length (aa)
homologs_49_741	151586	MIR domain	19	249
homologs_49_7516	175100	Nqo1 FMN-binding domain-like	21	485
homologs_49_7524	113219	Ribosomal protein L1	20	310
homologs_49_753	135579	Nqo5-like	20	272
homologs_49_7532	155468	PR-1-like	19	379
homologs_49_7561	192130	BAR domain	18	257
homologs_49_7611	214489	-	20	173
homologs_49_7675	236815	L30e/L7ae ribosomal proteins	19	116
homologs_49_7829	211364	NQO2-like	19	254
homologs_49_7891	205585	Tyrosine-dependent oxidoreductases	20	406
homologs_49_7919	110115	Calmodulin-like	20	512
homologs_49_793	163283	-	18	198
homologs_49_7952	58640	TBP-associated factors, TAFs	19	229
homologs_49_7968	78446	Ribosomal protein L30p/L7e	20	247
homologs_49_7987	229407	Proteasome subunits	20	246
homologs_49_7988	189106	Group II chaperonin (CCT, TRIC), ATPase domain	21	546
homologs_49_7993	201019	CAF1-like ribonuclease	20	293
homologs_49_8058	235765	Linker histone H1/H5	18	541
homologs_49_8077	152826	-	21	231
homologs_49_8093	208870	EMG1/NEP1-like	20	231
homologs_49_8096	203928	-	19	356
homologs_49_811	219308	Mago nashi protein	20	148
homologs_49_8113	161597	Canonical RBD	21	445
homologs_49_8114	140722	Tyrosine-dependent oxidoreductases	19	233
homologs_49_817	168894	-	20	142
homologs_49_8241	203978	HkH motif-containing C2H2 finger	20	126
homologs_49_8245	89428	-	18	219
homologs_49_8254	126319	-	20	303
homologs_49_8266	210661	Fe,Mn superoxide dismutase (SOD), C-terminal domain	20	224
homologs_49_8336	211907	Mitochondrial carrier	18	272
homologs_49_8351	196362	Aconitase iron-sulfur domain	21	785
homologs_49_845	222853	HMGL-like	20	333
homologs_49_8502	126607	mRNA capping enzyme	19	453
homologs_49_8550	200059	-	20	273
homologs_49_8606	173377	-	18	154
homologs_49_8626	116316	Ganglioside M2 (gm2) activator	20	233
homologs_49_8663	186993	Canonical RBD	18	158
homologs_49_8674	128587	STAT DNA-binding domain	20	790
homologs_49_8727	57020	Tetratricopeptide repeat (TPR)	21	391
homologs_49_8730	122770	-	18	206
homologs_49_8770	58748	Cold shock DNA-binding domain-like	18	146
homologs_49_8772	230099	Ribosomal L27 protein	18	150
homologs_49_8790	118333	Fibrinogen C-terminal domain-like	19	257
homologs_49_8865	68637	-	18	352

Agalma orthologous clusters	*L. gigantea* gene ID	*L. gigantea* family description	No. of Taxa	Length (aa)
homologs_49_8892	236136	Prefoldin	20	125
homologs_49_8897	191038	Glyoxalase I (lactoylglutathione lyase)	18	193
homologs_49_8910	185973	-	20	305
homologs_49_9025	130108	Phosphate binding protein-like	19	492
homologs_49_9038	89037	Frizzled cysteine-rich domain	18	329
homologs_49_9069	237930	RpoE2-like	19	117
homologs_49_9085	228995	Alcohol dehydrogenase-like, N-terminal domain	21	379
homologs_49_9164	-	-	19	235
homologs_49_9292	237294	-	18	308
homologs_49_9343	131082	DUSP, domain in ubiquitin-specific proteases	21	553
homologs_49_939	169544	-	18	226
homologs_49_9391	237630	Sedlin (SEDL)	19	217
homologs_49_9432	95635	CBM11	20	303
homologs_49_9461	129607	-	19	244
homologs_49_9544	206383	-	19	277
homologs_49_9548	217677	-	18	176
homologs_49_9618	239432	VPS37 C-terminal domain-like	18	281
homologs_49_9690	111730	Pumilio repeat	19	283
homologs_49_9723	184723	-	20	214
homologs_49_9764	140700	-	21	315
homologs_49_979	134785	Canonical RBD	18	223
homologs_49_9801	133714	-	21	300
homologs_49_9814	175108	MTH1598-like	20	163
homologs_49_9902	233342	DJ-1/PfpI	20	171
homologs_49_9994	182011	RBP11/RpoL	20	118

CHAPTER 3:

The pattern and pace of pulmonate evolution

3.1 ABSTRACT

The evolutionary relationships within the pulmonates, the air-breathing snails, have remained largely unresolved despite multiple morphological and molecular studies. Recent molecular studies have placed traditionally pulmonate and non-pulmonate taxa into Panpulmonata; however, the relationships within this new group are still poorly understood. Incongruence between studies has potentially resulted from morphological convergence, rapid cladogenesis, or a lack of informative loci. In this study I use a 500 nuclear gene dataset to investigate the pattern and timing of evolution within the highly diverse panpulmonate clade. I qualified the orthology of the 500 genes across a dataset of 79 newly sequenced and previously available transcriptomes. My dataset includes representatives of all major clades within Panpulmonata, including a wide representation of the stylommatophoran land snails, the most successful lineage of terrestrial molluscs. Maximum likelihood and Bayesian analyses confirm that Panpulmonata is monophyletic. Within Panpulmonata I reveal strong support for previously unsupported relationships, including Geophila, and the Pylopulmonata, a clade that unites the operculate panpulmonates. Molecular dating suggests a Permian or Early Triassic origin for Panpulmonata and a Triassic/Jurassic boundary origin for Eupulmonata and the freshwater Hygrophila. My analysis also suggests that Panpulmonata is indeed characterised by periods of relatively rapid cladogenesis, which occurred at the initial diversification of both Panpulmonata and the Stylommatophora.

3.2 INTRODUCTION

The pulmonates are a major lineage of snails and slugs within the Gastropoda that represent over 25,000 described species (Lydeard et al., 2010; Ponder and Lindberg, 2008). They are found globally (except Antarctica) in a wide range of habitats including marine, intertidal, mangrove, freshwater, and terrestrial environments, and are morphologically diverse, ranging from snails to limpets and slugs (Ponder and Lindberg, 2008). The evolutionary relationships among the pulmonate snails, however, have remained controversial despite a long history of scientific study (Haszprunar, 1985; Hubendick, 1979; Ponder and Lindberg, 2008; Ponder and Lindberg 2008, Schrödl, 2014; Solem, 1979; Tillier, 1984). Classically the pulmonates were considered a monophyletic lineage within the Heterobranchia – the ‘different-gilled’ snails within Gastropoda. A recent revision by Jörger et al. (2010) formally proposed the group Panpulmonata, which unites all pulmonate taxa with several lineages traditionally belonging to the Opisthobranchia, a polyphyletic lineage of non-air breathing marine slugs and related snails within the Heterobranchia, or the ‘Lower Heterobranchia’.

The monophyly of the Panpulmonata has since been supported by a number of molecular studies (Kocot et al., 2013b; Romero et al., 2016; Zapata et al., 2014). However, the major relationships within Panpulmonata remain largely unresolved (Figure 3.1). Specifically, that the Pulmonata do not form a monophyletic clade within Panpulmonata is yet to be adequately assessed with a phylogenomic dataset. While supporting Panpulmonata, the two phylogenomic studies with representatives of the panpulmonates have not been able to reject Pulmonata due to a lack of resolution (Zapata et al., 2014) or insufficient taxonomic sampling (Kocot et al., 2013b). Pulmonata has traditionally contained the Stylommatophora (terrestrial snails and slugs), the Systellommatophora (mostly intertidal and terrestrial slugs), and the Basommatophora, which comprise the Hygrophila (freshwater snails), the Siphonariidae (intertidal false limpets), and Amphiboloidea (intertidal and estuarine snails) (Hubendick, 1979; Solem, 1979). A polyphyletic Pulmonata would imply that morphological adaptions to breathing air have independently evolved multiple times across the pulmonates.

Several studies have suggested that the traditionally pulmonate Siphonariidae (marine false limpets) are basal within Panpulmonata and have a sister relationship with the Sacoglossa (sap-sucking sea slugs), termed ‘Siphoglossa’ (Jörger et al., 2010; Klussmann-Kolb et al., 2008). This relationship would imply that Pulmonata is not monophyletic and that the narrowed opening of the lung (the pneumostome) in the Siphonariidae evolved independent of the pneumostome in the Hygrophila and the Eupulmonata. Siphoglossa has only received support in Bayesian analyses based on limited gene datasets (Jörger et al., 2010; Klussmann-Kolb et al., 2008). Similarly, recent studies have also suggested a close relationship between the traditionally pulmonate Amphiboloidea, and the Glacidorbidae (minute freshwater snails) and Pyramidellidae (parasitic marine snails) (Dinapoli and Klussmann-Kolb, 2010; Jörger et al., 2010). This relationship would also imply that Pulmonata is not monophyletic but does unite the three Panpulmonate lineages which retain an operculum as adults. This clade has similarly only received support in Bayesian analyses based on small gene sets (Jörger et al., 2010; Klussmann-Kolb et al., 2008).

The Eupulmonata (sensu Bouchet and Rocroi, 2005) comprises three major lineages: 1) the Systellommatophora (intertidal and terrestrial slugs), 2) the Ellobiidea – marine, intertidal, and terrestrial snails which also include the Trimusculidae (intertidal limpets) and the Otinidae and Smeagolidae (intertidal snails and slugs), and, 3) the most successful molluscan lineage on land, the Stylommatophora. Eupulmonata is supported morphologically by the presence of a contractile pneumostome and characteristics of the central nervous system (Haszprunar and Huber, 1990), and has been supported in a number of molecular studies (Dinapoli and Klussmann-Kolb, 2010; Jörger et al., 2010; Klussmann-Kolb et al., 2008). However, the two most recent molecular studies to address the relationships within Eupulmonata found that Eupulmonata was not monophyletic (Dayrat et al., 2011; Romero et al., 2016). A non-monophyletic Eupulmonata would imply that the adaptations that have allowed the Systellommatophora and the Stylommatophora to transition to land have evolved independently. An alternative hypothesis, the Geophila, was first proposed by Férussac (1819). The Geophila, represented by a sister relationship between the Systellommatophora and the Stylommatophora, is based on several shared morphological characters and is supported in morphological analyses (Barker 2001; Ponder & Lindberg 2008), but no previous molecular study has supported this hypothesis.

The relationships within the Stylommatophora, the most speciose panpulmonate lineage, also remain largely unresolved (Tillier et al. 1996; Wade et al. 2001; 2006). The Stylommatophora were original divided into four separate groups based on the structure of the excretory system – the Sigmurethra, the Mesurethra, the Heterurethra, and the Orthurethra (Baker, 1955; Pilsbry, 1900). Of the four, only the Orthurethra and Heterurethra (as the Elasmognatha) are supported by molecular studies (Tillier et al. 1996; Wade et al. 2001; 2006). The only major relationship within the Stylommatophora that has received strong support in detailed phylogenetic analyses is the primary split between the achatinoid clade (including the families Achatinidae, Subulinidae, and Streptaxidae) and non-achantinoid clade (all other stylommatophorans, including the Orthurethra, Helicoidea, and Limacoidea; Wade et al., 2006). Morphological analyses have suggested that the Elasmognatha, a stylommatophoran lineage comprising the triangle slugs (Athoracophoridae) and the amber snails (Succineidae), are the basal stylommatophoran lineage (Barker 2001), however, this relationship has not been supported in molecular analyses (Wade et al. 2001; 2006). Given the fossil record, it has been suggested that the lack of resolution is due to a relatively rapid diversification event within the Stylommatophora (Tillier et al. 1996). No formal dating analysis has been conducted for the Stylommatophora and previous dating analyses for Panpulmonata have had either poor taxonomic representation (Zapata et al., 2014), or an insufficient number of molecular markers (Jörger et al., 2010; Tillier et al., 1996).

A lack of informative molecular markers for the pulmonates may be responsible for incongruence between studies (Figure 3.1). Several studies have shown that increasing the number of independent loci can aid phylogenetic reconstruction (Gontcharov et al., 2004; Leaché and Rannala, 2011; Wortley et al., 2005). The development of next generation sequencing technologies over the last decade have allowed for the fast and relatively inexpensive acquisition of large multi-locus phylogenetic datasets (e.g. Misof et al., 2014; O'Hara et al., 2014; Zapata et al., 2014). Recent phylogenomic studies which addressed relationships pulmonate relationships have either had limited representation of the major panpulmonate lineages (Kocot et al., 2013b) or have been unable to resolve the relationships with the group (Zapata et al., 2014). However, data from these studies, in conjunction with recent transcriptome studies focused on pulmonate lineages have greatly increased the resources available for phylogenetic analysis within Panpulmonata (Feldmeyer et al., 2011; Sadamoto et al., 2012; Teasdale et al., 2016). A recent study by Teasdale et al. (2016) identified a set of 500 orthologous genes suitable for phylogenetic analysis within the Eupulmonata. In this study I extend this orthologous gene set to investigate the patterns and timing of evolution across Panpulmonata. The dataset comprises transcriptome sequences for 79 taxa, representing the majority of the superfamilies within Panpulmonata, with a particular focus on the Stylommatophora. I resolve many of the evolutionary relationships within Panpulmonata using both maximum likelihood and Bayesian techniques, as well as different partitioning and subsetting schemes to investigate heterogeneity in phylogenetic signal within the dataset. I also conduct a dating analysis, using fossil calibrations, to investigate the timing of the diversification within Panpulmonata.

3.3 MATERIALS AND METHODS

3.3.1 Tissue collection and sequencing

I sequenced transcriptomes for 46 species and supplemented this dataset with 33 transcriptomes sequenced in previous studies (Table 3.1). The complete dataset contained transcriptomes from 79 species representing 65 families and 47 superfamilies, including 10 basal heterobranch species as outgroups (Table 3.1). This dataset includes representatives of 84% of all previously recognised superfamilies within Panpulmonata. To sequence the transcriptomes, total RNA was first extracted from foot or whole body tissue stored in RNAlater (Ambion Inc, USA) using the Qiagen RNeasy extraction kit (Qiagen, Hilden, Germany). Libraries were constructed using the TruSeq RNA sample preparation kit v2 (Illumina Inc., San Diego, CA), and sequenced on the Illumina HiSeq 2000 platform (100 bp paired end reads). I removed adaptor sequences and trimmed low quality bases from the reads using the program Trimmomatic (v0.22 and v0.32; Lohse et al. 2012). Reads shorter than 36 bp after trimming were discarded. I assembled the transcriptomes using the *de novo* transcriptome assembler Trinity (v2012-06-08 and r2013-08-14; Grabherr et al. 2011; Haas et al. 2013) using the default settings. The number of raw reads, trimmed reads, and assembled contigs for each sample are presented in Table 3.1. Of the 33 transcriptomes which were sequenced in previous studies, I trimmed and assembled the reads for 10 transcriptomes from Zapata et al. (2014) and used the published assemblies for the additional 23 transcriptomes (see Table 3.1).

3.3.2 Orthology determination and gene qualification

While the 500 nuclear genes have been qualified as single copy for the eupulmonates in Teasdale et al. (2016), the additional taxa in the current dataset may contain paralogous sequences for these genes. I therefore qualified orthology for the broader dataset using a procedure similar to that used in the original study (Teasdale et al., 2016), which included visual inspection of alignments of homologous sequences followed by screening gene trees for hidden paralogs. First, I identified all contigs homologous to the 500 genes by using BLAST (Blastx, cut off e-value of e-10; Camacho et al., 2009) to compare each transcriptome assembly to the predicted gene models from the owl limpet genome *(Lottia gigantea*; Simakov et al., 2013). All contigs with matches to the 500 genes were appended to the respective gene alignment of the 18 eupulmonate species from the original study (Teasdale et al., 2016). The sequences were then translated into amino acids and aligned using ClustalW in BioEdit v7.1.3 (Hall, 1999). I identified and removed untranslated regions and corrected frame shifts manually. Consensus sequences were produced where multiple overlapping contigs within a sample did not differ by more than three nucleotides. Non-overlapping fragments were also concatenated if there were no other contigs for that sample aligned to the same region and they were more similar to sequences from the same superfamily lineage.

Each alignment was then visually assessed for the presence of paralogous sequences. Neighbour joining trees constructed in MEGA were used to identify ‘out-paralogs’ (i.e. sequences resulting from a duplication event which occurred prior to the common ancestor of the study group), ‘in-paralogs’ (i.e., sequences resulting from a duplication event which occurred on a terminal branch; Remm et al. 2001), and contamination (i.e. mis-indexing and sequences which represented taxa from different phyla). The *Tornatellinops jacksonensis* (Achatinellidae) samples exhibited the highest level of contamination with fly sequences present for over half the 500 genes. In all cases out-paralogs and contamination were removed from the alignments. I identified 40 cases of in-paralogs, with *Triboniophorus graeffei* (Athoracophoridae) having the majority (21 cases), and in each case, as either in-paralog should reconstruct the same phylogenetic relationships, the longest sequence was retained. Within a sample, overlapping and divergent contigs that were not out-paralogs, in-paralogs or contaminants would have been regarded as paralogous sequences arising from duplication events younger the common ancestor of Panpulmonata, however, no such case of paralogy was found for this extended dataset. Finally, ambiguously aligned regions were manually masked and excluded from downstream analyses. I also used Aliscore (Kück et al., 2010), employing the default settings, to remove any remaining ambiguously aligned regions from the alignment for each individual gene.

As a final phylogenetic check for spurious sequences I screened gene trees for additional paralogs, contaminants, and miss-indexing, using TreSpEx (Struck, 2014). Specifically, for each gene I constructed a maximum likelihood tree in RAxML v8.2.4 (Stamatakis, 2014), using the LG + Γ model (100 fast bootstraps). I then used TreSpEx to search for well-supported nodes that were in conflict with several undisputed clades within Panpulmonata, including currently recognised families and superfamilies. Where any such conflict occurred, the relevant gene tree and alignment was visually assessed to determine whether a paralogous sequence was responsible for the conflict. The TreSpEx analysis detected no cases of undetected paralogy but did identify four cases of mis-indexing, where the sequence from one sample was present in the assembly of another. Mis-indexed sequences were removed from the dataset.

3.3.3 Dataset partitioning and phylogenetic analysis

The final data matrix represented 159,008 amino acids and was highly complete (Figure 3.2). I conducted the partitioning and phylogenetic analyses of the alignment in amino acids as 78% of the 500 genes were saturated at the nucleotide level (c-value; Kück and Struck, 2014). Partitioning of phylogenomic datasets typically involves grouping genes into larger partitions; however, it is likely that a lot of the variation relevant to choosing a substitution model occurs within a gene rather than among genes (Misof et al., 2014; O'Hara et al., 2014). Previous studies have shown that breaking up a gene into biophysical domains or exons results in better partitioning of variance and hence a better fitting partitioning scheme (Misof et al., 2014; O'Hara et al., 2014). Here I take a similar approach to O'Hara et al. (2014) and use exons as a simple scheme to break genes into smaller units.

I divided the concatenated amino acid alignment into exons based on boundaries delineated for the stylommatophoran land snail family the Camaenidae (see Teasdale et al., 2016) and the owl limpet genome (*Lottia gigantea*), resulting in 4,464 exons. The exons were reduced to 3,195 initial partitions by concatenating small exons (< 25 amino acids) to adjacent exons within the respective gene. Chi-squared tests for homogeneity of base composition conducted in BaCoCa v1.1 (Kück and Struck, 2014) showed no significant deviation from homogeneity for any exon and the overall relative composition frequency variability (RCFV) value, which represents the extent of compositional heterogeneity in amino acid frequency across clades, was low (0.0124; Appendix 3.1). To cluster the exons I first calculated the RCFV values, and the frequencies of hydrophobic, hydrophilic, polar, nonpolar, positive, neutral, and negative amino acids, per exon in BaCoCa. Using these per exon statistics I hierarchically clustered the 3,195 exons using Ward's method (Appendix 3.2), and then determined the optimum number of clusters (partitions) using the Kelley-Gardener-Sutcliffe penalty function (as implemented in maptree library in R; Kelley et al., 1996). The objective of this penalty function is to simultaneously minimise both the overall number of clusters and the dissimilarity among members within each cluster (Appendix 3.3). The clustering resulted in eight exon partitions that ranged in size from 4,898 to 42,785 amino acids.

For each partition, the best fitting amino acid substitution model was chosen based on the Bayesian information criterion (BIC), as implemented in PartitionFinder v1 (Lanfear et al., 2012). Using the resulting models, I conducted a Bayesian phylogenetic analysis using the program ExaBayes (Aberer et al., 2014). I ran four Metropolis-coupled ExaBayes replicates, with four chains each (three heated), for 600,000 generations sampling every 1,000 generations. Using the 'postProcParam' tool included with the ExaBayes package, I checked for convergence and adequate sampling of the posterior distribution of the parameter values by ensuring that the effective sample sizes (ESS) of all estimated parameters were greater than 200 and that the average standard deviation of split frequencies and potential scale reduction factors across runs were close to zero and one respectively. A consensus tree was created by combining the trees from the four separate runs, with the first 25% removed as burn-in in each case, using the ‘consensus’ tool included with the ExaBayes package. Maximum likelihood analysis was performed using the same eight exon partitioning and model scheme using RAXML v8.2.4 (Stamatakis, 2014). I considered a bootstrap of ≥ 75 as moderate support for a node, a bootstrap of ≥85 strongly supported, and a bootstrap of 100 unequivocal.

As the partitioning scheme can have an impact on the phylogenetic reconstruction (Kainer and Lanfear, 2015), I also constructed a maximum likelihood tree for the full concatenated dataset using the LG4X amino acid substitution model (Le et al., 2012) with no partitioning. This model allowing for heterogeneity along the amino acid sequences by using four different matrices (as opposed to one). With many loci we can also test the robustness of the phylogenetic reconstructions by examining the congruence among subsets of the data (Edwards, 2016). I assessed congruence between different subsets of the data using two approaches: 1) by comparing separate maximum phylogenies for each of the eight exon partitions (estimated using RAxML), and 2) by assessing the support for specific nodes using partitioned likelihood support (PLS; Lee and Hugall, 2003). I used PLS to compare strongly supported conflicting relationships among the eight exon partition phylogenies. I calculated per site likelihoods for each topology using RAxML and compared the summed likelihood values for each exon within each of the eight data partitions to determine whether a subset of the data was driving conflict in the analyses. The significance of the differences in likelihood support for the alternate hypotheses was tested using the Approximately Unbiased (AU) test as implemented in CONSEL (Shimodaira and Hasegawa, 2001).

3.3.4 Molecular dating and fossil calibration

I conducted the dating analysis using the approximate likelihood calculation algorithm implemented in MCMCTREE, part of the PAML package v4.8 (Yang, 2007). I used the topology resulting from the partitioned RAxML analysis as the fixed topology. Information from five fossils was used to set node age priors (Table 3.2). The ages of these fossils provided a minimum estimate for the time of divergence for the fossil’s assigned lineage and its sister clade (i.e. minimum stem calibrations). The minimum priors had soft bounds with a left tail probability of 2.5% and had a truncated Cauchy distribution, which approximates a uniform distribution using the default settings. To provide root constraints I used estimates derived from a broader phylogenomic study of Gastropoda (Zapata et al. 2014), which utilised fossil calibrations for deep nodes within the Gastropoda and the Mollusca. Specifically, I used the estimate of 344 Ma (95% range: 302.5 - 388.3 Ma) as a normally distributed prior for the basal split between the Architectonicidae and all remaining taxa (i.e. the root node). I set a maximum root age of 420 my, reflecting the estimated split between the Heterobranchia and the Caenogastropoda from Zapata et al. (2014). I conducted a partitioned analysis where substitution rates were estimated for each of the eight exon partitions separately, with the JTT + Γ model of sequence evolution assigned to each partition. I used the uncorrelated lognormal relaxed clock model and specified a birth–death speciation process as the tree prior with default parameters (death and growth rate parameters set as 1, and sampling parameter set as 0). I ran MCMCTREE twice independently, each time for 20 million generations, sampling every thousand and discarding the first 2000 samples as burn-in. I checked for convergence by plotting the correlation between the posterior means of the node ages for the two runs.

3.3.5 Morphological analyses

I reconstructed the ancestral states for three morphological characters, namely: 1) the presence of an operculum, 2) the structure of the opening of the pallial cavity, 3) the presence of a closed secondary ureter (see Appendix 3.25 for a detailed description of the states of each morphological character). The morphological characters were determined for each family represented in my data set and the data was obtained from the review of the extensive literature (>500 publications) and extensive observations of anatomy (G.M. Barker pers. observ.). Each morphological character was mapped onto the maximum likelihood topology obtained from the partitioned analysis with tips representing the same family collapsed. As some families were polymorphic for certain characters, i.e. contained species with different states, the ancestral state reconstructions were conducted using parsimony as implemented in MESQUITE v3.10 (Maddison and Maddison, 2016). I also mapped the current habitat type for each family onto the tree but did not re construct the ancestral states of these traits.

3.4 RESULTS

3.4.1 Deep relationships among pulmonates

The present analyses do not recover a monophyletic Pulmonata as traditionally defined (Figure 3.2), as Amphiboloidea forms a well-supported clade with traditionally non-pulmonate taxa, Glacidorbidae and Pyramidellidae. Instead, there is unequivocal support (BPP = 1, BS = 100) for Panpulmonata, uniting non-air-breathing lineages with pulmonates, and Sacoglossa as the basal lineage. These relationships were consistently recovered from all analyses, including the partitioned, unpartitioned, and individual partition maximum likelihood (ML) analyses, and the Bayesian analysis (Figure 3.2, Appendix 3.5, 3.6, 3.7). The relationships between the remaining panpulmonate clades, namely the Siphonariidae, the Acochlidiacea, the Hygrophila, and the Eupulmonata, remain uncertain. Based on the Bayesian analysis all relationships are well resolved, with the sacoglossans being most basal within Panpulmonata, followed by Siphonariidae, then by Amphiboloidea + Glacidorbidae + Pyramidellidae, then by Acochlidiacea, and finally the sister relationship between Hygrophila and Eupulmonata. While topologically consistent, the deep relationships within Panpulmonata were not strongly supported in the ML analyses. However, there was no strongly supported conflict in topology across the eight exon partitions. Given the size of the dataset, moderate support with little conflict is likely due to close divergences among lineages. A plot of bootstrap support against internode length supports this pattern (Appendix 3.19). Partitioned likelihood support (PLS) analyses showed that an alternative placement for Siphonariidae as sister to the Hygrophila could not be rejected (p-values ranged from 0.057 to 0.439, Figure 3.3 a). Per exon differences in likelihood showed that the bulk of the exons contained little phylogenetic information informative for discriminating between the two topologies in the PLS analysis (Figure 3.3 a).

3.4.2 Relationships within Stylommatophora

Within Eupulmonata, two major lineages are unequivocally supported by all analyses: 1) a clade comprising Ellobidae, Smeagolidae, and Trimusculidae, and 2) the Geophila, which comprises the monophyletic Systellomatophora and monophyletic Stylommatophora. Within the Stylommatophora itself there is unequivocal support for a sister relationship between the ‘achatinoid’ clade and the rest of the Stylommatophora confirming that the informal group Sigmurethra is paraphyletic. Most major lineages with multiple representatives in my dataset have unequivocal support, including the superfamilies Rhytidoidea, Punctoidea, Limacoidea, Orthalicoidea, and Helicoidea, and the unranked clades Elasmognatha and Orthurethra. The Bayesian and ML analyses both support a clade comprising the Helicoidea, Elasmognatha, and Orthalicoidea (Figure 3.2) that is sister to a strongly supported clade comprising the rest of the ‘non-achatinoid’ stylommatophorans. The only topological differences between the Bayesian and full dataset ML analyses regard the placement of two lineages: the Caryodidae and the Limacoidea + Testacellidae. The topological placement of these lineages is not supported in either analysis.

Individual phylogenies inferred for each of the eight exon partitions were largely consistent with the analyses based on the full dataset. However, one of the eight partitions showed strong support for a sister relationship between the Elasmognatha and the rest of the non-achatinoid Stylommatophora rather than a sister relationship with the Helicoidea as shown in the full dataset Bayesian and ML analyses. Partitioned likelihood support (PLS) analyses showed that only one partition significantly rejected the basal placement of the Elasmognatha (p-value = 0.001, Figure 3.3 b).

3.4.3 Molecular dating

The chronogram inferred by MCMCTREE suggests a probable Permian or Early Triassic origin for Panpulmonata (Figure 3.4). Panpulmonata diversified into the major lineages during the Triassic with Eupulmonata originating by the Early Jurassic and Hygrophila originating slightly later within the Jurassic (Figure 3.4). The most recent common ancestor of the Stylommatophora occurred in the Late Jurassic. The fossil calibrations used in the analysis all fell within the posterior distribution of the age of the relevant node except for the *Lymnaea* fossil. The median node age for the split between the Lymnaeidae and the Planorbidae was approximately 40 million years younger than the minimum age of the fossil calibration, but the boundary of the 95% confidence interval was only approx. 10 million years younger. This result suggests that the *Lymnaea* fossils might be incorrectly dated, or that they belong to an earlier hygrophilan lineage.

3.4.4 Morphological analyses

Ancestral state reconstructions of three morphological traits identified a number of key morphological transitions (Figure 3.5). I confirm that the Glacidorbidae, the Pyramidellidae, and the Amphiboloidea are the only panpulmonates to retain an operculum as adults. I also show that a narrowed opening to the mantle cavity (termed pneumostome) appears to have evolved three times: it is present in the brackish water Amphiboloidea, the marine to intertidal Siphonariidae, and the common ancestor of the fresh water Hygrophila and the mostly terrestrial Eupulmonata. The contractile pneumostome has evolved three times independently, once in the Hygrophila, once in the Siphonariidae, and once in the common ancestor of the mostly terrestrial Eupulmonata. The closed secondary ureter (or varying lengths within the mantle cavity) is shared by the majority of the Stylommatophora, except for the Orthurethra, where this feature of the excretory system has been lost.

3.5 DISCUSSION

The higher systematics of pulmonates and their phylogenetic position within the Heterobranchia have been controversial and in a state of flux for the greater part of the last century (Baker, 1955; Haszprunar, 1985; Klussmann-Kolb et al., 2008; Pilsbry, 1900; Ponder and Lindberg, 1997; Solem, 1979). A recent molecular phylogenetic analysis by Jörger et al. (2010) proposed the informal group Panpulmonata, grouping four lineages traditionally considered as opisthobranchs or 'lower heterobranchs' with the air-breathing pulmonates. In a similar manner, relationships within Stylommatophora, the most diverse lineage within the eupulmonates, have been difficult to resolve (Tillier et al., 1996; Wade et al., 2006, 2001). My analyses show clear support for the monophyly of the Panpulmonata, consistent with previous studies using mitochondrial and rRNA loci (Dinapoli and Klussmann-Kolb, 2010; Jörger et al., 2010; Klussmann-Kolb et al., 2008), as well as phylogenomic scale datasets (Kocot et al., 2013b; Zapata et al., 2014). I also show support for many major relationships within Panpulmonata (discussed below), and provide time estimates for divergences within the group, spanning 300 Ma of evolution.

3.5.1 Phylogenetic relationships within Panpulmonata

The sacoglossan sea slugs, traditionally regarded as opisthobranchs, were formally included in Panpulmonata by Jörger et al. (2010), with analyses suggesting a sister relationship between the Sacoglossa and the traditionally pulmonate Siphonariidae. Other molecular studies have also suggested this relationship – the combined clade termed Siphoglossa (Klussmann-Kolb et al., 2008; Medina et al., 2011) – however it has only received support based on a limited number of loci. A more recent phylogenomic study (Kocot et al., 2013b) recovered strong support for a sister group relationship between Siphonariidae and all sampled panpulmonates excluding the Sacoglossa, but many major panpulmonate lineages were not represented in their dataset. Here, based on greater taxonomic sampling, I confirm Kocot et al.'s (2013) finding of a basal Sacoglossa within Panpulmonata. The present analyses do not support the Siphoglossa hypothesis and a sister relationship between the Sacoglossa and Siphonariidae was not evident in any of the eight exon partition topologies. The Sacoglossa and Siphonariidae share a one-sided plicate gill which is absent from all other panpulmonates (Dayrat and Tillier, 2002). As discussed in Kocot et al. (2013), the basal position of the Sacoglossa and Siphonariidae is consistent with the hypothesis that the early panpulmonates retained an opisthobranch-type gill that was subsequently lost in the rest of the panpulmonates. However, in my analysis the relationships between Siphonariidae and the remaining major panpulmonate lineages are less certain.

While topologically consistent, the Bayesian analyses showed strong support for the deep relationships within Panpulmonata whereas the ML analyses only showed moderate support at best. Relationships with moderate support in the ML analyses included a sister relationship between the Eupulmonata and Hygrophila, and the monophyly of all panpulmonates excluding the Sacoglossa and Siphonariidae. Bayesian posterior probabilities and non-parametric bootstrap support, while both representing confidence in the phylogeny, are not equivalent and represent different properties of the dataset (García-Sandoval, 2014). High posterior probabilities but low bootstrap support at certain nodes suggests that a proportion of the dataset does not contain phylogenetic information relevant to the nodes in question (García-Sandoval, 2014). The partitioned likelihood support (PLS) analysis showed a general lack of phylogenetic information regarding the placement of the Siphonariidae and showed no evidence of strong conflict within the dataset. The genes are not highly conserved (Teasdale et al., 2016), therefore it is probable that the lack of phylogenetic information is the result of relatively rapid cladogenesis. This hypothesis is supported by the relationship between branch length and bootstrap support for the ML tree, with the shortest branches having smaller bootstrap support values. Increased taxonomic sampling may aid further interrogation of the rate of diversification within Panpulmonata and to determine whether a true polytomy likely exists. In addition, several of the lineages found to be problematic in my analyses are only represented by one (e.g. Siphonariidae) or two samples (e.g. Acochlidiacea).

Previous molecular studies have suggested, but with poor support, the grouping of the Pyramidellidae, the Glacidorbidae, and the traditionally pulmonate Amphiboloidea within one clade (Dinapoli et al., 2011; Jörger et al., 2010). This clade, uniting the only panpulmonate lineages that retain the plesiomorphic operculum as adults, received unequivocal support across my analyses. Hence, I here refer to this clade as the 'Pylopulmonata' (derived from pyle (Gr) – a gate). While this grouping has not been recovered in previous morphological analyses, the Glacidorbidae were originally suggested to be closely related to the Amphiboloidea due to the presence of the operculum and the lack of a separate bursa copulatrix (Ponder, 1986). Both the Glacidorbidae and the Pyramidellidae have a widely open pallial cavity, whereas the Amphiboloidea have a pallial cavity narrowed to a pneumostome. It is likely that the pneumostome in the Amphiboloidea is an adaptation to the semi-terrestrial brackish water environment and has evolved independently of the pneumostome that has become contractile in the higher freshwater Hygrophila and intertidal to terrestrial Eupulmonata (Barker 2001).

Our analyses show moderate support for the sister relationship between Eupulmonata (sensu Bouchet and Rocroi, 2005) and the Hygrophila. This relationship is not consistent with any previous molecular phylogeny (including those presented in Figure 3.1) or morphological analysis (Barker 2001; Dayrat and Tillier, 2002). Support for the clade Eupulmonata has been shown in several previous molecular studies (Dinapoli and Klussmann-Kolb, 2010; Holznagel et al., 2010; Jörger et al., 2010; Klussmann-Kolb et al., 2008). However, the most recent studies to address these relationships found that Eupulmonata was not monophyletic (Dayrat et al., 2011; Romero et al., 2016). In the present study, we show unequivocal support for the monophyly of the Eupulmonata, however, my analysis revealed phylogenetic relationships within Eupulmonata that differ from all previous molecular studies. The Geophila hypothesis was first proposed by Férussac (1819), and has since been recovered in at least one morphological analysis (Barker, 2001; Ponder and Lindberg, 2008), but has not been supported in a molecular phylogeny. My study provides unequivocal support for the 'Geophila' hypothesis, where the Systellommatophora is sister to the Stylommatophora as found by Barker (2001). By contrast, previous molecular studies revealed only weakly supported relationships within Eupulmonata or suggested that the Systellommatophora were sister to a clade containing the Ellobiidae, Otinidae and Trimusculidae (Dinapoli and Klussmann-Kolb, 2010; Jörger et al., 2010; Klussmann-Kolb et al., 2008) (Figure 3.1). The Systellommatophora and Stylommatophora share a number of morphological characters that are absent from the Ellobiidae, Trimusculidae and Otinidae. These shared characters include eyes at the end of the cephalic tentacles rather than at the base, an unpaired jaw, and a long pedal gland located on the floor of the visceral cavity (Ponder and Lindberg 2008).

3.5.2 Relationships within Stylommatophora

Consistent with Wade et al. (2006, 2001), the present analyses show unequivocal support for the monophyly of the Stylommatophora, with the 'achatinoid' clade – which in this dataset comprises the Achatinidae, the Subulinidae, and the carnivorous Streptaxidae – sister to all other Stylommatophora. The Wade et al. (2006, 2001) studies showed support for the monophyly of several major ‘non-achatinoid’ stylommatophoran lineages, including the Limacoidea, the Helicoidea, the Elasmognatha, the Orthurethra, and the Orthalicoidea, but were unable to resolve the relationships between these clades. In the present study, I confirm the monophyly of these major clades and suggest that the ‘non-achatinoid’ stylommatophorans form two major clades comprising: 1) the Helicoidea, the Orthalicoidea, and the Elasmognatha, and 2) the rest of the ‘non-achatinoid’ stylommatophorans including the Limacoidea, the Orthurethra, the Punctoidea, and southern hemisphere lineages such as the Caryodidae, Oopeltidae, and Rhytidoidea. Previous morphological analyses have suggested a close relationship between the Helicoidea and Orthalicoidea (represented by Bulimulidae; Barker 2001). However, the placement of Elasmognatha as sister to the Helicoidea is inconsistent with morphological analyses that suggested that the Elasmognatha were sister to all other stylommatophorans (Barker 2001). Wade et al. (2006, 2001) were also unable to resolve the basal non-achatinoid relationships. The partitioned likelihood support (PLS) analyses support the uncertainty in the placement of the Elasmognatha, as only one of the eight exon partitions significantly rejected a sister relationship between the Elasmognatha and the rest of the non-achatinoid Stylommatophora. Similar to the deeper relationships within Panpulmonata, it is plausible that this lack of resolution is due to short internode branch lengths, which may indicate relatively rapid cladogenesis early in the diversification of the non-achantinoid Stylommatophora. Previous molecular analyses (Tillier et al., 1996; Wade et al., 2006, 2001) have also shown short branch lengths early in the diversification of the non-achantinoid Stylommatophora but, as they relied on a small number of loci, where unable to determine whether the lack of resolution was due to a lack of data and the limitations of using nuclear ribosomal RNA, or a true lack of phylogenetic signal.

3.5.3 Timing of panpulmonate evolution

The molecular dating analysis suggests that Panpulmonata originated during the Permian or the Early Triassic (mean age = 262 Ma; 95% = 232 – 288 Ma). Two previous dating analyses suggested a slightly later origin during the Triassic (~223 Ma - Jörger et al., 2010; ~200 Ma - Zapata et al., 2014). The boundary between the Permian and Triassic represents the largest known mass-extinction event of marine organisms. Extinction of many bivalves and gastropods, as well as the trilobites, was putatively driven by ocean acidification (Clarkson et al., 2015). This extinction event has been linked to diversification events in other taxa (O'Hara et al., 2014) but an apparent diversification event after a mass-extinction event, according to simulation studies, may be explained by extinction alone without adaptive diversification (Crisp and Cook, 2009). Denser sampling of the major panpulmonate lineages would be needed to test these competing hypotheses. Understanding the drivers of diversification during the Triassic is difficult as there are very few confirmed early panpulmonates in the fossil record. The first confirmed appearance of the Siphonariidae occurs in the early Cretaceous (Kaim, 2004), but an origin as far back as the late Triassic would still be consistent with the present analysis. There are earlier records of the Siphonariidae assigned to the genus *Rhytidopilus* (see Tracey et al., 1993), however, they were later discounted (Sepkoski 2002; Dayrat et al. 2011). The family is presently not assigned to a superfamily (Bouchet and Rocroi, 2005). The sacoglossans only appear in the fossil record about 50 Ma (Le Renard, 1983) but this is unsurprising given that the shells (when present) are small and fragile, and, as is true for most panpulmonate taxa, the shells consist of aragonite which does not fossilise well (Ponder and Lindberg 2008).

The extant crown group of the Eupulmonata emerged around the Triassic-Jurassic boundary (~200 Ma), a date which is not in conflict with the fossil record. The earliest eupulmonate record dates to approx. 145 Ma in the late Jurassic (*Otina* sp.; Tracey et al., 1993). The Triassic-Jurassic boundary is associated with another major extinction event, which eliminated approximately half of all marine genera and many terrestrial vertebrates. Pangea also began to break up at this time, resulting in more coastline and a wetter climate in the Jurassic compared to the Triassic (Seton et al., 2012). The major driver of diversification in Eupulmonata might, however, be the acquisition of the contractile pneumostome, which allows the opening of the pallial cavity to be closed, aiding moisture retention and ventilation of the lung. This morphological adaptation, in conjunction with a high tolerance for physiological extremes (Little, 1983), allows the eupulmonates to inhabit a large variety of habitats. Several lineages within Eupulmonata are strictly terrestrial. The transition of the eupulmonates onto land has occurred relatively recently compared to other terrestrial lineages. Non-pulmonate molluscan lineages, including *Dawsonella* (Caenogastropoda), were terrestrial in the carboniferous (Gordon and Olson, 1995). In addition, the arthropods, one of the first clades to colonise land (Engel and Grimaldi, 2004), and the terrestrial tetrapods were present on land in the Devonian (Ahlberg and Milner, 1994).

The freshwater Hygrophila began to diversify in the Early Jurassic (~185 Ma), shortly after the eupulmonates. It is likely that the transition to freshwater happened at least once within the Hygrophila, potentially during the Jurassic, as the extant Hygrophila share the same adaptations to living in freshwater (e.g. a narrowed mantle cavity opening and a separate region of the kidney specialised for the reabsorption of water and salts). The other two freshwater lineages in Panpulmonata, the Glacidorbidae and the fresh water lineage in the Acochlidiacea, represent transitions independent from the Hygrophila. The Glacidorbidae have adaptations for life in freshwater but when this transition occurred is unclear as the first fossils are from the Miocene (Ponder and Avern, 2000). The Acochlidiacea potentially secondarily lost the pallial cavity as an adaptation to the interstitial existence (although a number of lineages are benthic). Despite this one acochlidiid lineage colonised freshwater in the Paleogene (Jörger et al., 2014) and a new species was recently discovered living on land in the humid tropical rainforest on a small island in Palau (Kano et al., 2015).

While many eupulmonates are terrestrial, the most successful molluscan lineage on land is the Stylommatophora (>20,000 species; Rosenberg, 2014). Previous studies have suggested a rapid diversification within the Stylommatophora, given that morphological and molecular studies have struggled to provide resolution and evidence from the fossil record (Tillier et al. 1996, Ponder and Lindberg 2008). As most extant stylommatophoran families appear in the fossil record in the Early Cenozoic, Tillier et al. (1996) suggested that an ‘explosive’ radiation from a single older eupulmonate lineage may have occurred approximately 60 Ma. My analysis shows that the crown diversification of the Stylommatophora began in the Late Jurassic with the split of the achatinoid and non-achatinoid lineages (~160 Ma). There are Stylommatophoran fossils present as early as the Upper Cretaceous (~85 Ma; Tracey et al., 1993) and often multiple genera from the same family are found in the same layer which is consistent with earlier diversification (Pan 1977; Salvador and Simone, 2013; Stworzewicz et al., 2009). Within the Stylommatophora there then appears to have been two putative concentrations of diversification. First a relatively rapid diversification in the Early Cretaceous (~130 Ma), as demonstrated by the relatively short internode branch lengths, topological uncertainty and the relatively low bootstrap support. Many of the modern families then appear at the Cretaceous/Cenozoic boundary (~60 Ma), consistent with the fossil record, although denser sampling of the stylommatophoran families is needed to assess diversification rates within the Stylommatophora. These dating estimates, however, provide a framework that will facilitate future studies investigating the pattern and rate of evolution within Panpulmonata.

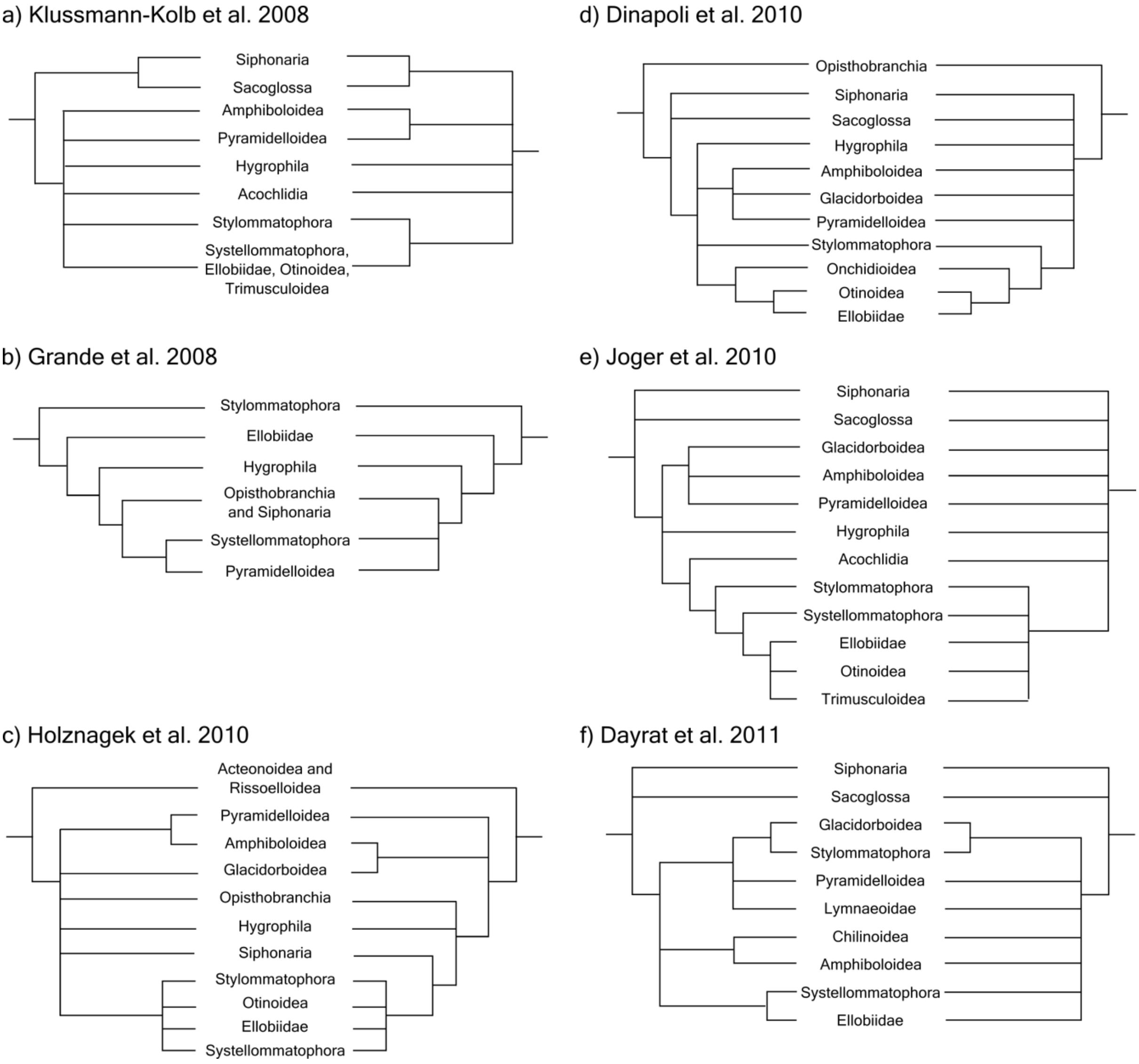


Figure 3.1. Summary of previous phylogenies for Panpulmonata. For each study the maximum likelihood analysis is presented on the left and the Bayesian analysis on the right. Only nodes with ≥ 75 bootstrap support are shown for the maximum likelihood analyses and only nodes with ≥ 0.95 posterior probabilities are shown for the Bayesian trees. The studies vary in the genetic data used: a) Klussmann-Kolb et al. (2008) – 18S and 28S rRNA, and mitochondrial 16S rRNA and CO1, b) Grande et al. (2008) – 12 mitochondrial protein-coding genes, c) Holznagel et al. (2010) – 28S rRNA, d) Dinapoli et al. (2010) – 18S and 28S rRNA and mitochondrial 16S rRNA and CO1, e) Jörger et al. (Jörger et al., 2010) – 18S and 28S rRNA and mitochondrial 16S rRNA and CO1, f) Dayrat et al. (2011) – 18S rRNA and mitochondrial 16S rRNA and CO1.

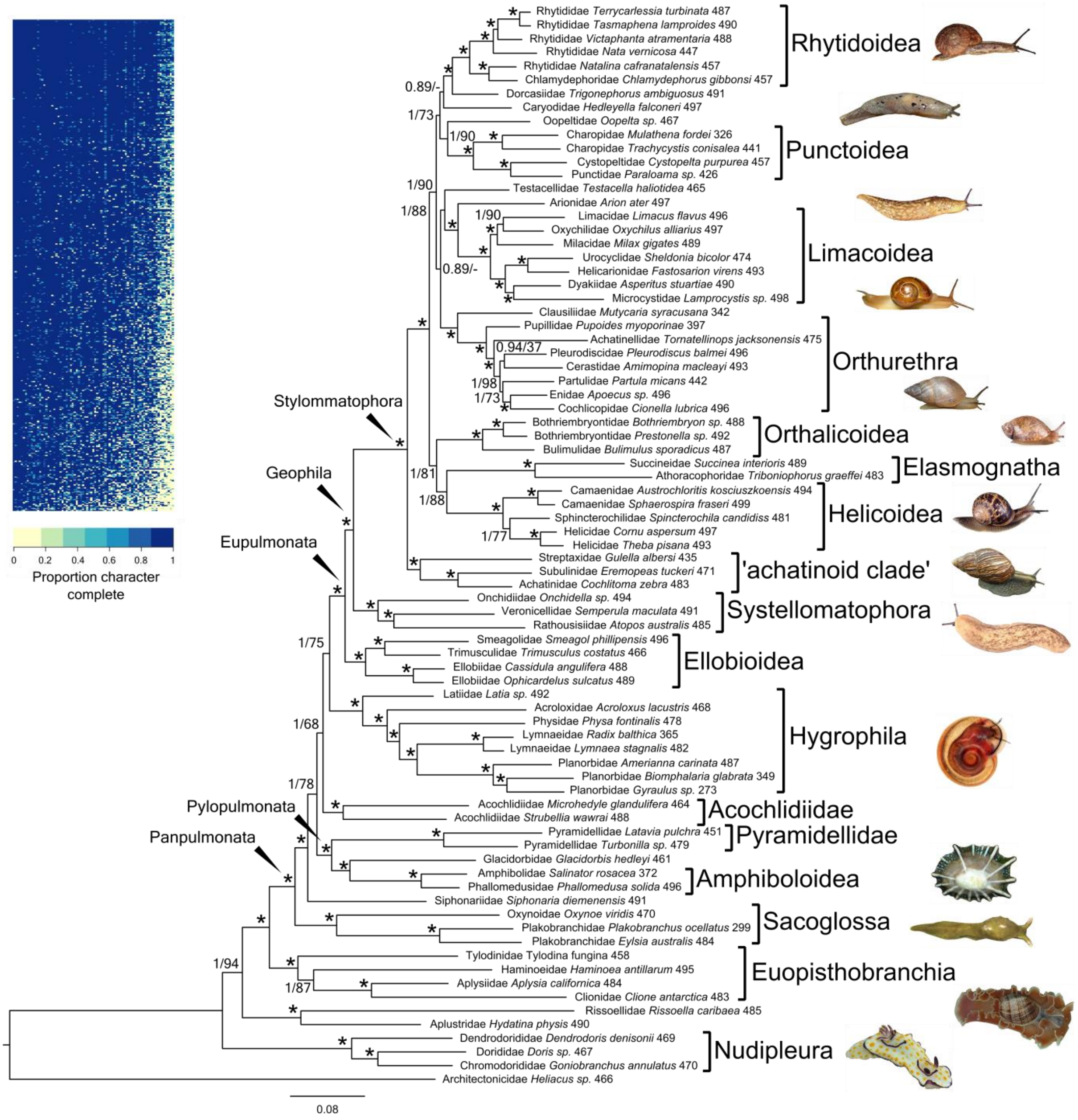
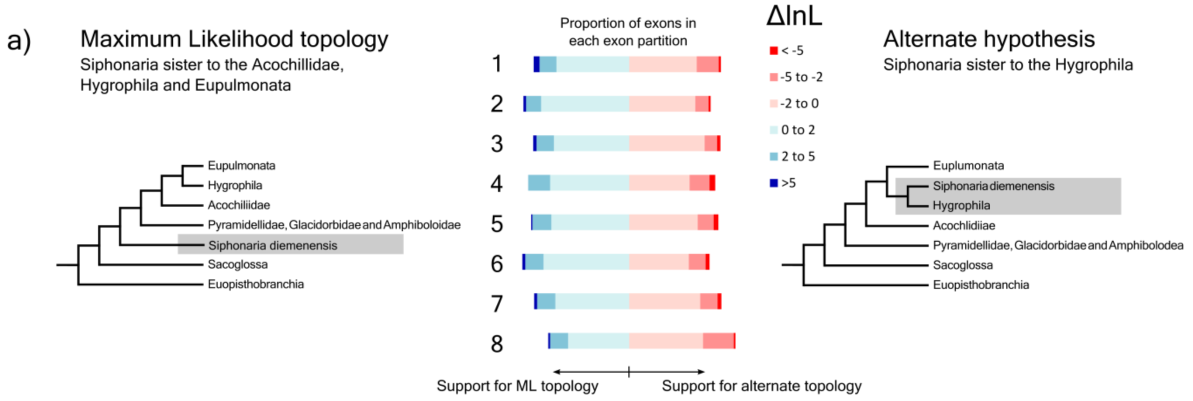

0
0.2
0.4
0.6
0.8
1
Proportion character complete
Rhytididae *Terrycarlessia turbinata* 487
Rhytididae *Tasmaphena lamproides* 490
Rhytididae *Victaphanta atramentaria* 488
Rhytididae *Nata vernicosa* 447
Rhytididae *Natalina cafranatalensis* 457
Chlamydephoridae *Chlamydephorus gibbonsi* 457
Dorcasiidae *Trigonephorus ambiguosus* 491
Caryodidae *Hedleyella falconeri* 497
Oopeltidae *Oopelta sp.* 467
Charopidae *Mulathena fordei* 326
Charopidae *Trachycystis conisalea* 441
Cystopeltidae *Cystopelta purpurea* 457
Punctidae *Paraloama sp.* 426
Testacellidae *Testacella haliotidea* 465
Arionidae *Arion ater* 497
Limacidae *Limacus flavus* 496
Oxychilidae *Oxychilus alliarius* 497
Milacidae *Milax gigates* 489
Urocyclidae *Sheldonia bicolor* 474
Helicarionidae *Fastosarion virens* 493
Dyakiidae *Asperitus stuartiae* 490
Microcystidae *Lamprocystis sp.* 498
Clausiliidae *Mutycaria syracusana* 342
Pupillidae *Pupoides myoporinae* 397
Achatinellidae *Tornatellinops jacksonensis* 475
Pleurodiscidae *Pleurodiscus balmei* 496
Cerastidae *Amimopina macleayi* 493
Partulidae *Partula micans* 442
Enidae *Apoecus sp.* 496
Cochlicopidae *Cionella lubrica* 496
Bothriembryontidae *Bothriembryon sp.* 488
Bothriembryontidae *Prestonella sp.* 492
Bulimulidae *Bulimulus sporadicus* 487
Succineidae *Succinea interioris* 489
Athoracophoridae *Triboniophorus graeffei* 483
Camaenidae *Austrochloritis kosciuszkoensis* 494
Camaenidae *Sphaerospira fraseri* 499
Sphincterochilidae *Spincterochila candidiss* 481
Helicidae *Cornu aspersum* 497
Helicidae *Theba pisana* 493
Streptaxidae *Gulella albersi* 435
Subulinidae *Eremopeas tuckeri* 471
Achatinidae *Cochlitoma zebra* 483
Onchidiidae *Onchidella sp.* 494
Veronicellidae *Semperula maculata* 491
Rathousisiidae *Atopos australis* 485
Smeagolidae *Smeagol phillipensis* 496
Trimusculidae *Trimusculus costatus* 466
Ellobiidae *Cassidula angulifera* 488
Ellobiidae *Ophicardelus sulcatus* 489
Latiidae *Latia sp.* 492
Acroloxidae *Acroloxus lacustris* 468
Physidae *Physa fontinalis* 478
Lymnaeidae *Radix balthica* 365
Lymnaeidae *Lymnaea stagnalis* 482
Planorbidae *Amerianna carinata* 487
Planorbidae *Biomphalaria glabrata* 349
Planorbidae *Gyraulus sp.* 273
Acochlidiidae *Microhedyle glandulifera* 464
Acochlidiidae *Strubellia wawrai* 488
Pyramidellidae *Latavia pulchra* 451
Pyramidellidae *Turbonilla sp.* 479
Glacidorbidae *Glacidorbis hedleyi* 461
Amphibolidae *Salinator rosacea* 372
Phallomedusidae *Phallomedusa solida* 496
Siphonariidae *Siphonaria diemenensis* 491
Oxynoidae *Oxynoe viridis* 470
Plakobranchidae *Plakobranchus ocellatus* 299
Plakobranchidae *Eylsia australis* 484
Tylodinidae Tylodina fungina 458
Haminoeidae *Haminoea antillarum* 495
Aplysiidae *Aplysia californica* 484
Clionidae *Clione antarctica* 483
Rissoellidae *Rissoella caribaea* 485
Aplustridae *Hydatina physis* 490
Dendrodorididae *Dendrodoris denisonii* 469
Dorididae *Doris sp.* 467
Chromodorididae *Goniobranchus annulatus* 470
Architectonicidae *Heliacus sp.* 466
Rhytidoidea
Punctoidea
Limacoidea
Orthurethra
Orthalicoidea
Elasmognatha
Helicoidea
'achatinoid clade'
Systellomatophora
Ellobioidea
Hygrophila
Acochlidiidae
Pyramidellidae
Amphiboloidea
Sacoglossa
Euopisthobranchia
Nudipleura
Stylommatophora
Geophila
Eupulmonata
Pylopulmonata
Panpulmonata
0.89/-
1/73
1/90
1/90
1/88
1/90
0.89/-
0.94/37
1/98
1/73
1/81
1/88
1/77
1/75
1/68
1/78
1/94
1/87
*
0.08

Figure 3.2. Bayesian phylogeny of Panpulmonata estimated using eight exon partitions. The node labels represent the posterior probabilities and bootstrap support from the maximum likelihood analysis. The asterisks represent nodes with 100 percent bootstrap support and Bayesian posterior probabilities of 1. The number after the species names represents the number of the 500 genes present for each taxon. The heat map shows the completeness of the dataset, sorted top to bottom from most to least complete gene (character complete) and left to right from most to least complete taxon. The colour key refers to the proportion of sequence present per taxon per gene.



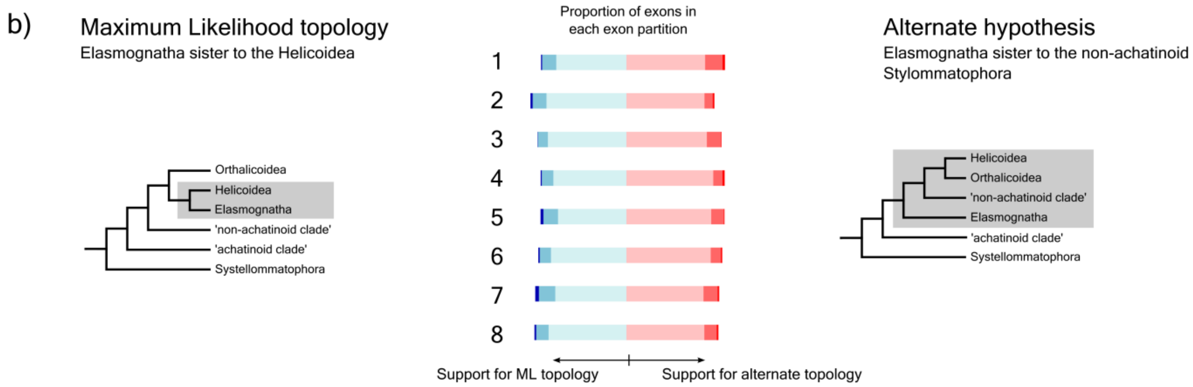


Figure 3.3. Partitioned likelihood support analyses comparing two alternate topological hypotheses amongst the major lineages within Panpulmonata. The first alternative hypothesis (a) regards the placement of the Siphonariidae. The second alternative topology (b) regards the placement of the Elasmognatha. The stacked bar charts show the proportion of exons in each of the eight exon clusters (labelled with numbers), which are in support of each hypothesis. The exons within the clusters are categorised by the summed differences in per site likelihoods (ΔlnL). Approximately Unbiased (AU) tests revealed that for all eight exon partitions could not reject the alternative hypothesis regarding the placement of the Siphonariidae (a) and only one partition (partition 2) could reject the alternative placement of the Elasmognatha.

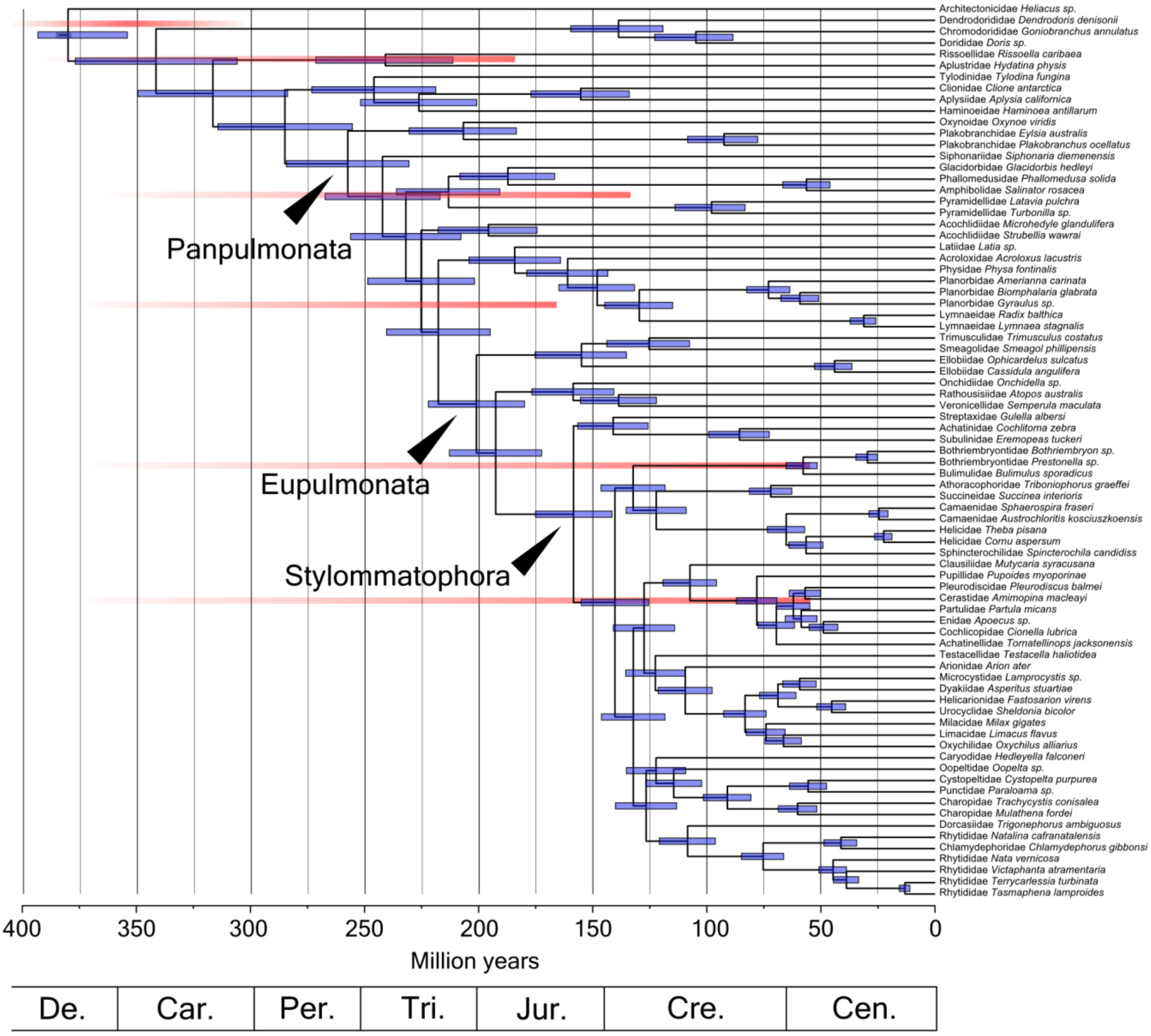


Figure 3.4. Chronogram for Panpulmonata inferred with MCMCTREE using a relaxed uncorrelated lognormal molecular clock model. The blue bars correspond to the 95% credibility intervals and the red bars represent the six calibrations used in the analysis (Table 3.2).

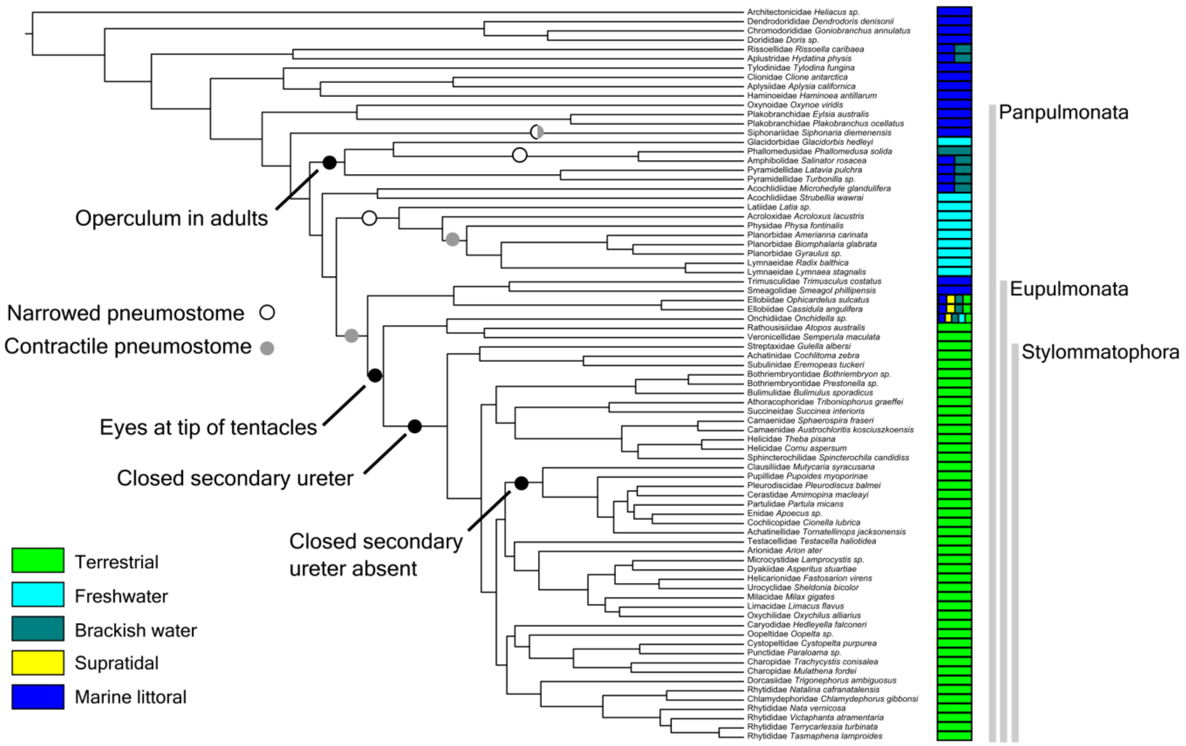

Architectonicidae *Heliacus sp.*
Dendrodorididae *Dendrodoris denisonii*
Chromodorididae *Goniobranchus annulatus*
Dorididae *Doris sp.*
Rissoellidae *Rissoella caribaea*
Aplustridae *Hydatina physis*
Tylodinidae *Tylodina fungina*
Clionidae *Clione antarctica*
Aplysiidae *Aplysia californica*
Haminoeidae *Haminoea antillarum*
Oxynoidae *Oxynoe viridis*
Plakobranchidae *Eylsia australis*
Plakobranchidae *Plakobranchus ocellatus*
Siphonariidae *Siphonaria diemenensis*
Glacidorbidae *Glacidorbis hedleyi*
Phallomedusidae *Phallomedusa solida*
Amphibolidae *Salinator rosacea*
Pyramidellidae *Latavia pulchra*
Pyramidellidae *Turbonilla sp.*
Acochlidiidae *Microhedyle glandulifera*
Acochlidiidae *Strubellia wawrai*
Latiidae *Latia sp.*
Acroloxidae *Acroloxus lacustris*
Physidae *Physa fontinalis*
Planorbidae *Amerianna carinata*
Planorbidae *Biomphalaria glabrata*
Planorbidae *Gyraulus sp.*
Lymnaeidae *Radix balthica*
Lymnaeidae *Lymnaea stagnalis*
Trimusculidae *Trimusculus costatus*
Smeagolidae *Smeagol phillipensis*
Ellobiidae *Ophicardelus sulcatus*
Ellobiidae *Cassidula angulifera*
Onchidiidae *Onchidella sp.*
Rathousisiidae *Atopos australis*
Veronicellidae *Semperula maculata*
Streptaxidae *Gulella albersi*
Achatinidae *Cochlitoma zebra*
Subulinidae *Eremopeas tuckeri*
Bothriembryontidae *Bothriembryon sp.*
Bothriembryontidae *Prestonella sp.*
Bulimulidae *Bulimulus sporadicus*
Athoracophoridae *Triboniophorus graeffei*
Succineidae *Succinea interioris*
Camaenidae *Sphaerospira fraseri*
Camaenidae *Austrochloritis kosciuszkoensis*
Helicidae *Theba pisana*
Helicidae *Cornu aspersum*
Sphincterochilidae *Spincterochila candidiss*
Clausiliidae *Mutycaria syracusana*
Pupillidae *Pupoides myoporinae*
Pleurodiscidae *Pleurodiscus balmei*
Cerastidae *Amimopina macleayi*
Partulidae *Partula micans*
Enidae *Apoecus sp.*
Cochlicopidae *Cionella lubrica*
Achatinellidae *Tornatellinops jacksonensis*
Testacellidae *Testacella haliotidea*
Arionidae *Arion ater*
Microcystidae *Lamprocystis sp.*
Dyakiidae *Asperitus stuartiae*
Helicarionidae *Fastosarion virens*
Urocyclidae *Sheldonia bicolor*
Milacidae *Milax gigates*
Limacidae *Limacus flavus*
Oxychilidae *Oxychilus alliarius*
Caryodidae *Hedleyella falconeri*
Oopeltidae *Oopelta sp.*
Cystopeltidae *Cystopelta purpurea*
Punctidae *Paraloama sp.*
Charopidae *Trachycystis conisalea*
Charopidae *Mulathena fordei*
Dorcasiidae *Trigonephorus ambiguosus*
Rhytididae *Natalina cafranatalensis*
Chlamydephoridae *Chlamydephorus gibbonsi*
Rhytididae *Nata vernicosa*
Rhytididae *Victaphanta atramentaria*
Rhytididae *Terrycarlessia turbinata*
Rhytididae *Tasmaphena lamproides*
Panpulmonata
Eupulmonata
Stylommatophora
Operculum in adults
Narrowed pneumostome ○
Contractile pneumostome ●
Eyes at tip of tentacles
Closed secondary ureter
Closed secondary ureter absent
🟩 Terrestrial
🟦 Freshwater
🟦 Brackish water
🟨 Supratidal
🟦 Marine littoral

Figure 3.5. Major morphological transitions within Panpulmonata mapped onto the chronogram resulting from the MCMCTREE analysis (Figure 3.4). The coloured bars to the right represent the current habitat usage at the family level. The operculum is plesiomorphic in Panpulmonata (i.e. it is the ancestral state), and the Glacidorbiidae, the Pyramellidae, and the Amphiboloidea are the only Panpulmonates to retain an operculum as adults.

Table 3.1. Species included in the study, including new and publicly available data, and sequencing, transcriptome assembly, and BLAST statistics.

Superfamily	**Family**	**Species**	**Source**	**Raw reads (pairs)**	**No. Trinity contigs**	**No. Blastx hits (e-10, *Lottia gigantea*)**	**No. of the 500 genes present**
Acavoidea	Caryodidae	*Hedleyella falconeri*	Sequenced herein	50136663	194605	24520	498
Acavoidea	Dorcasiidae	*Trigonephorus ambiguous*	Sequenced herein	12172881	115917	17475	493
Achatinelloidea	Achatinellidae	*Tornatellinops jacksonensis*	Sequenced herein	19685501	114186	24906	491
Achatinoidea	Achatinidae	*Cochlitoma zebra*	Sequenced herein	19890007	118129	16614	490
Achatinoidea	Subulinidae	*Eremopeas tuckeri*	Sequenced herein	29716463	130216	16407	476
Acochlidioidea	Parhedylidae	*Microhedyle glandulifera*	Zapata et al. 2014	6194970	261456	44575	464
Acochlidioidea	Acochlidiidae	*Strubellia wawrai*	Zapata et al. 2014	24132673	145096	33134	488
Acroloxoidea	Acroloxidae	*Acroloxus lacustris*	Sequenced herein	16964285	115029	24560	485
Acteonoidea	Aplustridae	*Hydatina physis*	Sequenced herein	18396478	102132	20433	491
Amphiboloidea	Amphibolidae	*Salinator rosacea*	Sequenced herein	28910244	85256	10485	377
Amphiboloidea	Phallomedusidae	*Phallomedusa solida*	Zapata et al. 2014	25685273	160685	31717	496
Aplysioidea	Aplysiidae	*Aplysia californica*	Broad Institute	-	26044	19647	484
Architectonicoidea	Architectonicidae	*Heliacus* sp.	Sequenced herein	-	105939	18162	466
Arionoidea	Oopeltidae	*Oopelta* sp.	Sequenced herein	16193557	86999	12472	467
Athoracophoroidea	Athoracophoridae	*Triboniophorus graeffei*	Sequenced herein	28542628	120022	20747	486
Chilinoidea	Latiidae	*Latia* sp.	Sequenced herein	19185851	105396	22926	493
Clausilioidea	Clausiliidae	*Muticaria syracusana*	Sequenced herein	-	88032	8629	342
Clionoidea	Clionidae	*Clione antarctica*	Zapata et al. 2014	20282761	100276	22831	483
Cochlicopoidea	Cochlicopidae	*Cionella lubrica*	Teasdale et al. 2016	8074560	111396	21675	497
Doridoidea	Chromodorididae	*Goniobranchus annulatus*	Sequenced herein	15983997	152752	20591	472
Doridoidea	Dorididae	*Doris kerguelenensis*	Zapata et al. 2014	14785164	194747	32526	467
Dyakioidea	Dyakiidae	*Asperitas cf stuartiae*	Sequenced herein	9322853	104942	15491	491
Ellobioidea	Ellobiidae	*Cassidula angulifera*	Teasdale et al. 2016	14281906	105803	16981	489
Ellobioidea	Ellobiidae	*Ophicardelus sulcatus*	Zapata et al. 2014	16026272	189708	30712	489
Enoidea	Cerastidae	*Amimopina macleayi*	Teasdale et al. 2016	7874195	93250	17258	494
Enoidea	Enidae	*Apoecus ramelauensis*	Teasdale et al. 2016	9362182	119711	21275	497

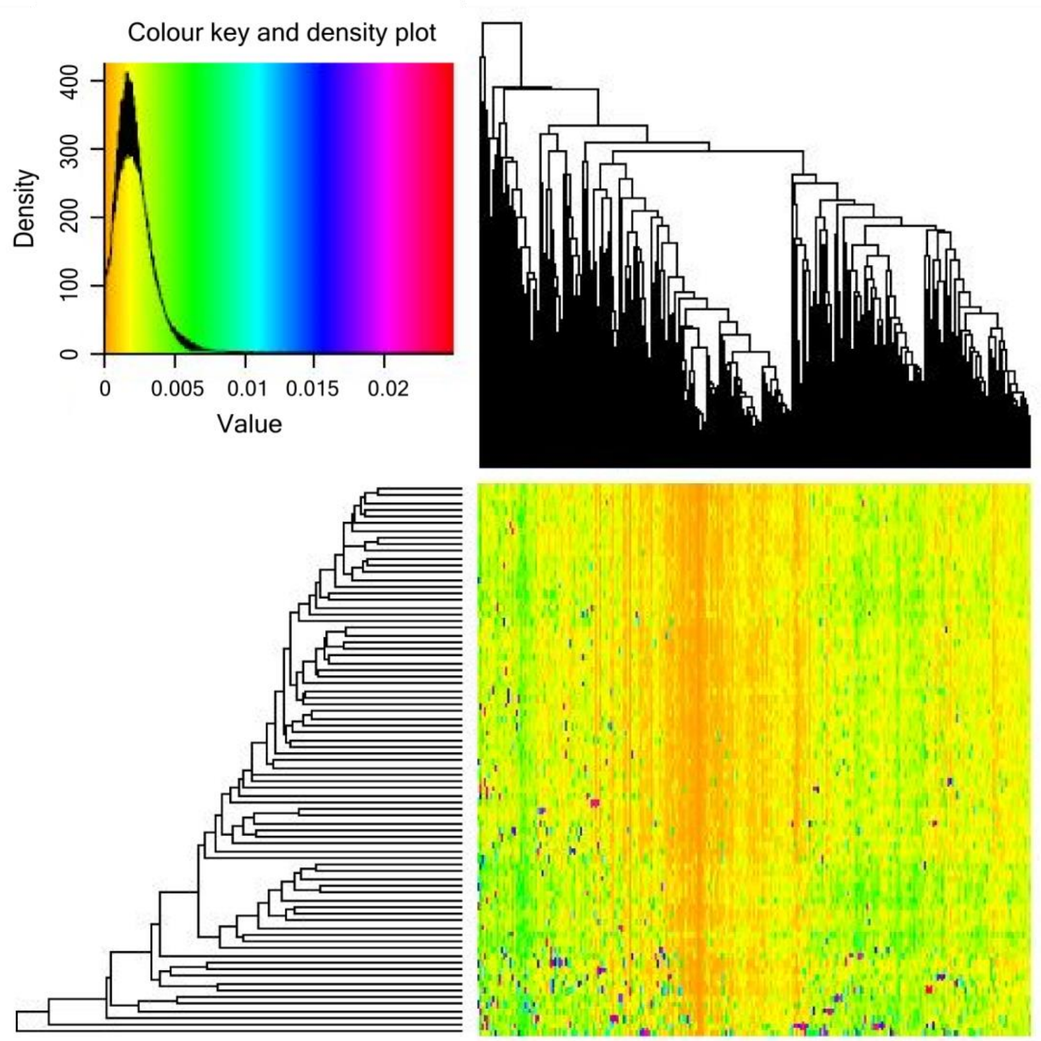
Superfamily	Family	Species	Source	Raw reads (pairs)	No. Trinity contigs	No. Blastx hits (e-10, *Lottia gigantea*)	No. of the 500 genes present
Euconulidae	Microcystidae	*Lamprocystis* sp.	Teasdale et al. 2016	22539699	128611	23797	499
Gastrodontoidea	Oxychilidae	*Oxychilus alliarius*	Teasdale et al. 2016	12925111	136044	21183	499
Glacidorboidea	Glacidorbidae	*Glacidorbis hedleyi*	Sequenced herein	18640140	114089	16641	464
Haminoeoidea	Haminoeidae	*Haminoea antillarum*	Zapata et al. 2014	11999771	186506	36926	495
Arionoidea	Arionidae	*Arion ater*	Sequenced herein	24232220	107782	23381	498
Helicarionoidea	Helicarionidae	*Fastosarion virens*	Teasdale et al. 2016	14904669	127454	18306	494
Helicarionoidea	Urocyclidae	*Sheldonia bicolor*	Sequenced herein	20406727	96641	12671	475
Helicoidea	Camaenidae	*Austrochloritis kosciuszkoensis*	Teasdale et al. 2016	11357080	107810	16238	495
Helicoidea	Camaenidae	*Sphaerospira fraseri*	Teasdale et al. 2016	31594841	179695	23910	500
Helicoidea	Helicidae	*Cornu aspersum*	Teasdale et al. 2016	21273910	160490	23114	498
Helicoidea	Helicidae	*Theba pisana*	Sequenced herein	-	157161	23543	493
Helicoidea	Sphincterochilidae	*Spincterochila candidissima*	Sequenced herein	-	132622	19013	481
Limacoidea	Limacidae	*Limacus flavus*	Teasdale et al. 2016	14907395	116088	19071	497
Lymnaeoidea	Lymnaeidae	*Lymnaea stagnalis*	Sadamoto et al. 2012	-	113250	18126	482
Lymnaeoidea	Lymnaeidae	*Radix balthica*	Feldmeyer et al. 2011	-	57986	11309	365
Onchidioidea	Onchidiidae	*Onchidella* sp.	Sequenced herein	38958754	125127	23995	495
Orthalicoidea	Bothriembryontidae	*Bothriembryon* sp .	Sequenced herein	15293953	108242	16663	489
Orthalicoidea	Bothriembryontidae	*Prestonella* sp.	Sequenced herein	17267990	136264	18337	493
Orthalicoidea	Bulimulidae	*Bulimulus sporadicus*	Sequenced herein	19198800	137216	16392	488
Otinoidea	Smeagolidae	*Smeagol phillipensis*	Teasdale et al. 2016	6393571	95429	23067	497
Oxynooidea	Oxynoidae	*Oxynoe viridis*	Sequenced herein	17130360	80018	13283	471
Parmacelloidea	Milacidae	*Milax gigates*	Teasdale et al. 2016	11263950	92337	16541	490
Partuloidea	Partulidae	*Partula micans*	Sequenced herein	12817661	102110	11905	443
Phyllidioidea	Dendrodorididae	*Dendrodoris kreusensternii*	Sequenced herein	17179394	99073	15445	473
Plakobranchoidea	Plakobranchidae	*Eylsia australis*	Sequenced herein	13998568	85855	16833	485
Plakobranchoidea	Plakobranchidae	*Plakobranchus ocellatus*	Wägele et al. 2010	-	77648	6114	299
Planorboidea	Physidae	*Physa fontinalis*	Sequenced herein	17463056	109148	31015	496
Planorboidea	Planorbidae	*Amerianna carinata*	Sequenced herein	22979776	105906	32478	494
Planorboidea	Planorbidae	*Biomphalaria glabrata*	Snaildb	-	43238	9034	349

Superfamily	**Family**	**Species**	**Source**	**Raw reads (pairs)**	**No. Trinity contigs**	**No. Blastx hits (e-10, *Lottia gigantea*)**	**No. of the 500 genes present**
Planorboidea	Planorbidae	*Gyraulus* sp.	Sequenced herein	6721952	17655	5969	276
Punctoidea	Charopidae	*Mulathena fordei*	Sequenced herein	29716463	81500	7114	328
Punctoidea	Charopidae	*Trachycystis conisalea*	Sequenced herein	20798108	109332	11670	441
Punctoidea	Cystopeltidae	*Cystopelta purpurea*	Sequenced herein	16926751	96589	15090	460
Punctoidea	Punctidae	*Paraloama* sp.	Sequenced herein	18103455	84235	11512	428
Pupilloidea	Pleurodiscidae	*Pleurodiscus balmei*	Sequenced herein	18789743	121688	20591	497
Pupilloidea	Pupillidae	*Pupoides myoporinae*	Sequenced herein	31709317	83814	10007	400
Pyramidelloidea	Pyramidellidae	*Latavia pulchra*	Sequenced herein	18296651	99737	15582	456
Pyramidelloidea	Pyramidellidae	*Turbonilla* sp.	Zapata et al. 2014	26619896	339517	49913	479
Rhytidoidea	Chlamydephoridae	*Chlamydephorus gibbonsi*	Sequenced herein	20029553	101413	12800	457
Rhytidoidea	Rhytididae	*Nata vernicosa*	Sequenced herein	15061888	91718	9969	447
Rhytidoidea	Rhytididae	*Natalina cafranatalensis*	Sequenced herein	17998595	100272	11298	458
Rhytidoidea	Rhytididae	*Tasmaphena lamproides*	Sequenced herein	20911998	124836	19313	490
Rhytidoidea	Rhytididae	*Terrycarlessia turbinata*	Teasdale et al. 2016	16985068	141421	17073	489
Rhytidoidea	Rhytididae	*Victaphanta atramenteria*	Teasdale et al. 2016	11312274	101127	16584	490
Rissoelloidea	Rissoellidae	*Rissoella caribaea*	Zapata et al. 2014	20666995	205074	47172	485
Siphonarioidea	Siphonariidae	*Siphonaria diemenensis*	Sequenced herein	24358578	114685	20656	493
Streptaxoidea	Streptaxidae	*Gulella albersi*	Sequenced herein	16428716	105625	10319	436
Succineoidea	Succineidae	*Succinea interioris*	Sequenced herein	33209269	95271	18854	491
Testacelloidea	Testacellidae	*Testacella haliotidea*	Sequenced herein	16099063	110915	16217	466
Trimusculoidea	Trimusculidae	*Trimusculus costatus*	Sequenced herein	12897216	95039	14791	468
Umbraculoidea	Tylodinidae	*Tylodina fungina*	Zapata et al. 2014	19608827	120587	24537	458
Veronicelloidea	Rathouisiidae	*Atopos australis*	Sequenced herein	15684330	104582	18727	487
Veronicelloidea	Veronicellidae	*Semperula maculata*	Teasdale et al. 2016	12461924	76847	21851	492

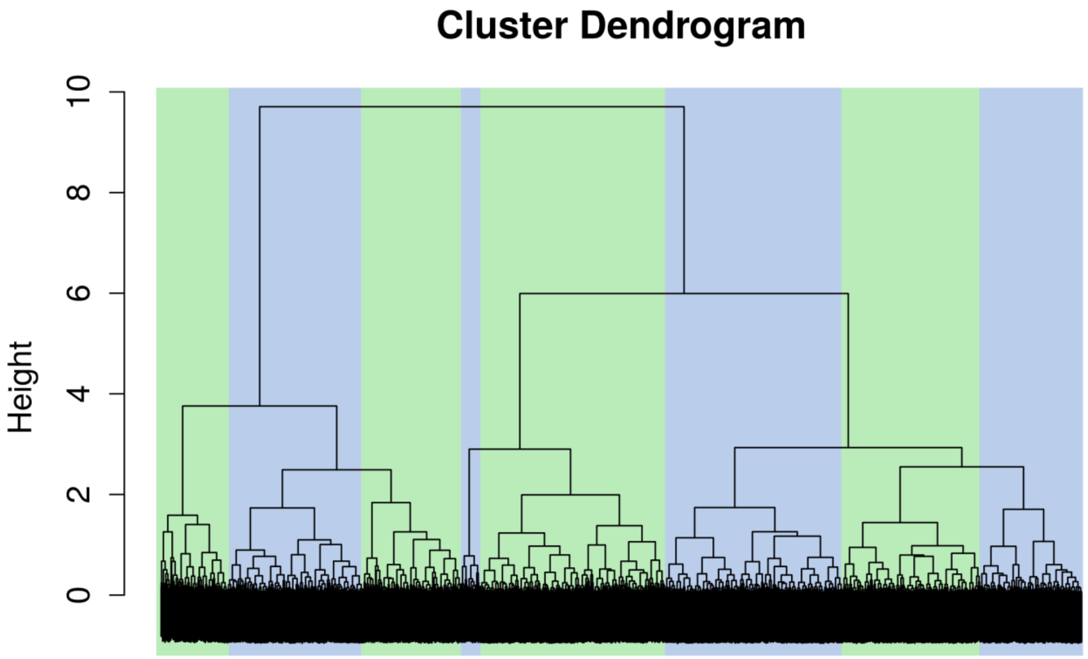
Table 3.2. Heterobranch fossils used for calibration in the MCMCTREE analysis. A sixth calibration point was obtained (see methods) from divergence estimates in Zapata et al. (2014).

Clade	**Type of calibration**	**Age (Ma)**	**Family**	**Species**	**Reference**
Bulimulus	Minimum	Middle Paleocene (57 – 59)	Bulimulidae	*Bulimulus fazendicus*	Salvador and Simone 2013
Siphonariidae	Minimum	Valanginian (132.9 – 139.8)	Siphonariidae	*Anisomyon* sp.	Kaim 2004
Lymnaeidae	Minimum	Bajocian (168.3 – 170.3)	Lymnaeidae	*Galba yunnanensis*	Pan 1977
Pupillidae	Minimum	Paleocene (66.0 – 56.0)	Pupillidae	*Albertanella minuta*	La Roque 1960
Acteonoidea	Minimum	Jamesoni (190.8 - 182.7)	Aplustridae	*Tornatellaea* cf. *fontis*	Rosenkrantz 1934

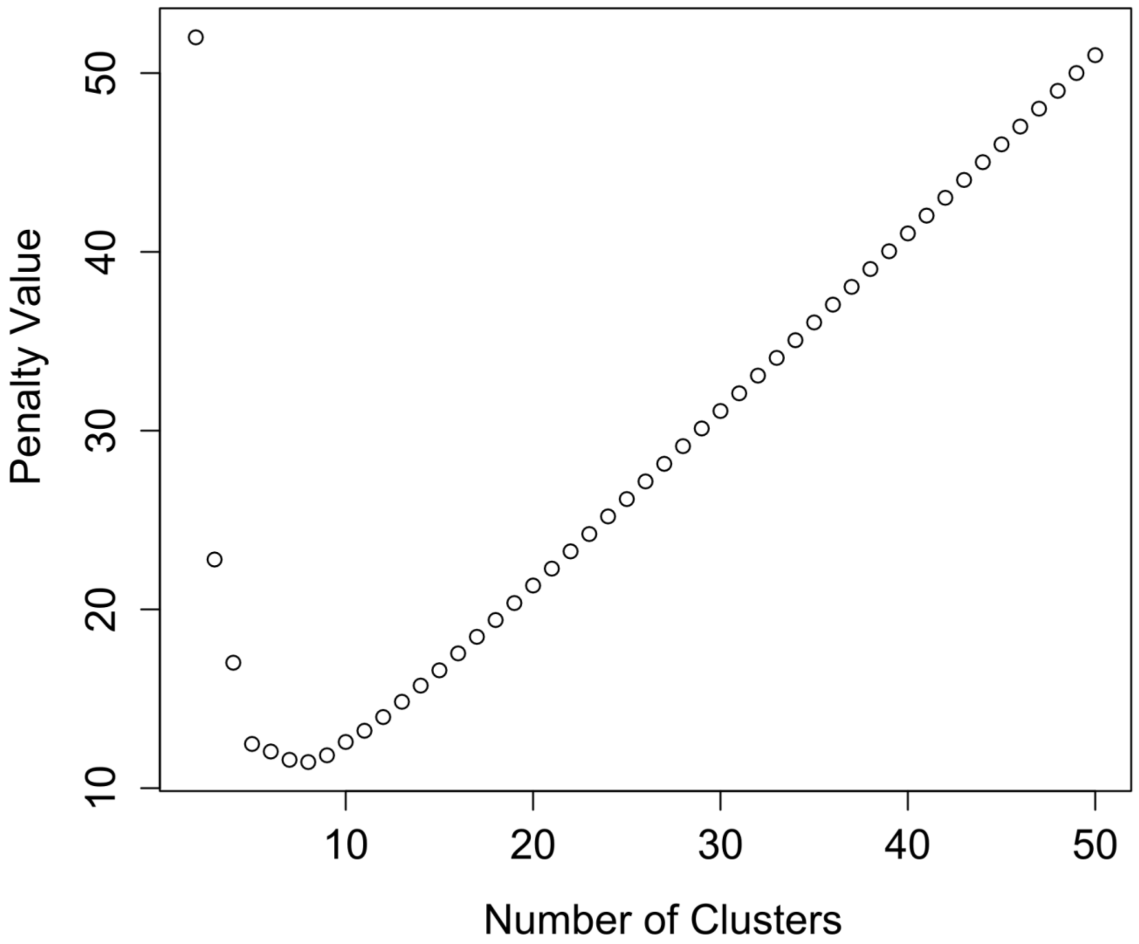
3.6 APPENDICES



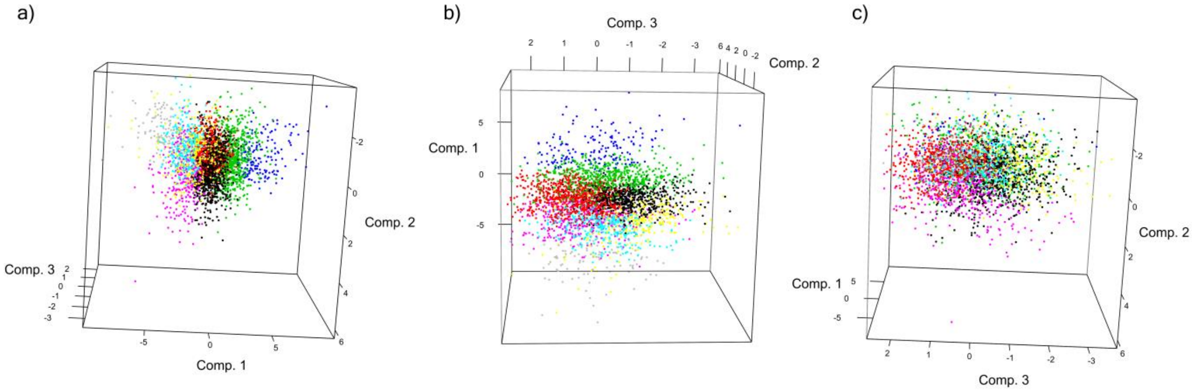
Appendix 3.1. Matrix showing the RCFV (Relative Composition Frequency Variability) value per taxon (top to bottom) per gene (left to right) for the exons in amino acid.



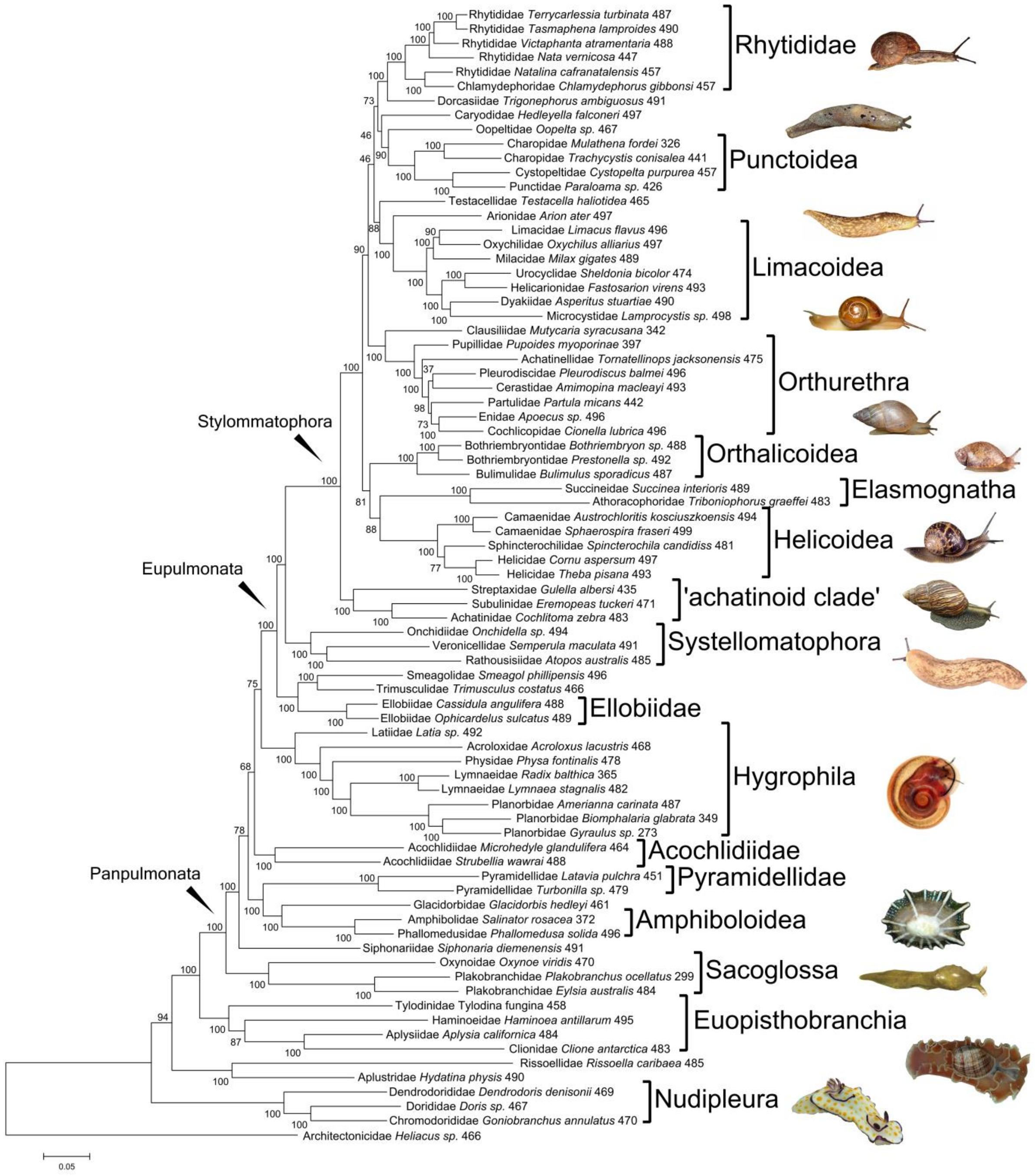
Appendix 3.2. Dendrogram resulting from Ward's hierarchical clustering based on the proportion of various types of amino acids within each exon. The coloured boxes depict the eight clusters used as data partitions for maximum likelihood phylogenetic analysis.



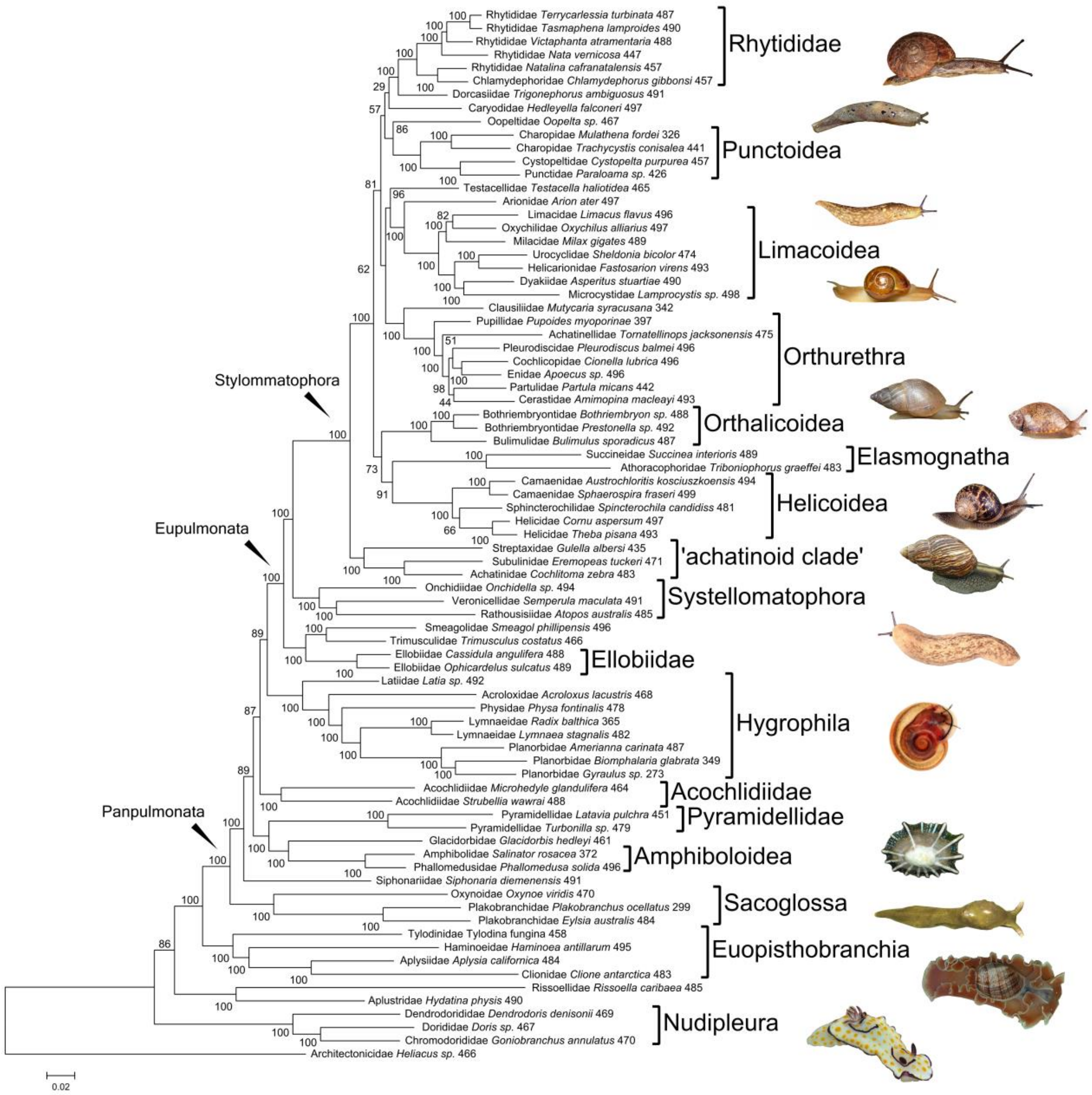
Appendix 3.3. Progress of the Kelley-Gardener-Sutcliffe penalty function during clustering of exons based on amino acid frequency. The minimum value of the penalty function is chosen as the cut-off point on the hierarchical tree (8 clusters). The minimum value represents the number of clusters where the dissimilarity between clusters was the highest and the dissimilarity within clusters was the lowest (Kelley et al., 1996).



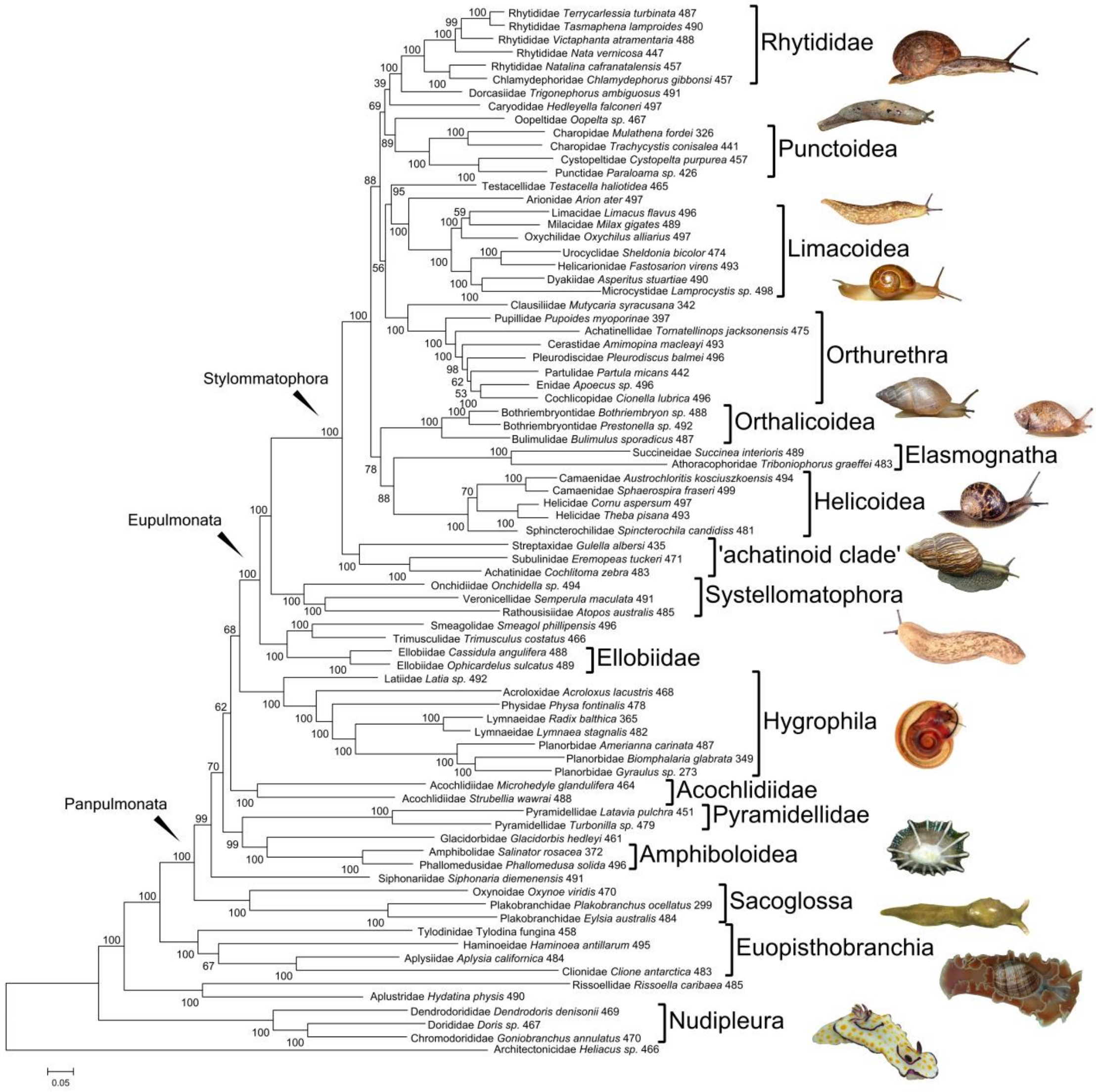
Appendix 3.4. Plots of the first three components of a PCA analysis for eight exon clusters based on the amino acid frequencies used to produce the clusters. Each data point represents one exon and each colour represents one of the eight exon clusters. The first three principle components explain 84% of the variation between the eight clusters. The majority of the variation between the eight exon clusters (58%) could be explained by the first principle component of the PCA analysis (where polarity and hydrophobicity have high contributions) and 95% could be explained by the first four principle components


Rhytididae Terrycarlessia turbinata 487
Rhytididae Tasmaphena lamproides 490
Rhytididae Victaphanta atramentaria 488
Rhytididae Nata vernicosa 447
Rhytididae Natalina cafranatalensis 457
Chlamydephoridae Chlamydephorus gibbonsi 457
Dorcasiidae Trigonephorus ambiguosus 491
Caryodidae Hedleyella falconeri 497
Oopeltidae Oopelta sp. 467
Charopidae Mulathena fordei 326
Charopidae Trachycystis conisalea 441
Cystopeltidae Cystopelta purpurea 457
Punctidae Paraloama sp. 426
Testacellidae Testacella haliotidea 465
Arionidae Arion ater 497
Limacidae Limacus flavus 496
Oxychilidae Oxychilus alliarius 497
Milacidae Milax gigates 489
Urocyclidae Sheldonia bicolor 474
Helicarionidae Fastosarion virens 493
Dyakiidae Asperitus stuartiae 490
Microcystidae Lamprocystis sp. 498
Clausiliidae Mutycaria syracusana 342
Pupillidae Pupoides myoporinae 397
Achatinellidae Tornatellinops jacksonensis 475
Pleurodiscidae Pleurodiscus balmei 496
Cerastidae Amimopina macleayi 493
Partulidae Partula micans 442
Enidae Apoecus sp. 496
Cochlicopidae Cionella lubrica 496
Bothriembryontidae Bothriembryon sp. 488
Bothriembryontidae Prestonella sp. 492
Bulimulidae Bulimulus sporadicus 487
Succineidae Succinea interioris 489
Athoracophoridae Triboniophorus graeffei 483
Camaenidae Austrochloritis kosciuszkoensis 494
Camaenidae Sphaerospira fraseri 499
Sphincterochilidae Spincterochila candidiss 481
Helicidae Cornu aspersum 497
Helicidae Theba pisana 493
Streptaxidae Gulella albersi 435
Subulinidae Eremopeas tuckeri 471
Achatinidae Cochlitoma zebra 483
Onchidiidae Onchidella sp. 494
Veronicellidae Semperula maculata 491
Rathousisiidae Atopos australis 485
Smeagolidae Smeagol phillipensis 496
Trimusculidae Trimusculus costatus 466
Ellobiidae Cassidula angulifera 488
Ellobiidae Ophicardelus sulcatus 489
Latiidae Latia sp. 492
Acroloxidae Acroloxus lacustris 468
Physidae Physa fontinalis 478
Lymnaeidae Radix balthica 365
Lymnaeidae Lymnaea stagnalis 482
Planorbidae Amerianna carinata 487
Planorbidae Biomphalaria glabrata 349
Planorbidae Gyraulus sp. 273
Acochlidiidae Microhedyle glandulifera 464
Acochlidiidae Strubellia wawrai 488
Pyramidellidae Latavia pulchra 451
Pyramidellidae Turbonilla sp. 479
Glacidorbidae Glacidorbis hedleyi 461
Amphibolidae Salinator rosacea 372
Phallomedusidae Phallomedusa solida 496
Siphonariidae Siphonaria diemenensis 491
Oxynoidae Oxynoe viridis 470
Plakobranchidae Plakobranchus ocellatus 299
Plakobranchidae Eylsia australis 484
Tylodinidae Tylodina fungina 458
Haminoeidae Haminoea antillarum 495
Aplysiidae Aplysia californica 484
Clionidae Clione antarctica 483
Rissoellidae Rissoella caribaea 485
Aplustridae Hydatina physis 490
Dendrodorididae Dendrodoris denisonii 469
Dorididae Doris sp. 467
Chromodorididae Goniobranchus annulatus 470
Architectonicidae Heliacus sp. 466
Rhytididae
Punctoidea
Limacoidea
Orthurethra
Orthalicoidea
Elasmognatha
Helicoidea
'achatinoid clade'
Systellomatophora
Ellobiidae
Hygrophila
Acochlidiidae
Pyramidellidae
Amphiboloidea
Sacoglossa
Euopisthobranchia
Nudipleura
Stylommatophora
Eupulmonata
Panpulmonata
0.05

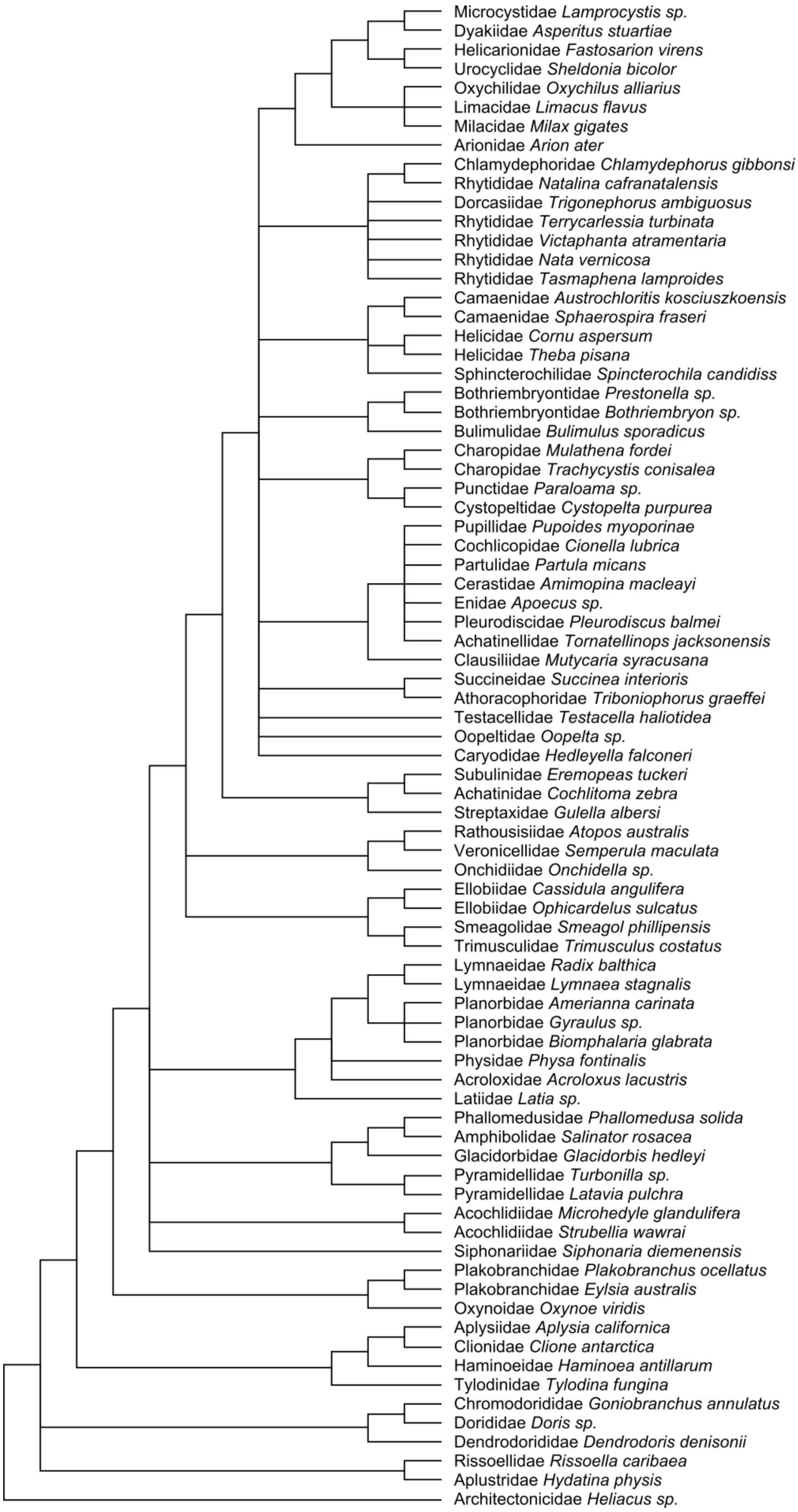
Appendix 3.5. Maximum likelihood phylogeny of Panpulmonata estimated using eight exon data partitions. The node labels represent the bootstrap support. The node labels summarise the boot strap support resulting from 255 thorough bootstraps.


Rhytididae Terrycarlessia turbinata 487
Rhytididae Tasmaphena lamproides 490
Rhytididae Victaphanta atramentaria 488
Rhytididae Nata vernicosa 447
Rhytididae Natalina cafranatalensis 457
Chlamydephoridae Chlamydephorus gibbonsi 457
Rhytididae
Dorcasiidae Trigonephorus ambiguosus 491
Caryodidae Hedleyella falconeri 497
Oopeltidae Oopelta sp. 467
Charopidae Mulathena fordei 326
Charopidae Trachycystis conisalea 441
Cystopeltidae Cystopelta purpurea 457
Punctidae Paraloama sp. 426
Punctoidea
Testacellidae Testacella haliotidea 465
Arionidae Arion ater 497
Limacidae Limacus flavus 496
Oxychilidae Oxychilus alliarius 497
Milacidae Milax gigates 489
Urocyclidae Sheldonia bicolor 474
Helicarionidae Fastosarion virens 493
Dyakiidae Asperitus stuartiae 490
Microcystidae Lamprocystis sp. 498
Limacoidea
Clausiliidae Mutycaria syracusana 342
Pupillidae Pupoides myoporinae 397
Achatinellidae Tornatellinops jacksonensis 475
Pleurodiscidae Pleurodiscus balmei 496
Cochlicopidae Cionella lubrica 496
Enidae Apoecus sp. 496
Partulidae Partula micans 442
Cerastidae Amimopina macleayi 493
Orthurethra
Stylommatophora
Bothriembryontidae Bothriembryon sp. 488
Bothriembryontidae Prestonella sp. 492
Bulimulidae Bulimulus sporadicus 487
Orthalicoidea
Succineidae Succinea interioris 489
Athoracophoridae Triboniophorus graeffei 483
Elasmognatha
Camaenidae Austrochloritis kosciuszkoensis 494
Camaenidae Sphaerospira fraseri 499
Sphincterochilidae Spincterochila candidiss 481
Helicidae Cornu aspersum 497
Helicidae Theba pisana 493
Helicoidea
Eupulmonata
Streptaxidae Gulella albersi 435
Subulinidae Eremopeas tuckeri 471
Achatinidae Cochlitoma zebra 483
'achatinoid clade'
Onchidiidae Onchidella sp. 494
Veronicellidae Semperula maculata 491
Rathousisiidae Atopos australis 485
Systellomatophora
Smeagolidae Smeagol phillipensis 496
Trimusculidae Trimusculus costatus 466
Ellobiidae Cassidula angulifera 488
Ellobiidae Ophicardelus sulcatus 489
Ellobiidae
Latiidae Latia sp. 492
Acroloxidae Acroloxus lacustris 468
Physidae Physa fontinalis 478
Lymnaeidae Radix balthica 365
Lymnaeidae Lymnaea stagnalis 482
Planorbidae Amerianna carinata 487
Planorbidae Biomphalaria glabrata 349
Planorbidae Gyraulus sp. 273
Hygrophila
Acochlidiidae Microhedyle glandulifera 464
Acochlidiidae Strubellia wawrai 488
Acochlidiidae
Panpulmonata
Pyramidellidae Latavia pulchra 451
Pyramidellidae Turbonilla sp. 479
Pyramidellidae
Glacidorbidae Glacidorbis hedleyi 461
Amphibolidae Salinator rosacea 372
Phallomedusidae Phallomedusa solida 496
Amphiboloidea
Siphonariidae Siphonaria diemenensis 491
Oxynoidae Oxynoe viridis 470
Plakobranchidae Plakobranchus ocellatus 299
Plakobranchidae Eylsia australis 484
Sacoglossa
Tylodinidae Tylodina fungina 458
Haminoeidae Haminoea antillarum 495
Aplysiidae Aplysia californica 484
Clionidae Clione antarctica 483
Euopisthobranchia
Rissoellidae Rissoella caribaea 485
Aplustridae Hydatina physis 490
Dendrodorididae Dendrodoris denisonii 469
Dorididae Doris sp. 467
Chromodorididae Goniobranchus annulatus 470
Nudipleura
Architectonicidae Heliacus sp. 466
0.02

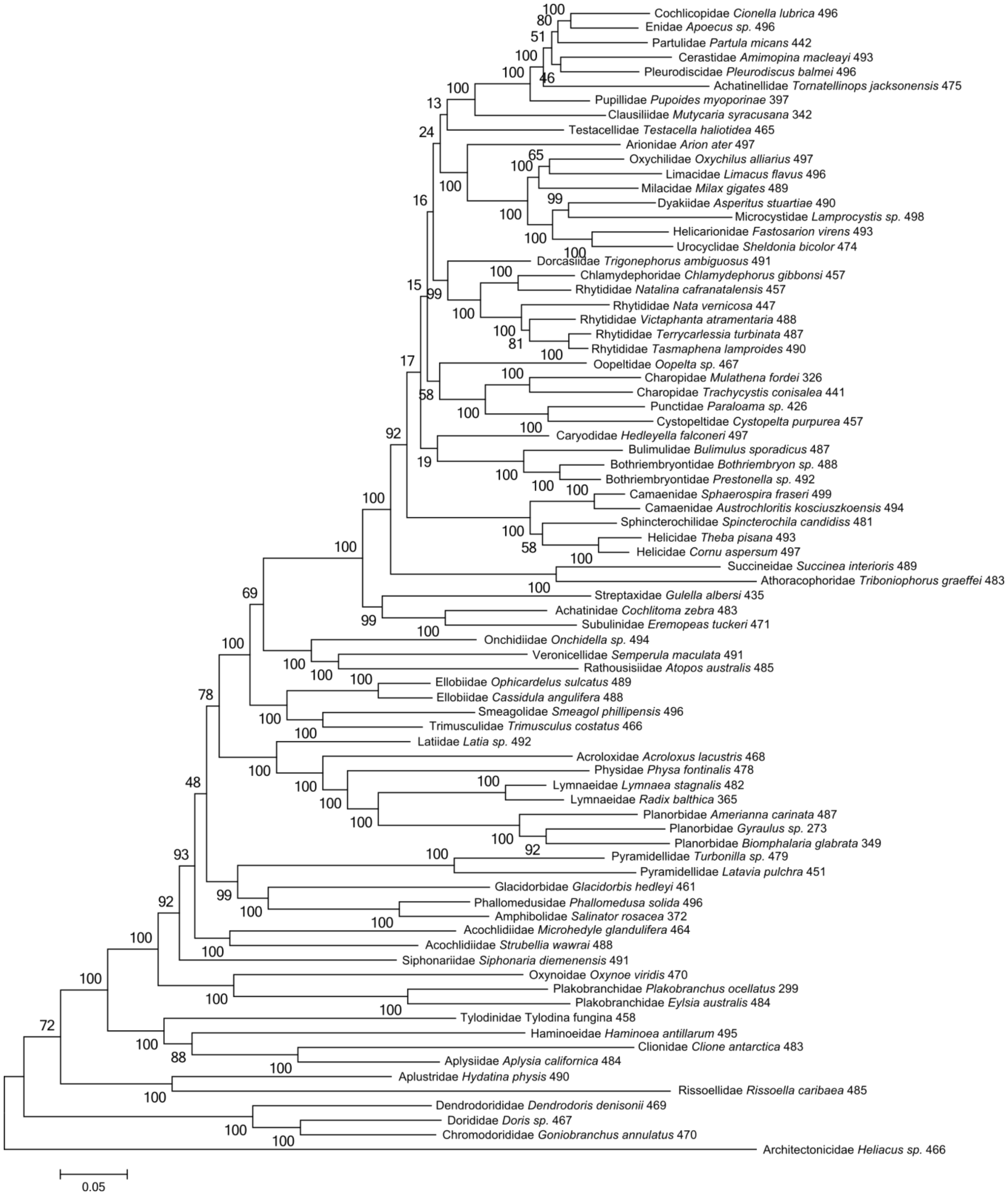
Appendix 3.6. Maximum likelihood tree constructed for the complete data set as analysed as a single partition but using the LG4X amino acid model which incorporates four separate substitution matrices to take into account heterogeneity. The node labels summarise the bootstrap support resulting from 255 thorough bootstraps.


Rhytididae *Terrycarlessia turbinata* 487
Rhytididae *Tasmaphena lamproides* 490
Rhytididae *Victaphanta atramentaria* 488
Rhytididae *Nata vernicosa* 447
Rhytididae *Natalina cafranatalensis* 457
Chlamydephoridae *Chlamydephorus gibbonsi* 457
Dorcasiidae *Trigonephorus ambiguosus* 491
Caryodidae *Hedleyella falconeri* 497
Oopeltidae *Oopelta sp.* 467
Charopidae *Mulathena fordei* 326
Charopidae *Trachycystis conisalea* 441
Cystopeltidae *Cystopelta purpurea* 457
Punctidae *Paraloama sp.* 426
Testacellidae *Testacella haliotidea* 465
Arionidae *Arion ater* 497
Limacidae *Limacus flavus* 496
Milacidae *Milax gigates* 489
Oxychilidae *Oxychilus alliarius* 497
Urocyclidae *Sheldonia bicolor* 474
Helicarionidae *Fastosarion virens* 493
Dyakiidae *Asperitus stuartiae* 490
Microcystidae *Lamprocystis sp.* 498
Clausiliidae *Mutycaria syracusana* 342
Pupillidae *Pupoides myoporinae* 397
Achatinellidae *Tornatellinops jacksonensis* 475
Cerastidae *Amimopina macleayi* 493
Pleurodiscidae *Pleurodiscus balmei* 496
Partulidae *Partula micans* 442
Enidae *Apoecus sp.* 496
Cochlicopidae *Cionella lubrica* 496
Bothriembryontidae *Bothriembryon sp.* 488
Bothriembryontidae *Prestonella sp.* 492
Bulimulidae *Bulimulus sporadicus* 487
Succineidae *Succinea interioris* 489
Athoracophoridae *Triboniophorus graeffei* 483
Camaenidae *Austrochloritis kosciuszkoensis* 494
Camaenidae *Sphaerospira fraseri* 499
Helicidae *Cornu aspersum* 497
Helicidae *Theba pisana* 493
Sphincterochilidae *Spincterochila candidiss* 481
Streptaxidae *Gulella albersi* 435
Subulinidae *Eremopeas tuckeri* 471
Achatinidae *Cochlitoma zebra* 483
Onchidiidae *Onchidella sp.* 494
Veronicellidae *Semperula maculata* 491
Rathousisiidae *Atopos australis* 485
Smeagolidae *Smeagol phillipensis* 496
Trimusculidae *Trimusculus costatus* 466
Ellobiidae *Cassidula angulifera* 488
Ellobiidae *Ophicardelus sulcatus* 489
Latiidae *Latia sp.* 492
Acroloxidae *Acroloxus lacustris* 468
Physidae *Physa fontinalis* 478
Lymnaeidae *Radix balthica* 365
Lymnaeidae *Lymnaea stagnalis* 482
Planorbidae *Amerianna carinata* 487
Planorbidae *Biomphalaria glabrata* 349
Planorbidae *Gyraulus sp.* 273
Acochlidiidae *Microhedyle glandulifera* 464
Acochlidiidae *Strubellia wawrai* 488
Pyramidellidae *Latavia pulchra* 451
Pyramidellidae *Turbonilla sp.* 479
Glacidorbidae *Glacidorbis hedleyi* 461
Amphibolidae *Salinator rosacea* 372
Phallomedusidae *Phallomedusa solida* 496
Siphonariidae *Siphonaria diemenensis* 491
Oxynoidae *Oxynoe viridis* 470
Plakobranchidae *Plakobranchus ocellatus* 299
Plakobranchidae *Eylsia australis* 484
Tylodinidae Tylodina fungina 458
Haminoeidae *Haminoea antillarum* 495
Aplysiidae *Aplysia californica* 484
Clionidae *Clione antarctica* 483
Rissoellidae *Rissoella caribaea* 485
Aplustridae *Hydatina physis* 490
Dendrodorididae *Dendrodoris denisonii* 469
Dorididae *Doris sp.* 467
Chromodorididae *Goniobranchus annulatus* 470
Architectonicidae *Heliacus sp.* 466
Rhytididae
Punctoidea
Limacoidea
Orthurethra
Orthalicoidea
Elasmognatha
Helicoidea
'achatinoid clade'
Systellomatophora
Ellobiidae
Hygrophila
Acochlidiidae
Pyramidellidae
Amphiboloidea
Sacoglossa
Euopisthobranchia
Nudipleura
Stylommatophora
Eupulmonata
Panpulmonata
0.05

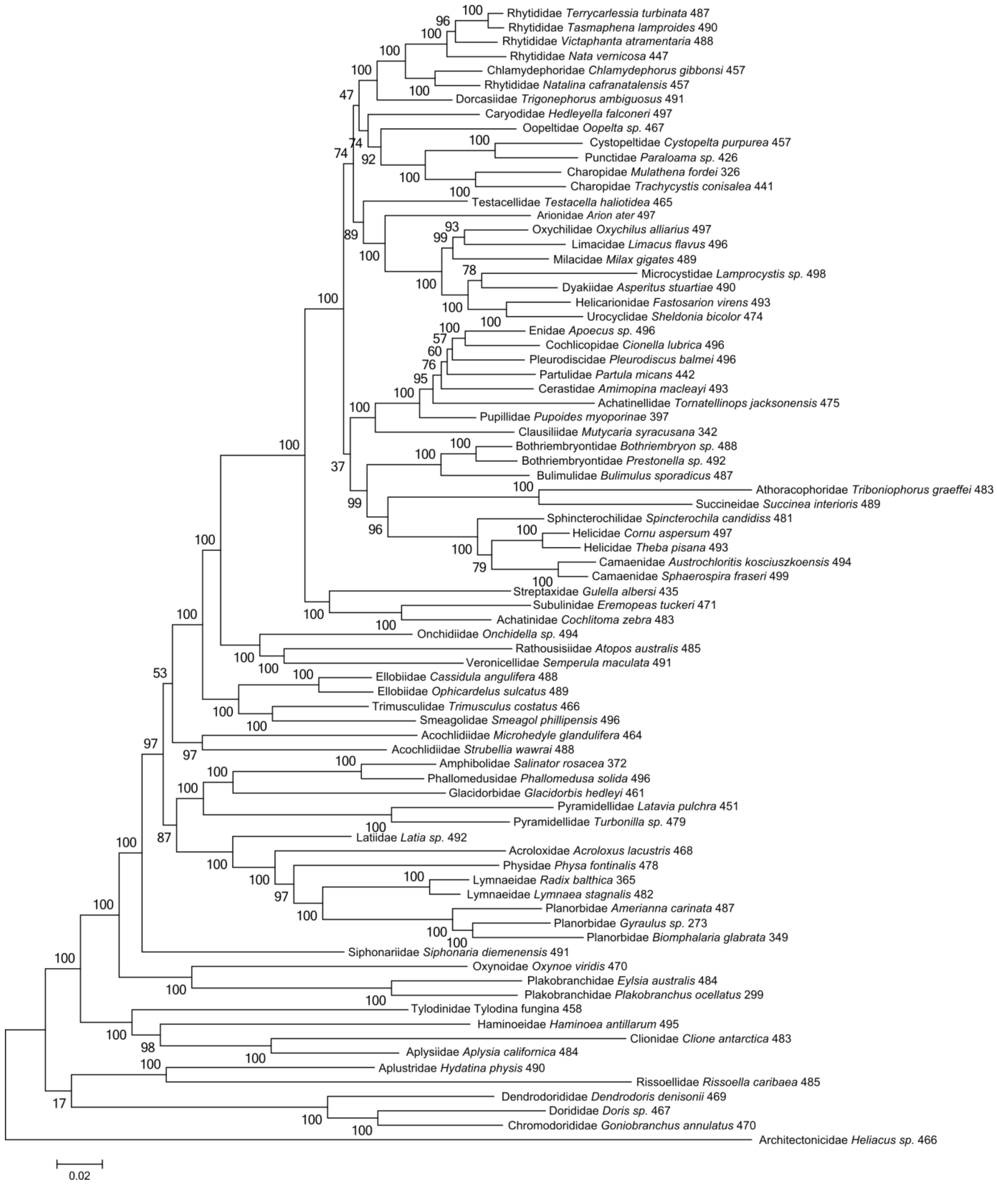
Appendix 3.7. Maximum likelihood tree constructed using a partitioning scheme of 82 clusters of exons. This partitioning scheme resulted from a Partition finder analysis which clustered 500 exon clusters resulting from Ward's hierarchical clustering. The node labels summarise the boot strap support resulting from 255 thorough bootstraps.


Microcystidae *Lamprocystis sp.*
Dyakiidae *Asperitus stuartiae*
Helicarionidae *Fastosarion virens*
Urocyclidae *Sheldonia bicolor*
Oxychilidae *Oxychilus alliarius*
Limacidae *Limacus flavus*
Milacidae *Milax gigates*
Arionidae *Arion ater*
Chlamydephoridae *Chlamydephorus gibbonsi*
Rhytididae *Natalina cafranatalensis*
Dorcasiidae *Trigonephorus ambiguosus*
Rhytididae *Terrycarlessia turbinata*
Rhytididae *Victaphanta atramentaria*
Rhytididae *Nata vernicosa*
Rhytididae *Tasmaphena lamproides*
Camaenidae *Austrochloritis kosciuszkoensis*
Camaenidae *Sphaerospira fraseri*
Helicidae *Cornu aspersum*
Helicidae *Theba pisana*
Sphincterochilidae *Spincterochila candidiss*
Bothriembryontidae *Prestonella sp.*
Bothriembryontidae *Bothriembryon sp.*
Bulimulidae *Bulimulus sporadicus*
Charopidae *Mulathena fordei*
Charopidae *Trachycystis conisalea*
Punctidae *Paraloama sp.*
Cystopeltidae *Cystopelta purpurea*
Pupillidae *Pupoides myoporinae*
Cochlicopidae *Cionella lubrica*
Partulidae *Partula micans*
Cerastidae *Amimopina macleayi*
Enidae *Apoecus sp.*
Pleurodiscidae *Pleurodiscus balmei*
Achatinellidae *Tornatellinops jacksonensis*
Clausiliidae *Mutycaria syracusana*
Succineidae *Succinea interioris*
Athoracophoridae *Triboniophorus graeffei*
Testacellidae *Testacella haliotidea*
Oopeltidae *Oopelta sp.*
Caryodidae *Hedleyella falconeri*
Subulinidae *Eremopeas tuckeri*
Achatinidae *Cochlitoma zebra*
Streptaxidae *Gulella albersi*
Rathousisiidae *Atopos australis*
Veronicellidae *Semperula maculata*
Onchidiidae *Onchidella sp.*
Ellobiidae *Cassidula angulifera*
Ellobiidae *Ophicardelus sulcatus*
Smeagolidae *Smeagol phillipensis*
Trimusculidae *Trimusculus costatus*
Lymnaeidae *Radix balthica*
Lymnaeidae *Lymnaea stagnalis*
Planorbidae *Amerianna carinata*
Planorbidae *Gyraulus sp.*
Planorbidae *Biomphalaria glabrata*
Physidae *Physa fontinalis*
Acroloxidae *Acroloxus lacustris*
Latiidae *Latia sp.*
Phallomedusidae *Phallomedusa solida*
Amphibolidae *Salinator rosacea*
Glacidorbidae *Glacidorbis hedleyi*
Pyramidellidae *Turbonilla sp.*
Pyramidellidae *Latavia pulchra*
Acochlidiidae *Microhedyle glandulifera*
Acochlidiidae *Strubellia wawrai*
Siphonariidae *Siphonaria diemenensis*
Plakobranchidae *Plakobranchus ocellatus*
Plakobranchidae *Eylsia australis*
Oxynoidae *Oxynoe viridis*
Aplysiidae *Aplysia californica*
Clionidae *Clione antarctica*
Haminoeidae *Haminoea antillarum*
Tylodinidae *Tylodina fungina*
Chromodorididae *Goniobranchus annulatus*
Dorididae *Doris sp.*
Dendrodorididae *Dendrodoris denisonii*
Rissoellidae *Rissoella caribaea*
Aplustridae *Hydatina physis*
Architectonicidae *Heliacus sp.*

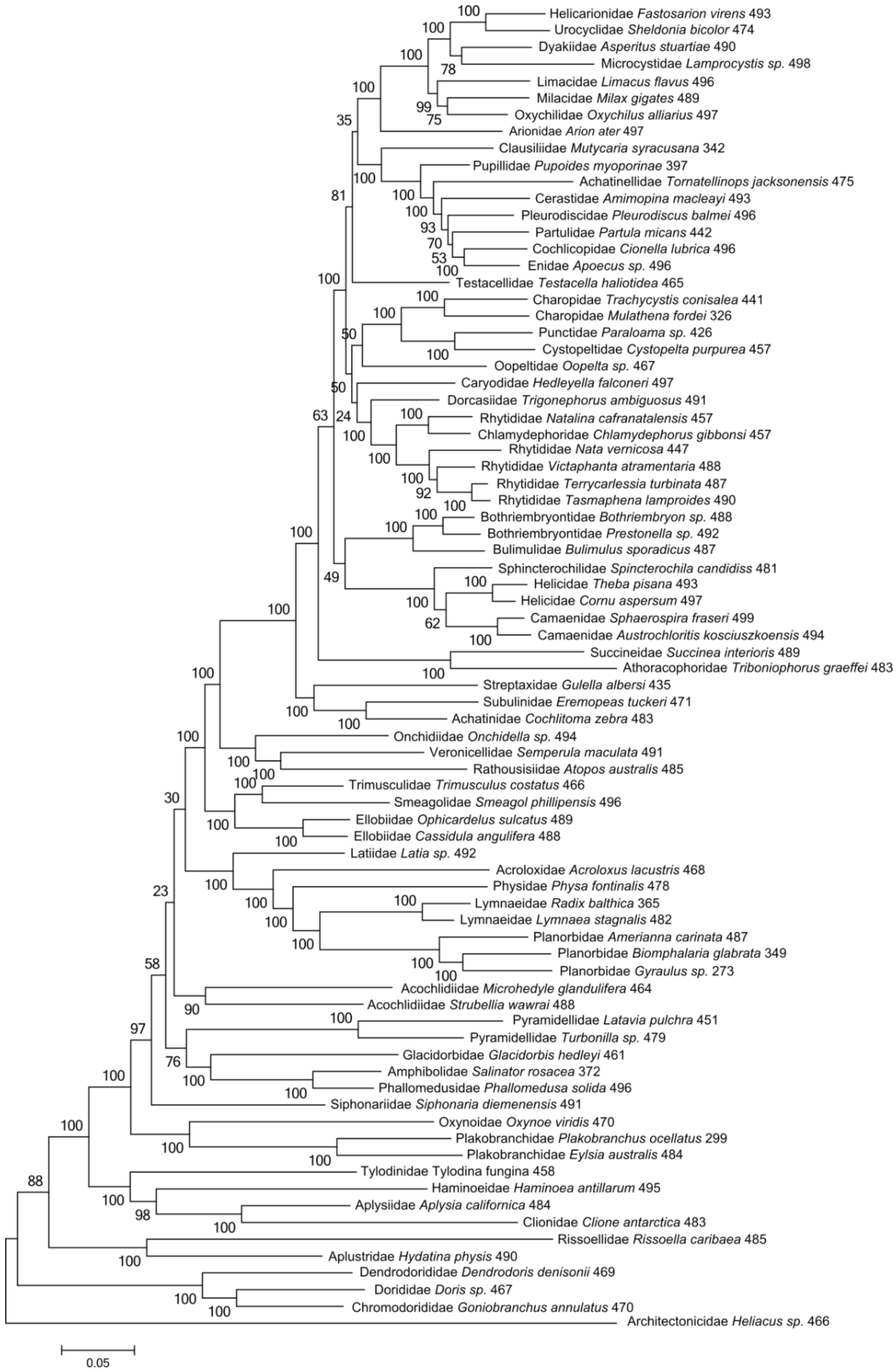
Appendix 3.8. Strict consensus of the maximum likelihood trees estimated for eight non-redundant clusters of exons for Panpulmonata.


100
80
51
100
100
46
100
13
24
Cochlicopidae Cionella lubrica 496
Enidae Apoecus sp. 496
Partulidae Partula micans 442
Cerastidae Amimopina macleayi 493
Pleurodiscidae Pleurodiscus balmei 496
Achatinellidae Tornatellinops jacksonensis 475
Pupillidae Pupoides myoporinae 397
Clausiliidae Mutycaria syracusana 342
Testacellidae Testacella haliotidea 465
Arionidae Arion ater 497
65
100
Oxychilidae Oxychilus alliarius 497
Limacidae Limacus flavus 496
100
Milacidae Milax gigates 489
16
99
Dyakiidae Asperitus stuartiae 490
100
Microcystidae Lamprocystis sp. 498
Helicarionidae Fastosarion virens 493
100
Urocyclidae Sheldonia bicolor 474
100
Dorcasiidae Trigonephorus ambiguosus 491
100
Chlamydephoridae Chlamydephorus gibbonsi 457
15
99
Rhytididae Natalina cafranatalensis 457
100
Rhytididae Nata vernicosa 447
Rhytididae Victaphanta atramentaria 488
100
Rhytididae Terrycarlessia turbinata 487
81
Rhytididae Tasmaphena lamproides 490
100
17
Oopeltidae Oopelta sp. 467
100
Charopidae Mulathena fordei 326
58
Charopidae Trachycystis conisalea 441
Punctidae Paraloama sp. 426
100
Cystopeltidae Cystopelta purpurea 457
92
100
Caryodidae Hedleyella falconeri 497
Bulimulidae Bulimulus sporadicus 487
19
Bothriembryontidae Bothriembryon sp. 488
100
Bothriembryontidae Prestonella sp. 492
100
100
100
Camaenidae Sphaerospira fraseri 499
Camaenidae Austrochloritis kosciuszkoensis 494
Sphincterochilidae Spincterochila candidiss 481
100
Helicidae Theba pisana 493
58
100
Helicidae Cornu aspersum 497
100
Succineidae Succinea interioris 489
Athoracophoridae Triboniophorus graeffei 483
100
69
Streptaxidae Gulella albersi 435
Achatinidae Cochlitoma zebra 483
99
Subulinidae Eremopeas tuckeri 471
100
Onchidiidae Onchidella sp. 494
100
Veronicellidae Semperula maculata 491
100
Rathousisiidae Atopos australis 485
100
100
Ellobiidae Ophicardelus sulcatus 489
78
Ellobiidae Cassidula angulifera 488
Smeagolidae Smeagol phillipensis 496
100
Trimusculidae Trimusculus costatus 466
100
Latiidae Latia sp. 492
Acroloxidae Acroloxus lacustris 468
100
Physidae Physa fontinalis 478
48
100
100
Lymnaeidae Lymnaea stagnalis 482
100
Lymnaeidae Radix balthica 365
100
Planorbidae Amerianna carinata 487
Planorbidae Gyraulus sp. 273
100
93
92
Planorbidae Biomphalaria glabrata 349
100
Pyramidellidae Turbonilla sp. 479
Pyramidellidae Latavia pulchra 451
99
Glacidorbidae Glacidorbis hedleyi 461
92
100
Phallomedusidae Phallomedusa solida 496
Amphibolidae Salinator rosacea 372
100
Acochlidiidae Microhedyle glandulifera 464
100
Acochlidiidae Strubellia wawrai 488
100
Siphonariidae Siphonaria diemenensis 491
Oxynoidae Oxynoe viridis 470
100
100
Plakobranchidae Plakobranchus ocellatus 299
Plakobranchidae Eylsia australis 484
100
72
Tylodinidae Tylodina fungina 458
Haminoeidae Haminoea antillarum 495
100
88
Clionidae Clione antarctica 483
100
Aplysiidae Aplysia californica 484
Aplustridae Hydatina physis 490
100
Rissoellidae Rissoella caribaea 485
Dendrodorididae Dendrodoris denisonii 469
100
Dorididae Doris sp. 467
100
Chromodorididae Goniobranchus annulatus 470
Architectonicidae Heliacus sp. 466
0.05

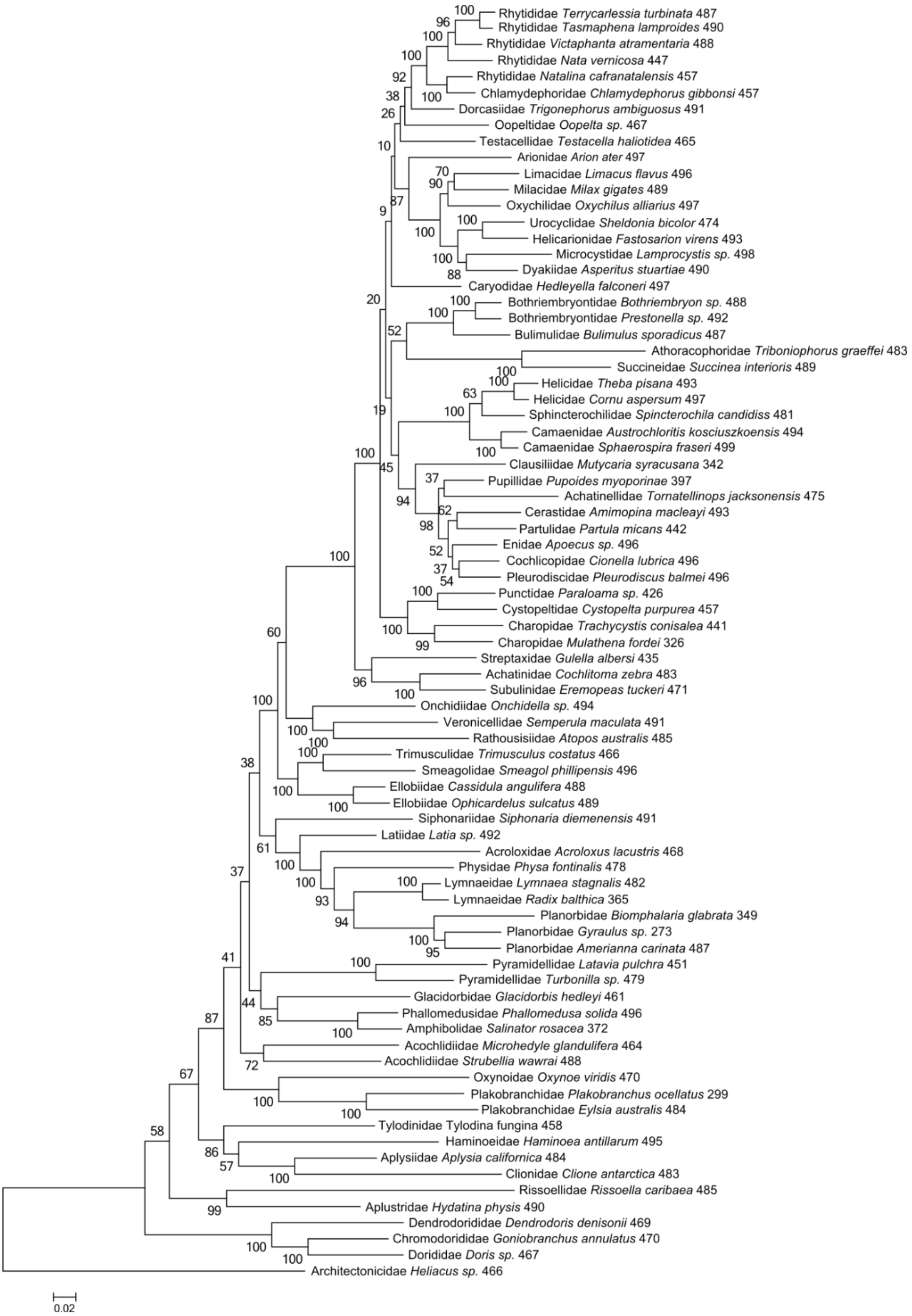
Appendix 3.9. Maximum likelihood tree constructed using the first of eight exon clusters produced by clustering the exons based on amino acid frequencies. This exon cluster consists of 32,384 amino acids. The node labels summarise the boot strap support resulting from 100 fast bootstraps in RAxML.


Rhytididae *Terrycarlessia turbinata* 487
Rhytididae *Tasmaphena lamproides* 490
Rhytididae *Victaphanta atramentaria* 488
Rhytididae *Nata vernicosa* 447
Chlamydephoridae *Chlamydephorus gibbonsi* 457
Rhytididae *Natalina cafranatalensis* 457
Dorcasiidae *Trigonephorus ambiguosus* 491
Caryodidae *Hedleyella falconeri* 497
Oopeltidae *Oopelta sp.* 467
Cystopeltidae *Cystopelta purpurea* 457
Punctidae *Paraloama sp.* 426
Charopidae *Mulathena fordei* 326
Charopidae *Trachycystis conisalea* 441
Testacellidae *Testacella haliotidea* 465
Arionidae *Arion ater* 497
Oxychilidae *Oxychilus alliarius* 497
Limacidae *Limacus flavus* 496
Milacidae *Milax gigates* 489
Microcystidae *Lamprocystis sp.* 498
Dyakiidae *Asperitus stuartiae* 490
Helicarionidae *Fastosarion virens* 493
Urocyclidae *Sheldonia bicolor* 474
Enidae *Apoecus sp.* 496
Cochlicopidae *Cionella lubrica* 496
Pleurodiscidae *Pleurodiscus balmei* 496
Partulidae *Partula micans* 442
Cerastidae *Amimopina macleayi* 493
Achatinellidae *Tornatellinops jacksonensis* 475
Pupillidae *Pupoides myoporinae* 397
Clausiliidae *Mutycaria syracusana* 342
Bothriembryontidae *Bothriembryon sp.* 488
Bothriembryontidae *Prestonella sp.* 492
Bulimulidae *Bulimulus sporadicus* 487
Athoracophoridae *Triboniophorus graeffei* 483
Succineidae *Succinea interioris* 489
Sphincterochilidae *Spincterochila candidiss* 481
Helicidae *Cornu aspersum* 497
Helicidae *Theba pisana* 493
Camaenidae *Austrochloritis kosciuszkoensis* 494
Camaenidae *Sphaerospira fraseri* 499
Streptaxidae *Gulella albersi* 435
Subulinidae *Eremopeas tuckeri* 471
Achatinidae *Cochlitoma zebra* 483
Onchidiidae *Onchidella sp.* 494
Rathousisiidae *Atopos australis* 485
Veronicellidae *Semperula maculata* 491
Ellobiidae *Cassidula angulifera* 488
Ellobiidae *Ophicardelus sulcatus* 489
Trimusculidae *Trimusculus costatus* 466
Smeagolidae *Smeagol phillipensis* 496
Acochlidiidae *Microhedyle glandulifera* 464
Acochlidiidae *Strubellia wawrai* 488
Amphibolidae *Salinator rosacea* 372
Phallomedusidae *Phallomedusa solida* 496
Glacidorbidae *Glacidorbis hedleyi* 461
Pyramidellidae *Latavia pulchra* 451
Pyramidellidae *Turbonilla sp.* 479
Latiidae *Latia sp.* 492
Acroloxidae *Acroloxus lacustris* 468
Physidae *Physa fontinalis* 478
Lymnaeidae *Radix balthica* 365
Lymnaeidae *Lymnaea stagnalis* 482
Planorbidae *Amerianna carinata* 487
Planorbidae *Gyraulus sp.* 273
Planorbidae *Biomphalaria glabrata* 349
Siphonariidae *Siphonaria diemenensis* 491
Oxynoidae *Oxynoe viridis* 470
Plakobranchidae *Eylsia australis* 484
Plakobranchidae *Plakobranchus ocellatus* 299
Tylodinidae Tylodina fungina 458
Haminoeidae *Haminoea antillarum* 495
Clionidae *Clione antarctica* 483
Aplysiidae *Aplysia californica* 484
Aplustridae *Hydatina physis* 490
Rissoellidae *Rissoella caribaea* 485
Dendrodorididae *Dendrodoris denisonii* 469
Dorididae *Doris sp.* 467
Chromodorididae *Goniobranchus annulatus* 470
Architectonicidae *Heliacus sp.* 466
0.02

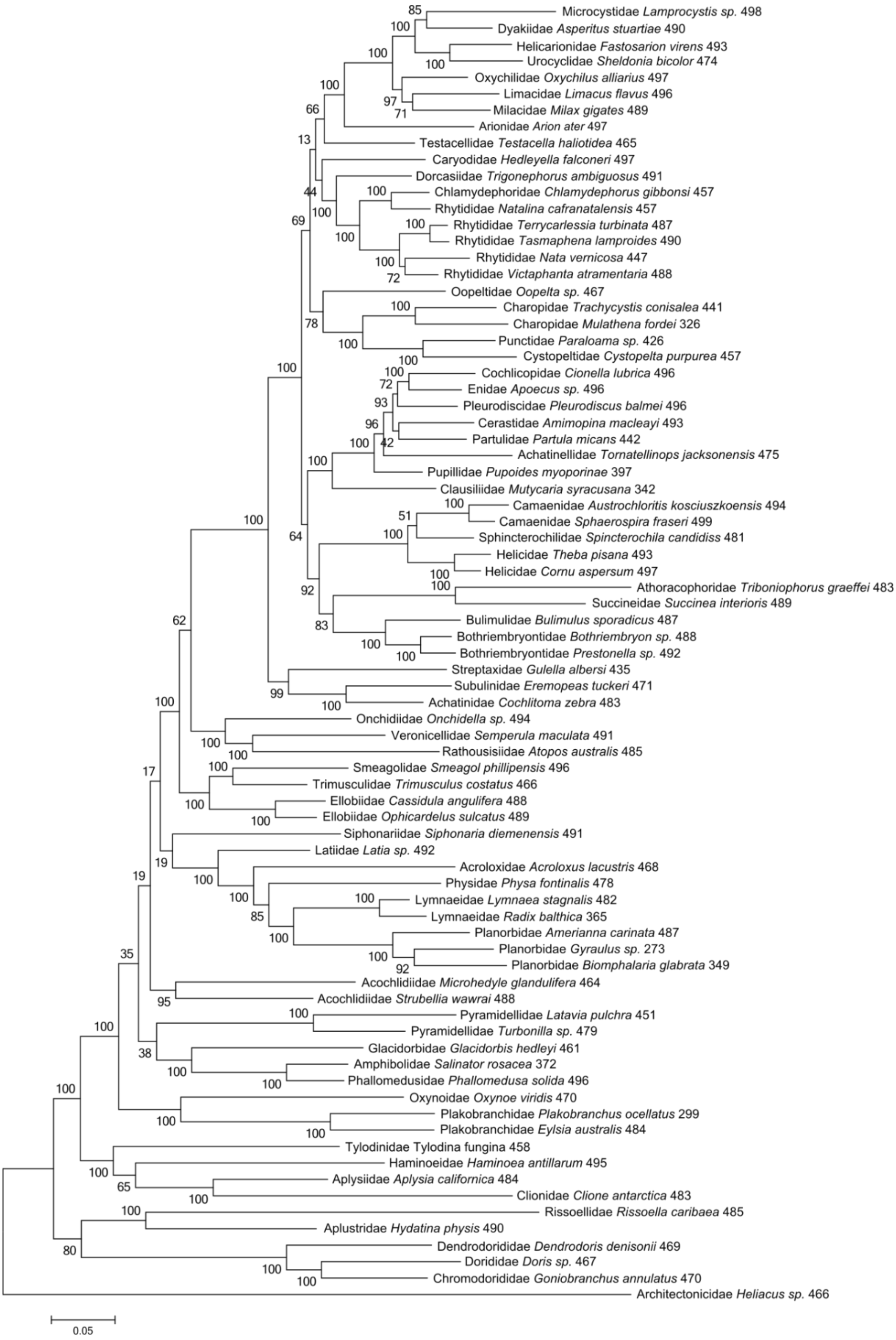
Appendix 3.10. Maximum likelihood tree constructed using the second of eight exon clusters produced by clustering the exons based on amino acid frequencies. This exon cluster consists of 42,785 amino acids. The node labels summarise the boot strap support resulting from 100 fast bootstraps in RAxML.


Helicarionidae Fastosarion virens 493
Urocyclidae Sheldonia bicolor 474
Dyakiidae Asperitus stuartiae 490
Microcystidae Lamprocystis sp. 498
Limacidae Limacus flavus 496
Milacidae Milax gigates 489
Oxychilidae Oxychilus alliarius 497
Arionidae Arion ater 497
Clausiliidae Mutycaria syracusana 342
Pupillidae Pupoides myoporinae 397
Achatinellidae Tornatellinops jacksonensis 475
Cerastidae Amimopina macleayi 493
Pleurodiscidae Pleurodiscus balmei 496
Partulidae Partula micans 442
Cochlicopidae Cionella lubrica 496
Enidae Apoecus sp. 496
Testacellidae Testacella haliotidea 465
Charopidae Trachycystis conisalea 441
Charopidae Mulathena fordei 326
Punctidae Paraloama sp. 426
Cystopeltidae Cystopelta purpurea 457
Oopeltidae Oopelta sp. 467
Caryodidae Hedleyella falconeri 497
Dorcasiidae Trigonephorus ambiguosus 491
Rhytididae Natalina cafranatalensis 457
Chlamydephoridae Chlamydephorus gibbonsi 457
Rhytididae Nata vernicosa 447
Rhytididae Victaphanta atramentaria 488
Rhytididae Terrycarlessia turbinata 487
Rhytididae Tasmaphena lamproides 490
Bothriembryontidae Bothriembryon sp. 488
Bothriembryontidae Prestonella sp. 492
Bulimulidae Bulimulus sporadicus 487
Sphincterochilidae Spincterochila candidiss 481
Helicidae Theba pisana 493
Helicidae Cornu aspersum 497
Camaenidae Sphaerospira fraseri 499
Camaenidae Austrochloritis kosciuszkoensis 494
Succineidae Succinea interioris 489
Athoracophoridae Triboniophorus graeffei 483
Streptaxidae Gulella albersi 435
Subulinidae Eremopeas tuckeri 471
Achatinidae Cochlitoma zebra 483
Onchidiidae Onchidella sp. 494
Veronicellidae Semperula maculata 491
Rathousisiidae Atopos australis 485
Trimusculidae Trimusculus costatus 466
Smeagolidae Smeagol phillipensis 496
Ellobiidae Ophicardelus sulcatus 489
Ellobiidae Cassidula angulifera 488
Latiidae Latia sp. 492
Acroloxidae Acroloxus lacustris 468
Physidae Physa fontinalis 478
Lymnaeidae Radix balthica 365
Lymnaeidae Lymnaea stagnalis 482
Planorbidae Amerianna carinata 487
Planorbidae Biomphalaria glabrata 349
Planorbidae Gyraulus sp. 273
Acochlidiidae Microhedyle glandulifera 464
Acochlidiidae Strubellia wawrai 488
Pyramidellidae Latavia pulchra 451
Pyramidellidae Turbonilla sp. 479
Glacidorbidae Glacidorbis hedleyi 461
Amphibolidae Salinator rosacea 372
Phallomedusidae Phallomedusa solida 496
Siphonariidae Siphonaria diemenensis 491
Oxynoidae Oxynoe viridis 470
Plakobranchidae Plakobranchus ocellatus 299
Plakobranchidae Eylsia australis 484
Tylodinidae Tylodina fungina 458
Haminoeidae Haminoea antillarum 495
Aplysiidae Aplysia californica 484
Clionidae Clione antarctica 483
Rissoellidae Rissoella caribaea 485
Aplustridae Hydatina physis 490
Dendrodorididae Dendrodoris denisonii 469
Dorididae Doris sp. 467
Chromodorididae Goniobranchus annulatus 470
Architectonicidae Heliacus sp. 466
0.05

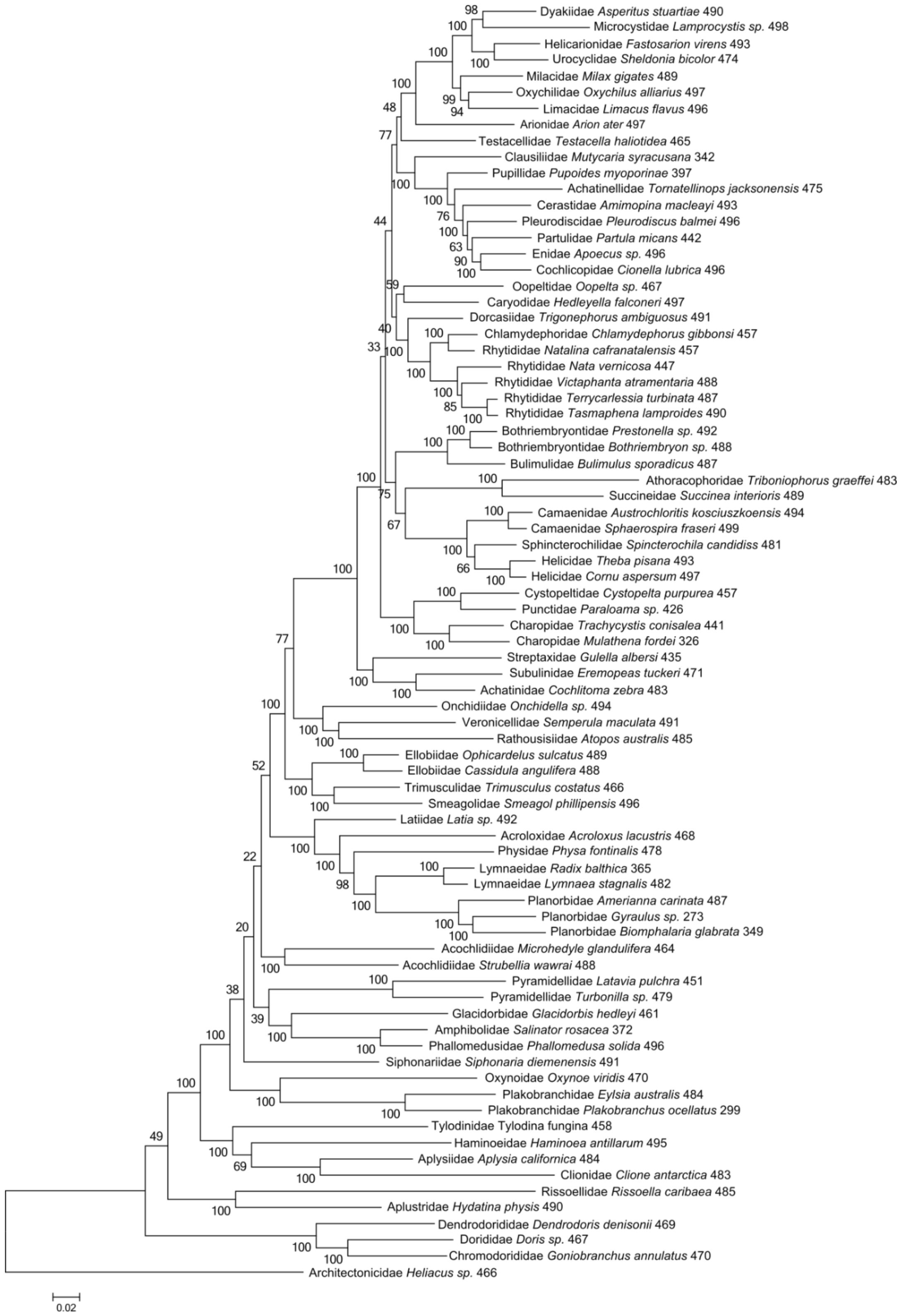
Appendix 3.11. Maximum likelihood tree constructed using the third of eight exon clusters produced by clustering the exons based on amino acid frequencies. This exon cluster consists of 26,044 amino acids. The node labels summarise the boot strap support resulting from 100 fast bootstraps in RAxML.


Rhytididae Terrycarlessia turbinata 487
Rhytididae Tasmaphena lamproides 490
Rhytididae Victaphanta atramentaria 488
Rhytididae Nata vernicosa 447
Rhytididae Natalina cafranatalensis 457
Chlamydephoridae Chlamydephorus gibbonsi 457
Dorcasiidae Trigonephorus ambiguosus 491
Oopeltidae Oopelta sp. 467
Testacellidae Testacella haliotidea 465
Arionidae Arion ater 497
Limacidae Limacus flavus 496
Milacidae Milax gigates 489
Oxychilidae Oxychilus alliarius 497
Urocyclidae Sheldonia bicolor 474
Helicarionidae Fastosarion virens 493
Microcystidae Lamprocystis sp. 498
Dyakiidae Asperitus stuartiae 490
Caryodidae Hedleyella falconeri 497
Bothriembryontidae Bothriembryon sp. 488
Bothriembryontidae Prestonella sp. 492
Bulimulidae Bulimulus sporadicus 487
Athoracophoridae Triboniophorus graeffei 483
Succineidae Succinea interioris 489
Helicidae Theba pisana 493
Helicidae Cornu aspersum 497
Sphincterochilidae Spincterochila candidiss 481
Camaenidae Austrochloritis kosciuszkoensis 494
Camaenidae Sphaerospira fraseri 499
Clausiliidae Mutycaria syracusana 342
Pupillidae Pupoides myoporinae 397
Achatinellidae Tornatellinops jacksonensis 475
Cerastidae Amimopina macleayi 493
Partulidae Partula micans 442
Enidae Apoecus sp. 496
Cochlicopidae Cionella lubrica 496
Pleurodiscidae Pleurodiscus balmei 496
Punctidae Paraloama sp. 426
Cystopeltidae Cystopelta purpurea 457
Charopidae Trachycystis conisalea 441
Charopidae Mulathena fordei 326
Streptaxidae Gulella albersi 435
Achatinidae Cochlitoma zebra 483
Subulinidae Eremopeas tuckeri 471
Onchidiidae Onchidella sp. 494
Veronicellidae Semperula maculata 491
Rathousiidae Atopos australis 485
Trimusculidae Trimusculus costatus 466
Smeagolidae Smeagol phillipensis 496
Ellobiidae Cassidula angulifera 488
Ellobiidae Ophicardelus sulcatus 489
Siphonariidae Siphonaria diemenensis 491
Latiidae Latia sp. 492
Acroloxidae Acroloxus lacustris 468
Physidae Physa fontinalis 478
Lymnaeidae Lymnaea stagnalis 482
Lymnaeidae Radix balthica 365
Planorbidae Biomphalaria glabrata 349
Planorbidae Gyraulus sp. 273
Planorbidae Amerianna carinata 487
Pyramidellidae Latavia pulchra 451
Pyramidellidae Turbonilla sp. 479
Glacidorbidae Glacidorbis hedleyi 461
Phallomedusidae Phallomedusa solida 496
Amphibolidae Salinator rosacea 372
Acochlidiidae Microhedyle glandulifera 464
Acochlidiidae Strubellia wawrai 488
Oxynoidae Oxynoe viridis 470
Plakobranchidae Plakobranchus ocellatus 299
Plakobranchidae Eylsia australis 484
Tylodinidae Tylodina fungina 458
Haminoeidae Haminoea antillarum 495
Aplysiidae Aplysia californica 484
Clionidae Clione antarctica 483
Rissoellidae Rissoella caribaea 485
Aplustridae Hydatina physis 490
Dendrodorididae Dendrodoris denisonii 469
Chromodorididae Goniobranchus annulatus 470
Dorididae Doris sp. 467
Architectonicidae Heliacus sp. 466
0.02

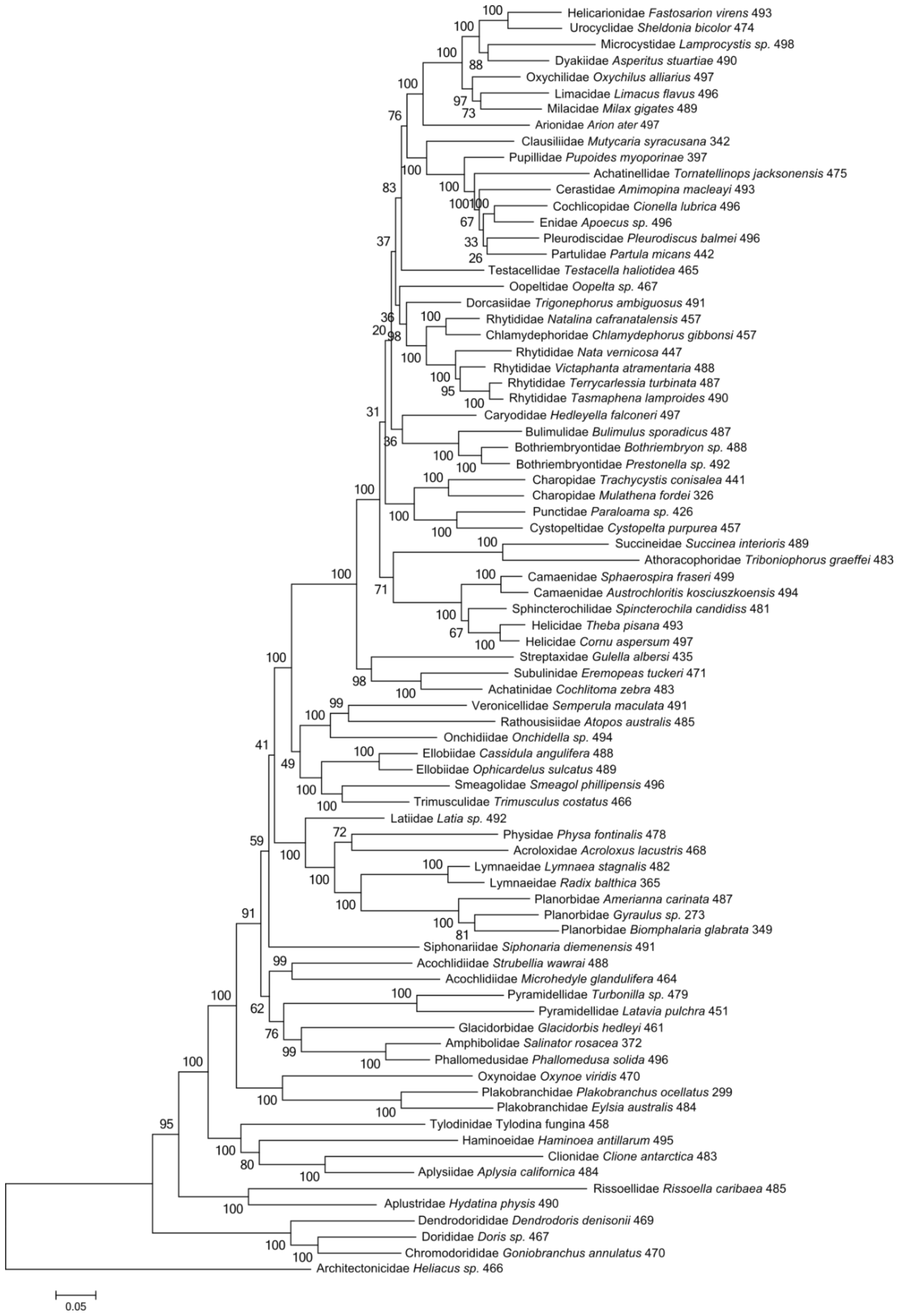
Appendix 3.12. Maximum likelihood tree constructed using the fourth of eight exon clusters produced by clustering the exons based on amino acid frequencies. This exon cluster consists of 7,562 amino acids. The node labels summarise the boot strap support resulting from 100 fast bootstraps in RAxML.


85
Microcystidae Lamprocystis sp. 498
100
Dyakiidae Asperitus stuartiae 490
100
Helicarionidae Fastosarion virens 493
100
Urocyclidae Sheldonia bicolor 474
100
Oxychilidae Oxychilus alliarius 497
97
Limacidae Limacus flavus 496
66
71
Milacidae Milax gigates 489
Arionidae Arion ater 497
13
Testacellidae Testacella haliotidea 465
Caryodidae Hedleyella falconeri 497
Dorcasiidae Trigonephorus ambiguosus 491
44
100
Chlamydephoridae Chlamydephorus gibbonsi 457
100
Rhytididae Natalina cafranatalensis 457
69
100
Rhytididae Terrycarlessia turbinata 487
100
Rhytididae Tasmaphena lamproides 490
100
Rhytididae Nata vernicosa 447
72
Rhytididae Victaphanta atramentaria 488
Oopeltidae Oopelta sp. 467
100
Charopidae Trachycystis conisalea 441
78
Charopidae Mulathena fordei 326
100
Punctidae Paraloama sp. 426
100
100
Cystopeltidae Cystopelta purpurea 457
100
Cochlicopidae Cionella lubrica 496
72
Enidae Apoecus sp. 496
93
Pleurodiscidae Pleurodiscus balmei 496
96
Cerastidae Amimopina macleayi 493
100
42
Partulidae Partula micans 442
Achatinellidae Tornatellinops jacksonensis 475
100
Pupillidae Pupoides myoporinae 397
Clausiliidae Mutycaria syracusana 342
100
Camaenidae Austrochloritis kosciuszkoensis 494
51
Camaenidae Sphaerospira fraseri 499
100
64
Sphincterochilidae Spincterochila candidiss 481
Helicidae Theba pisana 493
100
Helicidae Cornu aspersum 497
100
Athoracophoridae Triboniophorus graeffei 483
92
Succineidae Succinea interioris 489
62
83
Bulimulidae Bulimulus sporadicus 487
100
Bothriembryontidae Bothriembryon sp. 488
100
Bothriembryontidae Prestonella sp. 492
Streptaxidae Gulella albersi 435
99
Subulinidae Eremopeas tuckeri 471
100
100
Achatinidae Cochlitoma zebra 483
Onchidiidae Onchidella sp. 494
100
Veronicellidae Semperula maculata 491
100
Rathousiidae Atopos australis 485
100
Smeagolidae Smeagol phillipensis 496
17
Trimusculidae Trimusculus costatus 466
100
Ellobiidae Cassidula angulifera 488
100
Ellobiidae Ophicardelus sulcatus 489
Siphonariidae Siphonaria diemenensis 491
Latiidae Latia sp. 492
19
Acroloxidae Acroloxus lacustris 468
19
100
Physidae Physa fontinalis 478
100
100
Lymnaeidae Lymnaea stagnalis 482
85
Lymnaeidae Radix balthica 365
100
Planorbidae Amerianna carinata 487
35
100
Planorbidae Gyraulus sp. 273
92
Planorbidae Biomphalaria glabrata 349
Acochlidiidae Microhedyle glandulifera 464
95
Acochlidiidae Strubellia wawrai 488
100
Pyramidellidae Latavia pulchra 451
100
Pyramidellidae Turbonilla sp. 479
38
Glacidorbidae Glacidorbis hedleyi 461
100
Amphibolidae Salinator rosacea 372
100
100
Phallomedusidae Phallomedusa solida 496
Oxynoidae Oxynoe viridis 470
100
Plakobranchidae Plakobranchus ocellatus 299
100
Plakobranchidae Eylsia australis 484
Tylodinidae Tylodina fungina 458
100
Haminoeidae Haminoea antillarum 495
65
Aplysiidae Aplysia californica 484
100
Clionidae Clione antarctica 483
100
Rissoellidae Rissoella caribaea 485
Aplustridae Hydatina physis 490
80
Dendrodorididae Dendrodoris denisonii 469
Dorididae Doris sp. 467
100
100
Chromodorididae Goniobranchus annulatus 470
Architectonicidae Heliacus sp. 466
0.05

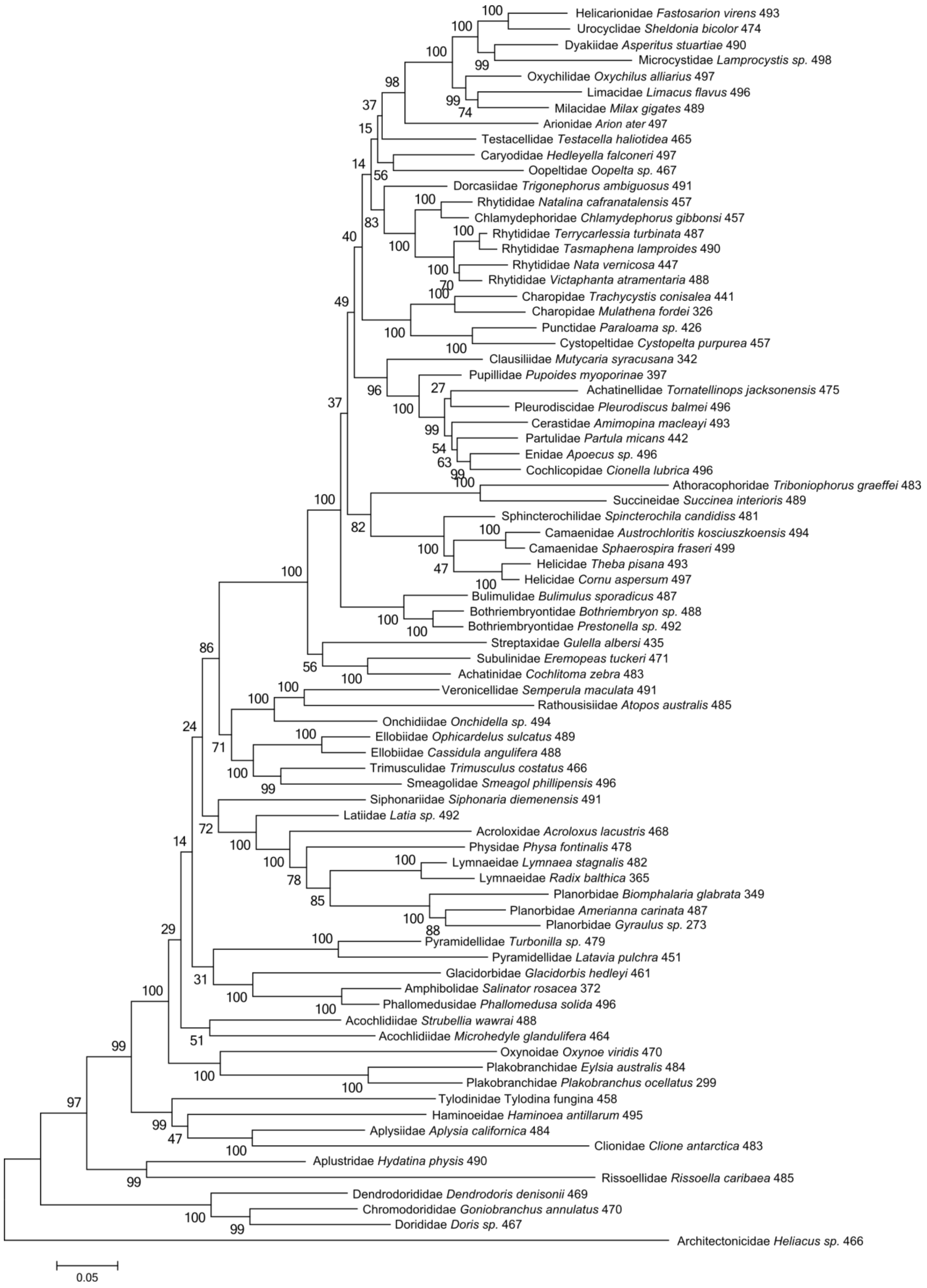
Appendix 3.13. Maximum likelihood tree constructed using the fifth of eight exon clusters produced by clustering the exons based on amino acid frequencies. This exon cluster consists of 15,953 amino acids. The node labels summarise the boot strap support resulting from 100 fast bootstraps in RAxML.


98
Dyakiidae Asperitus stuartiae 490
100
Microcystidae Lamprocystis sp. 498
100
Helicarionidae Fastosarion virens 493
100
Urocyclidae Sheldonia bicolor 474
100
Milacidae Milax gigates 489
48
99
Oxychilidae Oxychilus alliarius 497
94
Limacidae Limacus flavus 496
Arionidae Arion ater 497
77
Testacellidae Testacella haliotidea 465
Clausiliidae Mutycaria syracusana 342
Pupillidae Pupoides myoporinae 397
100
Achatinellidae Tornatellinops jacksonensis 475
100
Cerastidae Amimopina macleayi 493
76
44
Pleurodiscidae Pleurodiscus balmei 496
100
Partulidae Partula micans 442
63
Enidae Apoecus sp. 496
90
Cochlicopidae Cionella lubrica 496
100
59
Oopeltidae Oopelta sp. 467
Caryodidae Hedleyella falconeri 497
Dorcasiidae Trigonephorus ambiguosus 491
40
100
Chlamydephoridae Chlamydephorus gibbonsi 457
33
100
Rhytididae Natalina cafranatalensis 457
Rhytididae Nata vernicosa 447
100
Rhytididae Victaphanta atramentaria 488
100
Rhytididae Terrycarlessia turbinata 487
85
Rhytididae Tasmaphena lamproides 490
100
100
Bothriembryontidae Prestonella sp. 492
100
Bothriembryontidae Bothriembryon sp. 488
Bulimulidae Bulimulus sporadicus 487
100
100
Athoracophoridae Triboniophorus graeffei 483
75
Succineidae Succinea interioris 489
100
Camaenidae Austrochloritis kosciuszkoensis 494
67
Camaenidae Sphaerospira fraseri 499
Sphincterochilidae Spincterochila candidiss 481
100
66
Helicidae Theba pisana 493
100
Helicidae Cornu aspersum 497
100
100
Cystopeltidae Cystopelta purpurea 457
Punctidae Paraloama sp. 426
100
Charopidae Trachycystis conisalea 441
100
Charopidae Mulathena fordei 326
77
Streptaxidae Gulella albersi 435
100
Subulinidae Eremopeas tuckeri 471
100
Achatinidae Cochlitoma zebra 483
100
Onchidiidae Onchidella sp. 494
Veronicellidae Semperula maculata 491
100
100
Rathousisiidae Atopos australis 485
100
Ellobiidae Ophicardelus sulcatus 489
52
Ellobiidae Cassidula angulifera 488
Trimusculidae Trimusculus costatus 466
100
100
Smeagolidae Smeagol phillipensis 496
Latiidae Latia sp. 492
Acroloxidae Acroloxus lacustris 468
100
Physidae Physa fontinalis 478
22
100
100
Lymnaeidae Radix balthica 365
98
Lymnaeidae Lymnaea stagnalis 482
100
Planorbidae Amerianna carinata 487
Planorbidae Gyraulus sp. 273
100
20
100
Planorbidae Biomphalaria glabrata 349
Acochlidiidae Microhedyle glandulifera 464
100
Acochlidiidae Strubellia wawrai 488
100
Pyramidellidae Latavia pulchra 451
38
Pyramidellidae Turbonilla sp. 479
Glacidorbidae Glacidorbis hedleyi 461
39
Amphibolidae Salinator rosacea 372
100
100
Phallomedusidae Phallomedusa solida 496
100
Siphonariidae Siphonaria diemenensis 491
Oxynoidae Oxynoe viridis 470
100
Plakobranchidae Eylsia australis 484
100
100
Plakobranchidae Plakobranchus ocellatus 299
Tylodinidae Tylodina fungina 458
49
Haminoeidae Haminoea antillarum 495
100
Aplysiidae Aplysia californica 484
69
100
Clionidae Clione antarctica 483
Rissoellidae Rissoella caribaea 485
100
Aplustridae Hydatina physis 490
Dendrodorididae Dendrodoris denisonii 469
100
Dorididae Doris sp. 467
100
Chromodorididae Goniobranchus annulatus 470
Architectonicidae Heliacus sp. 466
0.02

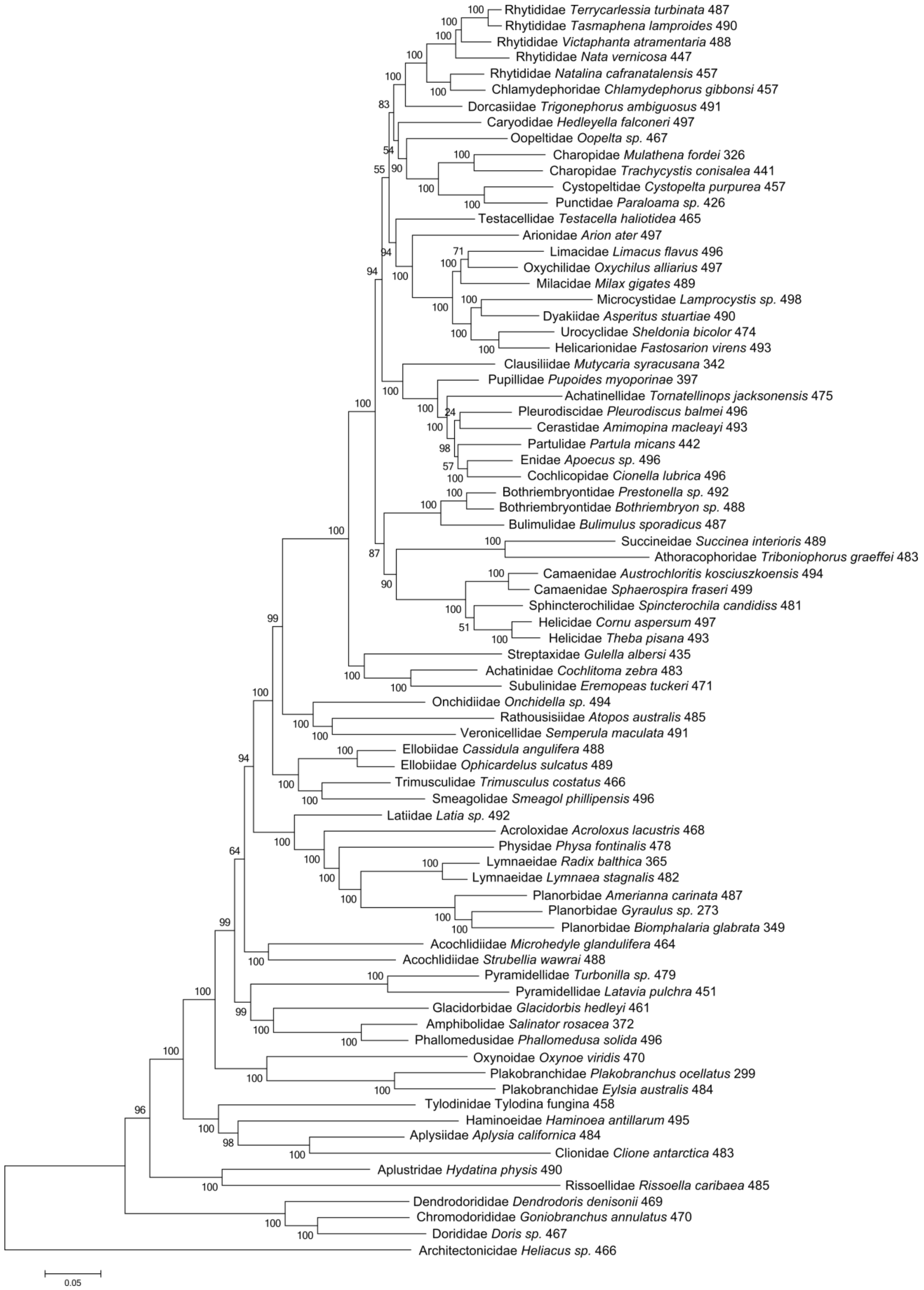
Appendix 3.14. Maximum likelihood tree constructed using the sixth of eight exon clusters produced by clustering the exons based on amino acid frequencies. This exon cluster consists of 19,487 amino acids. The node labels summarise the boot strap support resulting from 100 fast bootstraps in RAxML.


Helicarionidae Fastosarion virens 493
Urocyclidae Sheldonia bicolor 474
Microcystidae Lamprocystis sp. 498
Dyakiidae Asperitus stuartiae 490
Oxychilidae Oxychilus alliarius 497
Limacidae Limacus flavus 496
Milacidae Milax gigates 489
Arionidae Arion ater 497
Clausiliidae Mutycaria syracusana 342
Pupillidae Pupoides myoporinae 397
Achatinellidae Tornatellinops jacksonensis 475
Cerastidae Amimopina macleayi 493
Cochlicopidae Cionella lubrica 496
Enidae Apoecus sp. 496
Pleurodiscidae Pleurodiscus balmei 496
Partulidae Partula micans 442
Testacellidae Testacella haliotidea 465
Oopeltidae Oopelta sp. 467
Dorcasiidae Trigonephorus ambiguosus 491
Rhytididae Natalina cafranatalensis 457
Chlamydephoridae Chlamydephorus gibbonsi 457
Rhytididae Nata vernicosa 447
Rhytididae Victaphanta atramentaria 488
Rhytididae Terrycarlessia turbinata 487
Rhytididae Tasmaphena lamproides 490
Caryodidae Hedleyella falconeri 497
Bulimulidae Bulimulus sporadicus 487
Bothriembryontidae Bothriembryon sp. 488
Bothriembryontidae Prestonella sp. 492
Charopidae Trachycystis conisalea 441
Charopidae Mulathena fordei 326
Punctidae Paraloama sp. 426
Cystopeltidae Cystopelta purpurea 457
Succineidae Succinea interioris 489
Athoracophoridae Triboniophorus graeffei 483
Camaenidae Sphaerospira fraseri 499
Camaenidae Austrochloritis kosciuszkoensis 494
Sphincterochilidae Spincterochila candidiss 481
Helicidae Theba pisana 493
Helicidae Cornu aspersum 497
Streptaxidae Gulella albersi 435
Subulinidae Eremopeas tuckeri 471
Achatinidae Cochlitoma zebra 483
Veronicellidae Semperula maculata 491
Rathousiidae Atopos australis 485
Onchidiidae Onchidella sp. 494
Ellobiidae Cassidula angulifera 488
Ellobiidae Ophicardelus sulcatus 489
Smeagolidae Smeagol phillipensis 496
Trimusculidae Trimusculus costatus 466
Latiidae Latia sp. 492
Physidae Physa fontinalis 478
Acroloxidae Acroloxus lacustris 468
Lymnaeidae Lymnaea stagnalis 482
Lymnaeidae Radix balthica 365
Planorbidae Amerianna carinata 487
Planorbidae Gyraulus sp. 273
Planorbidae Biomphalaria glabrata 349
Siphonariidae Siphonaria diemenensis 491
Acochlidiidae Strubellia wawrai 488
Acochlidiidae Microhedyle glandulifera 464
Pyramidellidae Turbonilla sp. 479
Pyramidellidae Latavia pulchra 451
Glacidorbidae Glacidorbis hedleyi 461
Amphibolidae Salinator rosacea 372
Phallomedusidae Phallomedusa solida 496
Oxynoidae Oxynoe viridis 470
Plakobranchidae Plakobranchus ocellatus 299
Plakobranchidae Eylsia australis 484
Tylodinidae Tylodina fungina 458
Haminoeidae Haminoea antillarum 495
Clionidae Clione antarctica 483
Aplysiidae Aplysia californica 484
Rissoellidae Rissoella caribaea 485
Aplustridae Hydatina physis 490
Dendrodorididae Dendrodoris denisonii 469
Dorididae Doris sp. 467
Chromodorididae Goniobranchus annulatus 470
Architectonicidae Heliacus sp. 466
0.05

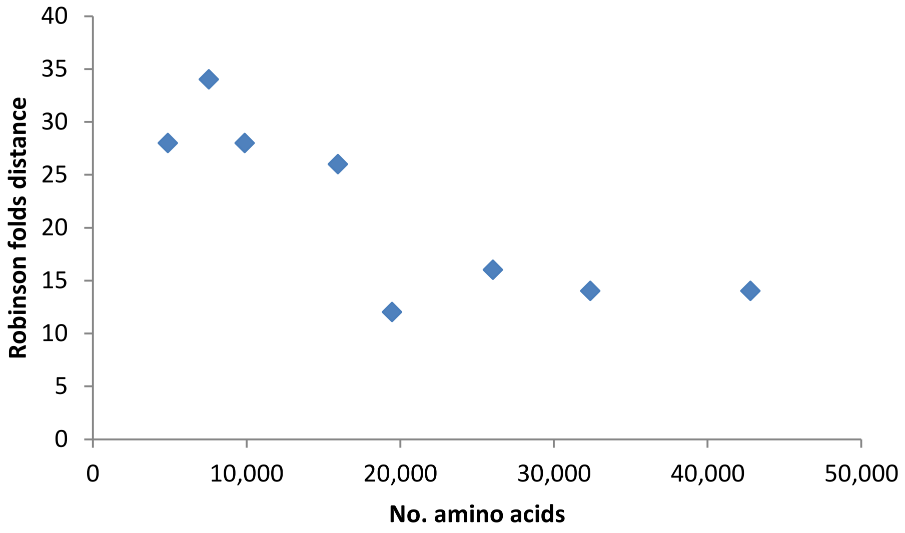
Appendix 3.15. Maximum likelihood tree constructed using the seventh of eight exon clusters produced by clustering the exons based on amino acid frequencies. This exon cluster consists of 9,895 amino acids. The node labels summarise the boot strap support resulting from 100 fast bootstraps in RAxML.


Helicarionidae Fastosarion virens 493
Urocyclidae Sheldonia bicolor 474
Dyakiidae Asperitus stuartiae 490
Microcystidae Lamprocystis sp. 498
Oxychilidae Oxychilus alliarius 497
Limacidae Limacus flavus 496
Milacidae Milax gigates 489
Arionidae Arion ater 497
Testacellidae Testacella haliotidea 465
Caryodidae Hedleyella falconeri 497
Oopeltidae Oopelta sp. 467
Dorcasiidae Trigonephorus ambiguosus 491
Rhytididae Natalina cafranatalensis 457
Chlamydephoridae Chlamydephorus gibbonsi 457
Rhytididae Terrycarlessia turbinata 487
Rhytididae Tasmaphena lamproides 490
Rhytididae Nata vernicosa 447
Rhytididae Victaphanta atramentaria 488
Charopidae Trachycystis conisalea 441
Charopidae Mulathena fordei 326
Punctidae Paraloama sp. 426
Cystopeltidae Cystopelta purpurea 457
Clausiliidae Mutycaria syracusana 342
Pupillidae Pupoides myoporinae 397
Achatinellidae Tornatellinops jacksonensis 475
Pleurodiscidae Pleurodiscus balmei 496
Cerastidae Amimopina macleayi 493
Partulidae Partula micans 442
Enidae Apoecus sp. 496
Cochlicopidae Cionella lubrica 496
Athoracophoridae Triboniophorus graeffei 483
Succineidae Succinea interioris 489
Sphincterochilidae Spincterochila candidiss 481
Camaenidae Austrochloritis kosciuszkoensis 494
Camaenidae Sphaerospira fraseri 499
Helicidae Theba pisana 493
Helicidae Cornu aspersum 497
Bulimulidae Bulimulus sporadicus 487
Bothriembryontidae Bothriembryon sp. 488
Bothriembryontidae Prestonella sp. 492
Streptaxidae Gulella albersi 435
Subulinidae Eremopeas tuckeri 471
Achatinidae Cochlitoma zebra 483
Veronicellidae Semperula maculata 491
Rathousisiidae Atopos australis 485
Onchidiidae Onchidella sp. 494
Ellobiidae Ophicardelus sulcatus 489
Ellobiidae Cassidula angulifera 488
Trimusculidae Trimusculus costatus 466
Smeagolidae Smeagol phillipensis 496
Siphonariidae Siphonaria diemenensis 491
Latiidae Latia sp. 492
Acroloxidae Acroloxus lacustris 468
Physidae Physa fontinalis 478
Lymnaeidae Lymnaea stagnalis 482
Lymnaeidae Radix balthica 365
Planorbidae Biomphalaria glabrata 349
Planorbidae Amerianna carinata 487
Planorbidae Gyraulus sp. 273
Pyramidellidae Turbonilla sp. 479
Pyramidellidae Latavia pulchra 451
Glacidorbidae Glacidorbis hedleyi 461
Amphibolidae Salinator rosacea 372
Phallomedusidae Phallomedusa solida 496
Acochlidiidae Strubellia wawrai 488
Acochlidiidae Microhedyle glandulifera 464
Oxynoidae Oxynoe viridis 470
Plakobranchidae Eylsia australis 484
Plakobranchidae Plakobranchus ocellatus 299
Tylodinidae Tylodina fungina 458
Haminoeidae Haminoea antillarum 495
Aplysiidae Aplysia californica 484
Clionidae Clione antarctica 483
Aplustridae Hydatina physis 490
Rissoellidae Rissoella caribaea 485
Dendrodorididae Dendrodoris denisonii 469
Chromodorididae Goniobranchus annulatus 470
Dorididae Doris sp. 467
Architectonicidae Heliacus sp. 466
100
100
100
99
98
37
99
74
15
14
56
100
83
40
100
100
100
70
100
49
100
100
96
27
37
100
99
54
63
99
100
100
82
100
100
47
100
100
100
100
86
56
100
100
100
24
100
71
100
99
72
14
100
100
100
78
85
100
88
29
100
31
100
100
100
51
99
100
100
97
99
47
100
99
100
99
0.05

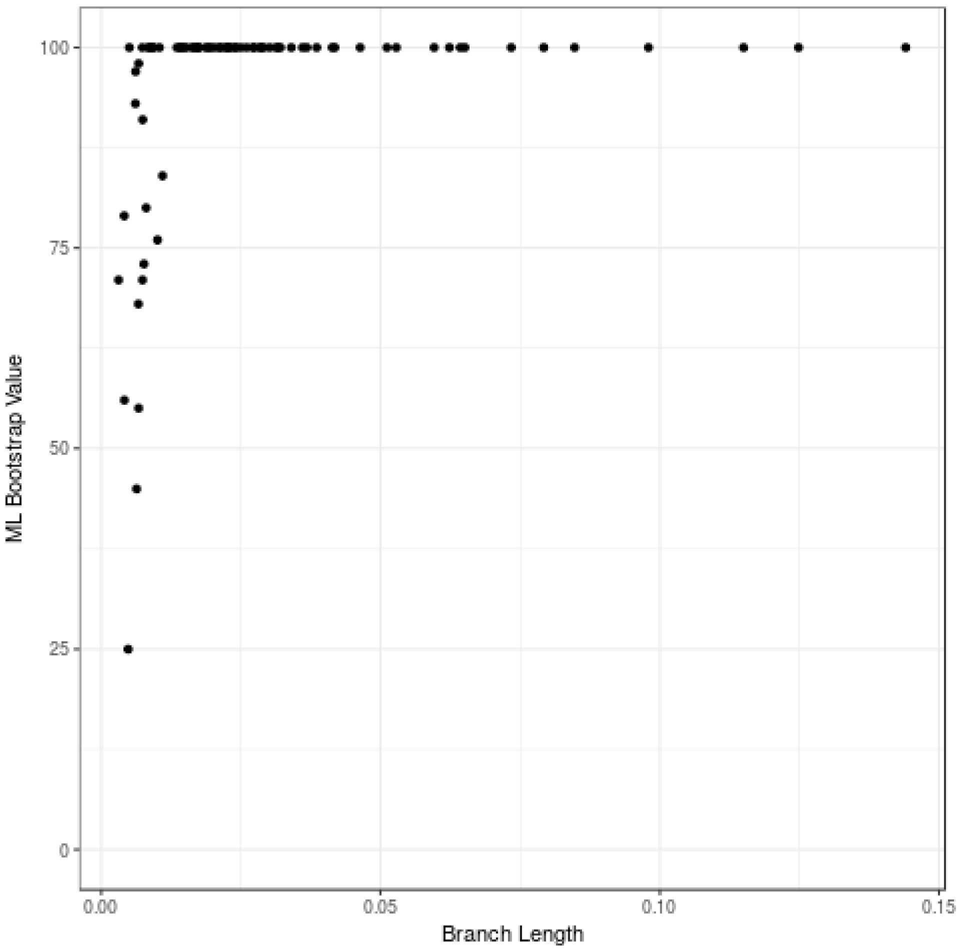
Appendix 3.16. Maximum likelihood tree constructed using the eighth of eight exon clusters produced by clustering the exons based on amino acid frequencies. This exon cluster consists of 4,898 amino acids. The node labels summarise the boot strap support resulting from 100 fast bootstraps in RAxML.


Rhytididae Terrycarlessia turbinata 487
Rhytididae Tasmaphena lamproides 490
Rhytididae Victaphanta atramentaria 488
Rhytididae Nata vernicosa 447
Rhytididae Natalina cafranatalensis 457
Chlamydephoridae Chlamydephorus gibbonsi 457
Dorcasiidae Trigonephorus ambiguosus 491
Caryodidae Hedleyella falconeri 497
Oopeltidae Oopelta sp. 467
Charopidae Mulathena fordei 326
Charopidae Trachycystis conisalea 441
Cystopeltidae Cystopelta purpurea 457
Punctidae Paraloama sp. 426
Testacellidae Testacella haliotidea 465
Arionidae Arion ater 497
Limacidae Limacus flavus 496
Oxychilidae Oxychilus alliarius 497
Milacidae Milax gigates 489
Microcystidae Lamprocystis sp. 498
Dyakiidae Asperitus stuartiae 490
Urocyclidae Sheldonia bicolor 474
Helicarionidae Fastosarion virens 493
Clausiliidae Mutycaria syracusana 342
Pupillidae Pupoides myoporinae 397
Achatinellidae Tornatellinops jacksonensis 475
Pleurodiscidae Pleurodiscus balmei 496
Cerastidae Amimopina macleayi 493
Partulidae Partula micans 442
Enidae Apoecus sp. 496
Cochlicopidae Cionella lubrica 496
Bothriembryontidae Prestonella sp. 492
Bothriembryontidae Bothriembryon sp. 488
Bulimulidae Bulimulus sporadicus 487
Succineidae Succinea interioris 489
Athoracophoridae Triboniophorus graeffei 483
Camaenidae Austrochloritis kosciuszkoensis 494
Camaenidae Sphaerospira fraseri 499
Sphincterochilidae Spincterochila candidiss 481
Helicidae Cornu aspersum 497
Helicidae Theba pisana 493
Streptaxidae Gulella albersi 435
Achatinidae Cochlitoma zebra 483
Subulinidae Eremopeas tuckeri 471
Onchidiidae Onchidella sp. 494
Rathousisiidae Atopos australis 485
Veronicellidae Semperula maculata 491
Ellobiidae Cassidula angulifera 488
Ellobiidae Ophicardelus sulcatus 489
Trimusculidae Trimusculus costatus 466
Smeagolidae Smeagol phillipensis 496
Latiidae Latia sp. 492
Acroloxidae Acroloxus lacustris 468
Physidae Physa fontinalis 478
Lymnaeidae Radix balthica 365
Lymnaeidae Lymnaea stagnalis 482
Planorbidae Amerianna carinata 487
Planorbidae Gyraulus sp. 273
Planorbidae Biomphalaria glabrata 349
Acochlidiidae Microhedyle glandulifera 464
Acochlidiidae Strubellia wawrai 488
Pyramidellidae Turbonilla sp. 479
Pyramidellidae Latavia pulchra 451
Glacidorbidae Glacidorbis hedleyi 461
Amphibolidae Salinator rosacea 372
Phallomedusidae Phallomedusa solida 496
Oxynoidae Oxynoe viridis 470
Plakobranchidae Plakobranchus ocellatus 299
Plakobranchidae Eylsia australis 484
Tylodinidae Tylodina fungina 458
Haminoeidae Haminoea antillarum 495
Aplysiidae Aplysia californica 484
Clionidae Clione antarctica 483
Aplustridae Hydatina physis 490
Rissoellidae Rissoella caribaea 485
Dendrodorididae Dendrodoris denisonii 469
Chromodorididae Goniobranchus annulatus 470
Dorididae Doris sp. 467
Architectonicidae Heliacus sp. 466
0.05

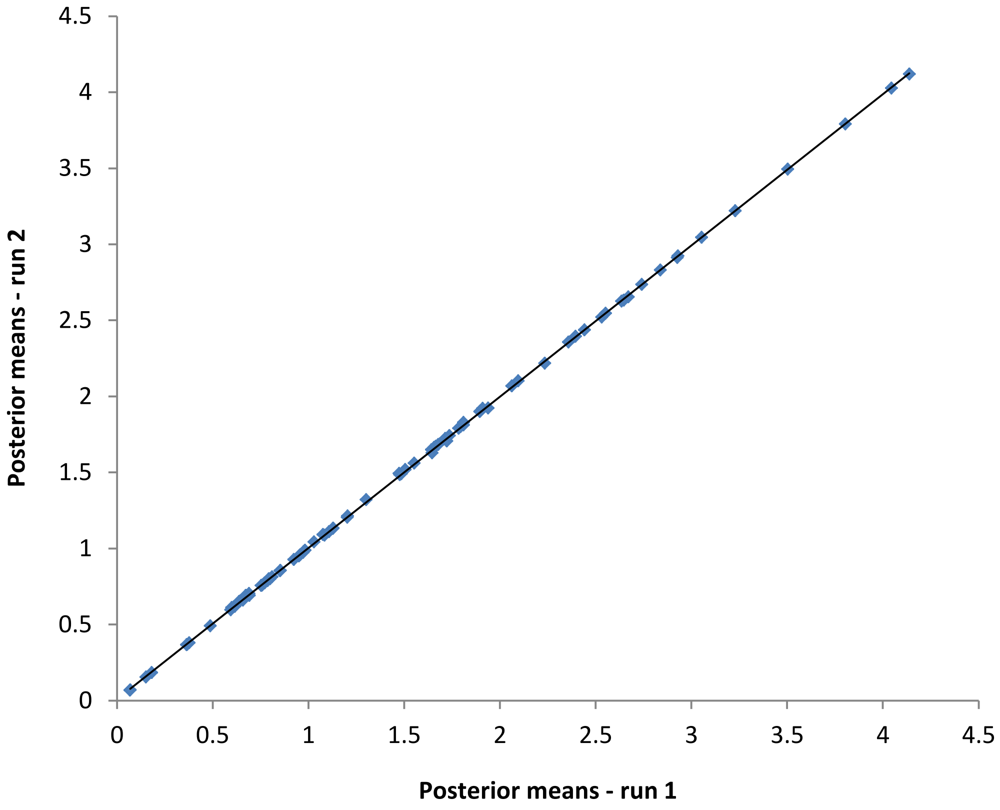
Appendix 3.17. Maximum likelihood phylogeny of Panpulmonata estimated from using eight exon data partitions with *Siphonaria diemenensis* removed from the dataset. The node labels represent the fast bootstrap support. The number after the species names represents the number of the 500 genes present for each taxon.



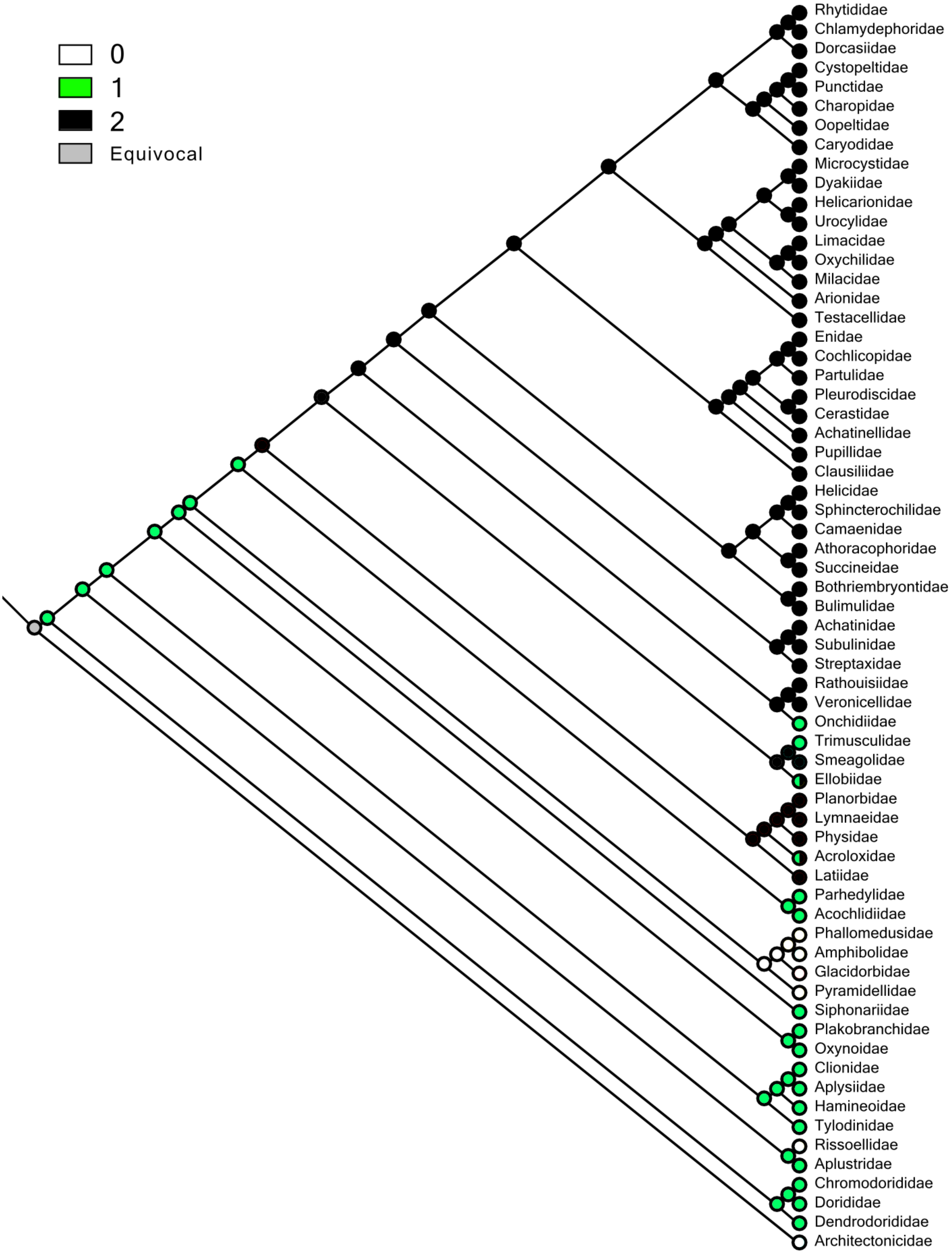
Appendix 3.18. Correlation between exon cluster size and similarity to the maximum likelihood tree produced using all eight exon clusters.



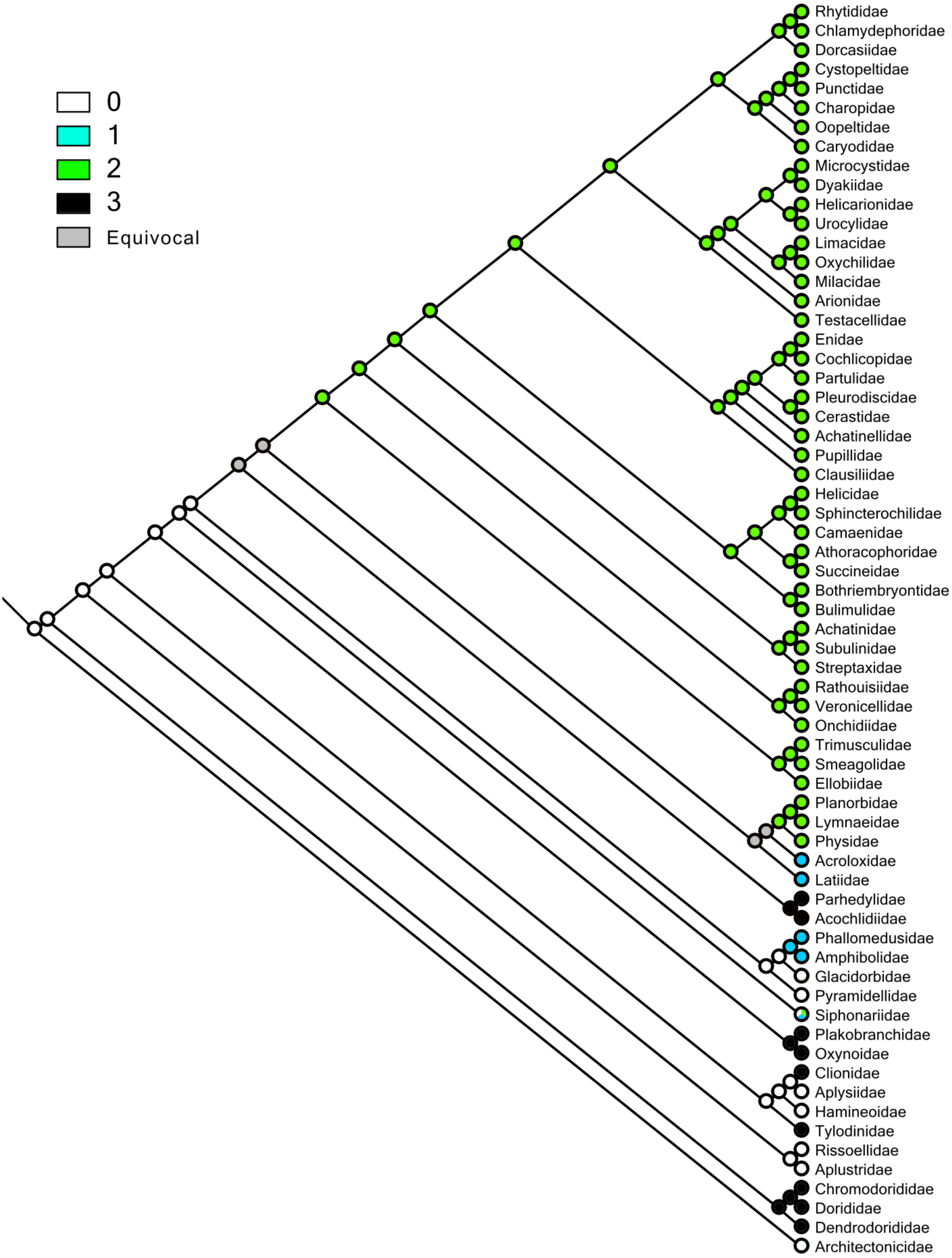
Appendix 3.19. The Correlation between the branch lengths, estimated with maximum likelihood analysis and the eight exon partitioning scheme, and the corresponding bootstrap support values.



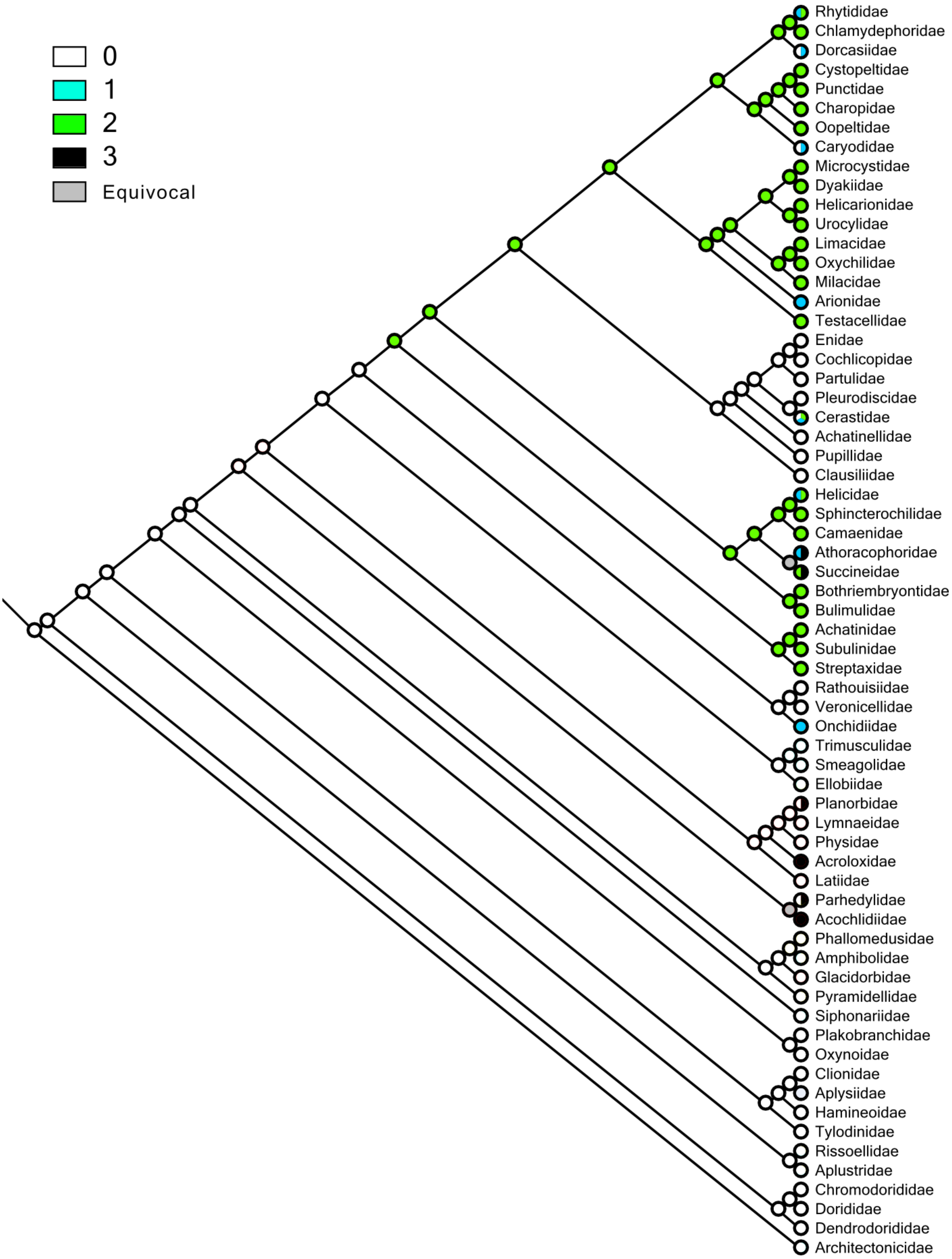
Appendix 3.20. Correlation between the posterior means of two independent MCMCTREE chains ($R^2 = 0.999$). A linear relationship shows that convergence has been reached.



Appendix 3.21. Ancestral state reconstruction of the opening of operculum across Panpulmonata (see Appendix 3.25 for detailed description of the character states).



Appendix 3.22. Ancestral state reconstruction of the opening of the pallial cavity across Panpulmonata (see Appendix 3.25 for detailed description of the character states).



Appendix 3.23. Ancestral state reconstruction of the closed secondary ureter across Panpulmonata (see Appendix 3.25 for detailed description of the character states).

Appendix 3.24. Sample information.

Superfamily	Family	Species	Source	Collector	Specimen voucher	Sequence accession	Origin
Acavoidea	Caryodidae	*Hedleyella falconeri*	Sequenced herein	Adnan Moussalli	-	-	QLD, Australia
Acavoidea	Dorcasiidae	*Trigonephorus ambiguous*	“	Adnan Moussalli	-	-	South Africa
Achatinelloidea	Achatinellidae	*Tornatellinops jacksonensis*	“	Frank Köhler	not vouchered	-	NSW, Australia
Achatinoidea	Achatinidae	*Cochlitoma zebra*	“	Dai Herbert	-	-	South Africa
Achatinoidea	Subulinidae	*Eremopeas tuckeri*	“	Adnan Moussalli	-	-	NT, Australia
Acochlidioidea	Parhedylidae	*Microhedyle glandulifera*	Zapata et al. 2014	-	-	SRR1505118	-
Acochlidioidea	Acochlidiidae	*Strubellia wawrai*	“	-	-	SRR1505137	-
Acroloxoidea	Acroloxidae	*Acroloxus lacustris*	Sequenced herein	Christian Albrecht	not vouchered	-	Mecklenburg-Lower Pommerania, Germany
Acteonoidea	Aplustridae	*Hydatina physis*	“	Adnan Moussalli	-	-	South Africa
Amphiboloidea	Amphibolidae	*Salinator rosacea*	“	Adnan Moussalli	-	-	Darwin, NT, Australia
Amphiboloidea	Phallomedusidae	*Phallomedusa solida*	Zapata et al. 2014	-	-	SRR1505127	-
Aplysioidea	Aplysiidae	*Aplysia californica*	Broad Institute	-	-	PRJNA209509	-
Architectonicoidea	Architectonicidae	*Heliacus sp.*	Sequenced herein	Frank Köhler	AM C.480254	-	Long Reef, NSW, Australia
Arionoidea	Oopeltidae	*Oopelta sp.*	“	Adnan Moussalli	-	-	South Africa
Athoracophoroidea	Athoracophoridae	*Triboniophorus graeffei*	“	Adnan Moussalli	-	-	Mt Warning , NSW, Australia
Chilinoidea	Latiidae	*Latia sp.*	“	Adnan Moussalli	-	-	New Zealand
Clausilioidea	Clausiliidae	*Muticaria syracusana*	“	Danilo Scuderi	AM C.478879	-	Sicily, Italy
Clionoidea	Clionidae	*Clione antarctica*	Zapata et al. 2014	-	-	SRR1505107	-
Cochlicopoidea	Cochlicopidae	*Cionella lubrica*	Teasdale et al. 2016	Frank Köhler	MV614	-	Blue Mountains, NSW, Australia
Doridoidea	Chromodorididae	*Goniobranchus annulatus*	Sequenced herein	Adnan Moussalli	-	-	South Africa
Doridoidea	Dorididae	*Doris kerguelenensis*	Zapata et al. 2014	-	-	SRR1505108	-
Dyakioidea	Dyakiidae	*Asperitas cf stuartiae*	Sequenced herein	Frank Köhler	NMV F193286	-	Dili, Timor-Leste
Ellobioidea	Ellobiidae	*Cassidula angulifera*	Teasdale et al. 2016	Adnan Moussalli	NMV F193289	-	Manatuto, Timor-Leste
Ellobioidea	Ellobiidae	*Ophicardelus sulcatus*	Zapata et al. 2014	-	-	SRR1505124	-
Enoidea	Cerastidae	*Amimopina macleayi*	Teasdale et al. 2016	Adnan Moussalli	NMV F193290	-	Darwin, NT, Australia

Superfamily	Family	Species	Source	Collector	Specimen voucher	Sequence accession	Origin
Enoidea	Enidae	*Apoecus ramelauensis*	“	Frank Köhler	AM C.488753	-	Timor-Leste
Gastrodontoidea	Microcystidae	*Lamprocystis sp.*	“	Frank Köhler	AM C.476947	-	Timor-Leste
Gastrodontoidea	Oxychilidae	*Oxychilus alliarius*	“	Adnan Moussalli	NMV F226626	-	Melbourne, VIC, Australia
Glacidorboidea	Glacidorbidae	*Glacidorbis hedleyi*	Sequenced herein	Frank Köhler	AM C.478607	-	Blue Mountains, NSW, Australia
Haminoeoidea	Haminoeidae	*Haminoea antillarum*	Zapata et al. 2014	-	-	SRR1505111	-
Helicarionoidea	Arionidae	*Arion ater*	Sequenced herein	Stephen Teasdale	-	-	Mt Dandenong, Vic, Australia
Helicarionoidea	Helicarionidae	*Fastosarion virens*	Teasdale et al. 2016	Adnan Moussalli	NMV F193282	-	Noosa, QLD, Australia
Helicarionoidea	Urocyclidae	*Sheldonia bicolor*	Sequenced herein	Dai Herbert	-	-	South Africa
Helicoidea	Camaenidae	*Austrochloritis kosciuszkoensis*	Teasdale et al. 2016	Adnan Moussalli	NMV F193285	-	Sylvia Creek, VIC, Australia
Helicoidea	Camaenidae	*Sphaerospira fraseri*	“	Adnan Moussalli	NMV F193284	-	Noosa, QLD, Australia
Helicoidea	Helicidae	*Cornu aspersum*	“	Adnan Moussalli	NMV F193280	-	Melbourne, VIC, Australia
Helicoidea	Helicidae	*Theba pisana*	Sequenced herein	Frank Köhler	WAM S66455	-	WA, Australia
Helicoidea	Sphincterochilidae	*Spincterochila candidissima*	“	Danilo Scuderi	AM C.478873	-	Sicily, Italy
Limacoidea	Limacidae	*Limax flavus*	Teasdale et al. 2016	Adnan Moussalli	NMV F193283	-	Melbourne, VIC, Australia
Lymnaeoidea	Lymnaeidae	*Lymnaea stagnalis*	Sadamoto et al. 2012	-	-	PRJDB98	-
Lymnaeoidea	Lymnaeidae	*Radix balthica*	Feldmeyer et al. 2011	-	-	-	-
Onchidioidea	Onchidiidae	*Onchidella sp.*	Sequenced herein	Tim O'Hara	-	-	-
Orthalicoidea	Bothriembryontidae	*Bothriembryon sp* .	“	-	-	-	-
Orthalicoidea	Bothriembryontidae	*Prestonella sp.*	“	Adnan Moussalli	-	-	South Africa
Orthalicoidea	Bulimulidae	*Bulimulus sporadicus*	“	-	-	-	South Africa
Otinoidea	Smeagolidae	*Smeagol phillipensis*	Teasdale et al. 2016	Adnan Moussalli	MVR13_138	-	Phillip Is. VIC, Australia
Oxynooidea	Oxynoidae	*Oxynoe viridis*	Sequenced herein	Frank Köhler	AM C.478603	-	Narrabeen Beach, NSW, Australia
Parmacelloidea	Milacidae	*Milax gigates*	Teasdale et al. 2016	Adnan Moussalli	NMV F226625	-	Melbourne, VIC, Australia
Partuloidea	Partulidae	*Partula micans*	Sequenced herein	Diarmaid Ó Foighil	UMMZ304355	-	Solomon Islands
Phyllidioidea	Dendrodorididae	*Dendrodoris denisonii*	“	Adnan Moussalli		-	South Africa
Plakobranchoidea	Plakobranchidae	*Eylsia australis*	“	Frank Köhler	AM C.478604	-	Long Reef, NSW, Australia
Plakobranchoidea	Plakobranchidae	*Plakobranchus ocellatus*	Wägele et al. 2010	-	-	PRJNA52099	-
Planorboidea	Physidae	*Physa fontinalis*	Sequenced herein	Frank Köhler	not vouchered	-	Brandenburg, Germany

Superfamily	**Family**	**Species**	**Source**	**Collector**	**Specimen voucher**	**Sequence accession**	**Origin**
Planorboidea	Planorbidae	*Amerianna carinata*	“	Adnan Moussalli	-	-	Fogg Dam, NT, Australia
Planorboidea	Planorbidae	*Biomphalaria glabrata*	Snaildb	-	-	-	-
Planorboidea	Planorbidae	*Gyraulus sp.*	Sequenced herein	Adnan Moussalli	-	-	Fogg Dam, NT, Australia
Punctoidea	Charopidae	*Mulathena fordei*	“	Adnan Moussalli	-	-	Wilsons Promontory, VIC, Australia
Punctoidea	Charopidae	*Trachycystis conisalea*	“	Dai Herbert	-	-	South Africa
Punctoidea	Cystopeltidae	*Cystopelta purpurea*	“	Adnan Moussalli	-	-	-
Punctoidea	Punctidae	*Paraloama sp.*	“	-	-	-	-
Pupilloidea	Pleurodiscidae	*Pleurodiscus balmei*	“	Frank Köhler	AM C.487473	-	Sydney, NSW, Australia
Pupilloidea	Pupillidae	*Pupoides myoporinae*	“	Adnan Moussalli	-	-	Ned's Corner, VIC, Australia
Pyramidelloidea	Pyramidellidae	*Latavia pulchra*	“	Frank Köhler	AM C.480250	-	Long Reef, NSW, Australia
Pyramidelloidea	Pyramidellidae	*Turbonilla sp.*	Zapata et al. 2014	-	-	SRR1505139	-
Rhytidoidea	Chlamydephoridae	*Chlamydephorus gibbonsi*	Sequenced herein	Dai Herbert	-	-	South Africa
Rhytidoidea	Rhytididae	*Nata vernicosa*	“	Dai Herbert	-	-	South Africa
Rhytidoidea	Rhytididae	*Natalina cafranatalensis*	“	Dai Herbert	-	-	South Africa
Rhytidoidea	Rhytididae	*Tasmaphena lamproides*	“	Adnan Moussalli	-	-	Wilsons Promontory, VIC, Australia
Rhytidoidea	Rhytididae	*Terrycarlessia turbinata*	Teasdale et al. 2016	Adnan Moussalli	NMV F193292	-	Comboyne, NSW, Australia
Rhytidoidea	Rhytididae	*Victaphanta atramenteria*	“	Adnan Moussalli	NMV F226627	-	Toolangi, VIC, Australia
Rissoelloidea	Rissoellidae	*Rissoella caribaea*	Zapata et al. 2014	-	-	SRR1505135	-
Siphonarioidea	Siphonariidae	*Siphonaria diemenensis*	Sequenced herein	Adnan Moussalli	-	-	-
Streptaxoidea	Streptaxidae	*Gulella albersi*	“	Dai Herbert	-	-	South Africa
Succineoidea	Succineidae	*Succinea interioris*	“	Adnan Moussalli	-	-	Mt Brown, SA, Australia
Testacelloidea	Testacellidae	*Testacella haliotidea*	“	Winston Ponder	AM C.478525	-	Mosman, NSW, Australia
Trimusculoidea	Trimusculidae	*Trimusculus costatus*	“	Adnan Moussalli	-	-	South Africa
Umbraculoidea	Tylodinidae	*Tylodina fungina*	Zapata et al. 2014	-	-	SRR1505140	-
Veronicelloidea	Rathouisiidae	*Atopos australis*	Sequenced herein	Adnan Moussalli	-	-	Brisbane, QLD, Australia
Veronicelloidea	Veronicellidae	*Semperula maculata*	Teasdale et al. 2016	Frank Köhler	AM C.476934	-	Manatuto, Timor-Leste

Appendix 3.25. Morphological and life history characters.

Character	Score	State
Operculum	0	present in early ontogeny and retained in adult
	1	present early in ontogeny cast, near or at metamorphosis, or retained through metamorphosis, but shed early in post-embryonic life, thus absent in adult
	2	absent throughout ontogeny
Pallial opening	0	widely open
	1	opening narrowed (to pneumostome-like passage)
	2	opening narrowed to contractile pneumostome
	3	cavity reduced, opening minute or without opening to body exterior
Closed secondary ureter	0	absent
	1	short (<1.5x long axis of kidney)
	2	long (~1.5-5x long axis of kidney), for most part running alongside rectum
	3	very long (>2x long axis of kidney), convoluted, secondarily disassociated with rectum
Habitat	0	marine littoral
	1	supratidal
	2	brackish-water
	3	freshwater
	4	terrestrial

Appendix 3.26. Morphological characters scored per family. The ampersands indicate families that contain species that have different states.

Family	Operculum	Pallial opening	Closed secondary ureter	Habitat
Caryodidae	2	2	0 & 1	4
Dorcasiidae	2	2	0 & 1	4
Achatinellidae	2	2	0	4
Arionidae	2	2	1	4
Athoracophoridae	2	2	1 & 3	4
Succineidae	2	2	2 & 3	4
Bulimulidae	2	2	2	4
Bothriembryontidae	2	2	2	4
Camaenidae	2	2	2	4
Helicidae	2	2	1 & 2	4
Sphincterochilidae	2	2	2	4
Charopidae	2	2	2	4
Punctidae	2	2	2	4
Cystopeltidae	2	2	2	4
Clausiliidae	2	2	0	4
Cochlicopidae	2	2	0	4
Dyakiidae	2	2	2	4
Cerastidae	2	2	0 & 1 & 2	4
Enidae	2	2	0	4
Microcystidae	2	2	2	4
Helicarionidae	2	2	2	4
Urocylidae	2	2	2	4
Limacidae	2	2	2	4
Milacidae	2	2	2	4
Oxychilidae	2	2	2	4
Oopeltidae	2	2	2	4
Partulidae	2	2	0	4
Pleurodiscidae	2	2	0	4
Pupillidae	2	2	0	4
Rhytididae	2	2	1 & 2	4
Chlamydephoridae	2	2	2	4
Streptaxidae	2	2	2	4
Achatinidae	2	2	2	4
Subulinidae	2	2	2	4
Testacellidae	2	2	2	4
Siphonariidae	1	0 & 1 & 2	0	0
Trimusculidae	1	2	0	0
Phallomedusidae	0	1	0	2
Amphibolidae	0	1	0	0& 2
Ellobiidae	1 & 2	2	0	0 & 1 & 2 & 4
Latiidae	2	1	0	3
Smeagolidae	2	2	0	0

Family	Operculum	Pallial opening	Closed secondary ureter	Habitat
Onchidiidae	1	2	1	0 & 1 & 2 & 3 & 4
Rathouisiidae	2	2	0	4
Veronicellidae	2	2	0	4
Acroloxidae	1 & 2	1	3	3
Physidae	2	1	0	3
Planorbidae	2	1	0 & 3	3
Lymnaeidae	2	1	0	3
Glacidorbidae	0	0	0	3
Aplustridae	1	0	0	0 & 2
Hamineoidae	1	0	0	0
Parhedylidae	1	3	0 & 3	0 & 2 & 3
Acochlidiidae	1	3	3	3
Oxynoidae	1	3	0	0
Plakobranchidae	1	3	0	0
Aplysiidae	1	0	?	0
Tylodinidae	1	3	0	0
Dendrodorididae	1	3	0	0
Chromodorididae	1	3	0	0
Dorididae	1	3	0	0
Clionidae	1	3	0	0
Pyramidellidae	0	0	0	0 & 2
Rissoellidae	0	0	0	0 & 2
Architectonicidae	0	0	0	0

CHAPTER 4:

Phylogenetic relationships of the Australian carnivorous land snails (Rhytididae: Stylommatophora) using exon capture

4.1 ABSTRACT

Australia has the highest taxonomic diversity of rhytidids, a family of carnivorous land snails with a Gondwanan distribution. Previous higher classifications of the Australian Rhytididae are based on limited morphological characters and have not been assessed with molecular evidence. I present a molecular phylogeny of the Australian Rhytididae based on a large multi-locus dataset comprising nuclear exons sequenced using exon capture. Using both Bayesian and maximum likelihood analyses I identified four monophyletic lineages within the Australian Rhytididae, as well as an unresolved group of southern temperate lineages. I also show that there is unrecognised diversity across the Australian rhytidids, particularly in the smaller rhytidids. Contrary to shell morphology, on which the current taxonomy is based, a number of currently recognised genera are shown to be either polyphyletic or paraphyletic. The Australian Rhytididae all resulted from an apparent pulse of diversification approx. 45-30 Ma. Given that the South African *Nata* and the New Zealand *Delos* and *Schizoglossa* also belong to this clade, this date suggests that cross-water dispersal has played a major role in the evolution of this group.

4.2 INTRODUCTION

The Rhytididae are a family of carnivorous land snail (Stylommatophora) with a Gondwanan distribution. They are found in South Africa, Australia, New Zealand, Papua New Guinea, and some Pacific islands, however, the centre of taxonomic diversity is Australia. The most recent study to address the species level taxonomy of the Australian Rhytididae (Stanisic et al., 2010) described 60 new species and 15 new genera based on shell morphology, bringing the total number of described species to 78 (25 genera) compared to 19 described species (5 genera) in South Africa (Herbert and Moussalli, 2010, 2016) and 33 described species (10 genera) in New Zealand (Spencer et al., 2006). Phylogenetic studies based on mitochondrial markers and/or the nuclear gene 28S have greatly advanced and revised the taxonomy within the major South African (Moussalli and Herbert, 2016; Moussalli et al., 2009) and New Zealand (Efford et al., 2002; Spencer et al., 2006) lineages but no study has examined the phylogenetic relationships of the Australian Rhytididae in any detail. Unanswered questions therefore remain regarding the evolutionary relationships within the Australian lineages, including the pattern and timing of diversification.

Based on shell morphology, including shell sculpture and size – the Australian Rhytididae range from minute (2mm) to large (45mm) – Solem (1959) suggested that the Australian rhytidids formed several major groups that included both large and small Rhytididae from Australia, New Zealand, and the Pacific islands; although he noted that these groups may not represent phylogeny. Typically the larger Australian Rhytididae, including the genera *Austrorhytida, Tasmaphena, Strangesta*, and *Murphitella* have been considered closely related, as have the smaller genera such as the *Montidelos*, *Echotrida*, and *Saladelos* (Climo, 1977; Iredale, 1933; Smith, 1979; Solem, 1959). The large genus *Victaphanta*, found in Tasmania and Victoria, have been considered distinct from other Australian species and potentially related to the New Zealand *Powelliphanta* as both feature a highly proteinous shell (Solem, 1959). Subsequent classifications of the Australian Rhytididae have largely upheld these broad groupings to varying degrees (Climo, 1977; Iredale, 1933; Solem 1959; Smith 1987) but the higher classification of the Australian genera has not been assessed with molecular evidence. In addition, the new genera described by Stanisic et al. (2010), based on more thorough sampling and shell morphology, have not been assessed with molecular phylogenetic analyses.

The Australian Rhytididae are key to our understanding of the pattern of diversification of the Rhytididae as a whole, given the high species diversity within Australia and the possibility that the Australian Rhytididae are not monophyletic. The only study to address the age of the Australian Rhytididae (Moussalli and Herbert, 2016) had limited taxonomic sampling and relied on a biogeographic calibration, the date that Africa split from east Gondwana (~120 Ma; Chatterjee and Scotese, 1999). Assuming this biogeographic calibration, the study suggested that the Australian lineages split from the African *Nata* approximately 100 Ma (Moussalli and Herbert, 2016). A recent phylogenomic study, which estimated a fossil calibrated phylogeny of Panpulmonata, included a limited number of Australian and South African rhytidids and suggested later dates for the diversification of the Rhytididae (Chapter 4). Understanding of the timing and pattern of evolutionary relationships within putatively Gondwanan lineages is essential to assess the role of vicariance verses dispersal in the diversification of the Rhytididae.

Here I present a molecular phylogeny of the Australian Rhytididae with the aim of identifying major lineages, assessing the current generic classification, and to explore the biogeographic pattern and timing of diversification across eastern Australia. By sequencing new transcriptomes and utilising those already available (Teasdale et al. 2016; Chapter 3), I designed an exon capture probe set that allows the efficient capture and sequencing of large multi-locus nuclear datasets from both fresh and museum tissue, preserved in ethanol. I targeted 500 genes qualified as orthologous for the Eupulmonata (Teasdale et al. 2016) and additional 325 genes identified for the Rhytididae using automated methods of orthology determination herein. I also utilised data from a recent pulmonate phylogenomics study to provide a backbone for dating and calibrating the tree (Chapter 3).

4.3 METHODS AND MATERIALS

4.3.1 Orthology identification and Probe design

I designed exon capture probes to target 825 nuclear genes across the family Rhytididae. Exon capture probes can only tolerate up to 12% sequence divergence (Hugall *et al.* 2015 but see the protocol used in Li *et al.* 2013) therefore baits were designed from ten representative taxa which collectively spaned the most divergent lineages within the family including New Zealand and South Africa taxa. In addition to seven previously sequenced transcriptomes (Teasdale et al. 2016, Chapter 3), I sequenced three additional species, *Montidelos exiguus* from Australia, and *Schizoglossa sp.* and *Delos sp.* from New Zealand (Table 4.1). These additional transcriptomes were sequenced and assembled using the same protocol detailed in Teasdale et al. (2016). Briefly, RNA was extracted from tissue preserved in RNAlater using the RNeasy extraction kit (Qiagen). Libraries were constructed using the TruSeq RNA library preparation kits (Illumina) and sequenced using the Illumina HiSeq 2000 sequencing platform (100 bp, paired end reads). Poor quality sequences and adaptors were trimmed from the reads using Trimmomatic (Bolger et al., 2014) and the transcriptomes were assembled using Trinity (r2013-08-14; Grabherr et al., 2011; Haas et al., 2013).

The orthology of the 500 gene set used in this study had been qualified by Teasdale et al. (2016). In this same study an additional 375 genes were identified as potentially orthologous and single copy across Eupulmonata using the automated pipeline Agalma (Dunn et al., 2013; Teasdale et al., 2016). Given I sequenced lineages not considered in the previous studies, I decided to rerun the Agalma pipeline, with just the 10 rhytidid transcriptomes. To identify contigs representing the 500 genes from Teasdale et al (2016) I used Blastx (Blast+;

Camacho et al., 2009) to compare each transcriptome to the owl limpet genome (*L. gigantea*) predicted gene models. All contigs with a blast match (e-value cut off of e-10) with the relevant genes were trimmed of the untranslated regions (UTRs), placed in the correct reading frame, and then collated into a single fasta file per gene using a custom python script (https://github.com/lteasdale/pullexons_EC). Each gene was then aligned using the L-INS-i method in MAFFT (Katoh and Standley, 2013). For cases where the target gene was fragmented and represented by multiple assembled contigs, sequences were merged into a single final consensus sequence. In such cases, overlapping fragments would only be merged if the proportion of mismatch was <2%. If no overlap existed between fragments, a single, final sequence would only be created if there were no other competing contigs (custom python script: https://github.com/lteasdale/consensus_maker_LT). Each alignment was then visually assessed to check the quality of the alignment, to identify any potentially paralogous sequences, and create additional consensus sequences where warranted. Sequences for an additional 325 genes were identified by analysing the 10 transcriptomes with the fully automated orthology determination pipeline Agalma (Dunn et al. 2013) using default settings (from the 'postassemble' step). From this analysis I only retained orthologous clusters which were represented in at least 9 of the 10 transcriptomes and when no other orthologous clusters were produced from the respective homologous cluster. The 500 genes from Teasdale et al. (2016) were delineated into 2,294 exons based on the *L. gigantea* genome exon boundaries using Exonerate (Slater and Birney, 2005). The sequences for the 325 additional genes were kept whole as a previous study showed that novel exon boundaries did not affect capture efficiency unless the exons were particularly small (<40 bp; Teasdale et al., 2016).

To increase the chance of capturing sequence from lineages not represented in the transcriptome dataset I reconstructed the ancestral state sequences for the 825 genes (see Hugall et al., 2016). Marginal ancestral reconstructions were conducted using the program FastML (Ashkenazy et al., 2012) and a guide tree constructed using the concatenated alignment for all 825 genes in RAxML (Stamatakis, 2014; partitioned by codon position, GTR+Γ, 100 fast bootstraps). I included multiple copies of each exon in the probe design to maximise sequence capture across the Rhytididae. When taking into account pairwise distances for each exon (Figure 4.1 a; https://github.com/lteasdale/p-distance_script) and the ~12% probe mismatch threshold I decided to include six copies of each target exon in the probe design. Three of the copies were ancestral state reconstructions: 1) the common ancestor of *Chlamydephorus gibbonsi* (which is nested withint the Rhytididae (Moussalli and

Herbert, 2016)) and *Natalina cafra natalensis*, 2) the common ancestor of the Australian and New Zealand taxa sequenced, as well as *Nata vernicosa*, and 3) the common ancestor between all ten taxa (Figure 4.1 b). The other three copies were sequences for the species 4) *Nata vernicosa*, 5) *Austrorhytida lamproides*, and 6) *Chlamydephorus gibbonsi*. These taxa were chosen as they represented the tips of the major rhytidid clades and had high sequence coverage of the exons. Where exons were missing from *C. gibbonsi* or *A. lamproides*, I included sequences for *N. cafra natalensis* or any of the other Australian and New Zealand taxa respectively.

Probes for the target sequences were designed and manufactured by MYcroarray (Ann Arbor, Michigan) using the custom MYbaits Target Enrichment kit (biotinylated 120 bp RNA baits at 2X tiling). As the probes are 120 bp, I excluded all exons less than 100 bp from the design. All exons between 100 bp and 120 bp were padded out with T's to ensure a 120 bp probe was constructed. The final probe design targeted 2,483 exons representing 687,621 bp.

4.3.2 Tissue extractions and sequencing

I extracted DNA from museum and freshly collected tissue preserved in ethanol for 115 samples representing 62 of the 78 currently recognised Australian species of Rhytididae, and 25 of the 29 known Australian genera (Table 1). DNA extractions were performed using the standard tissue protocol for the DNeasy Blood and Tissue extraction kit (QIAGEN). I quantified the resulting DNA using the Qubit fluorometer (Invitrogen) and the QIAxpert spectrophotometer system (QIAGEN). DNA library preparations were conducted using a modified version of the NEBNext Ultra DNA Library Prep kit for Illumina sequencing with additional PCRs for samples with low DNA concentration as needed. Up to 12 libraries were pooled per capture, and hybridised to the baits (at one-quarter dilution) for 36 hours, following the MYbait protocol v3. The captured fragments for all samples were sequenced on a half lane of the HiSeq 2500 Illumina sequencing platform (125 bp, paired end reads).

4.3.3 Exon capture data analysis

Duplicate reads, which may represent PCR duplicates, were removed from each sample using the program FastUniq v1.1 (Xu et al., 2012). Low quality sequence and adaptors were trimmed from the reads using Trimmomatic (Bolger et al., 2014). For read mapping I produced a sample specific reference for each exon by first merging and assembling the reads using BBMerge and Tadpole (https://sourceforge.net/projects/bbmap), with a kmer size of 130, respectively. I then identified contigs within each assembly that matched the target exons using tblastx (blast+; Camacho et al., 2009). I selected the contig with the best e-value match for each target exon (an e-value of at least e-10), and removed the UTRs to produce a sample specific reference for read mapping (custom python script: https://github.com/lteasdale/pullexons_EC.py).

The reads for each sample were then mapped to the respective sample specific exon reference using BBMap (https://sourceforge.net/projects/bbmap). The initial BAM alignments showed that exon boundaries were often unconserved between *L. gigantea* and the Rhytididae, as was the case for a previous exon capture set for the Camaenidae (Teasdale et al., 2016). Where novel exon boundaries were present the reference exons were split to reflect the actual exon boundaries within the Rhytididae. The reads were then remapped to the revised sample specific exon references. I called variants from the resulting BAM files using 'HaplotypeCaller' in GATK with the contamination fraction to filter set at 0.1 (McKenna et al., 2010). I then produced a consensus sequence for each exon, per sample, using 'consensus' in BCFtools v1.3.1 (Li, 2011) with all sites with coverage <5 masked (custom python script: https://github.com/lteasdale/mask_low_cov). The consensus sequences, the respective sequences for the 10 Rhytididae transcriptomes, and the transcriptome sequences for one outgroup, *Trigonephorus ambiguosus* (Chapter 4), were collated into fasta files per exon using a custom python script (https://github.com/lteasdale/fasta_formatter_general). The exons were then aligned using the frameshift aware alignment program MACSE (Ranwez et al., 2011).

Once aligned, sequences which contained <30% of the exon or comprised more than three percent ambiguous sites were removed from the alignments (custom python script: https://github.com/lteasdale/ambig_counter). Ambiguously aligned regions of the alignments were removed using GBLOCKS (Castresana, 2000) with codon information retained. The final data matrix contained 4,126 exons, but was relatively sparse (57% complete). I therefore produced a second matrix that only contained exons represented by ≥ 80% of the samples ('highly complete matrix', 1,276 exons, 84% complete, 185,388 bp). Finally, to assess alignment quality I calculated the number of variable sites per exon using AMAS (Borowiec, 2016) and the average p-distance per exon (Capella-Gutiérrez et al., 2009), and visually assessed the alignments for errors. Using the program BaCoCa (Kück and Struck, 2014) I also assessed the phylogenetic utility of the final data matrix by calculating levels of saturation (C-value) and deviation from sequence homogeneity (χ^2 test of sequence homogeneity) per exon.

4.3.4 Phylogenetic and dating analyses

To reconstruct the phylogenetic relationships within the Rhytididae I conducted both maximum likelihood and Bayesian analyses. I partitioned the nucleotide alignment by codon position. For the amino acid analyses I ran a single partition analysis and a partitioning scheme where the exons were clustered into seven partitions based on amino acid frequencies (calculated in BaCoCa) using Wards Hierarchical Clustering and the Kelley-Gardener-Sutcliffe penalty function (Kelley et al., 1996; calculated using the maptree library in R). For each partitioning scheme I selected the suitable substitution model for each partition using PartitionFinder (Lanfear et al., 2012), with RAxML and BIC model selection. PartitionFinder selected GTR+I+Γ as the best substitution model for each codon position, however, as I could not run invariant site estimation in the Bayesian analyses, I ran two maximum likelihood analyses in RAxML using the 84% complete data matrix for comparison, with the models GTR+Γ and GTR+I+Γ assigned to each partition respectively. I also produced a maximum likelihood phylogeny for the sparse nucleotide matrix, containing all captured exons, with the GTR+Γ model assigned to each partition. Both the single partition and seven exon partition amino acid analyses were conducted with the amino acid substitution model JTT assigned to each partition.

I conducted a Bayesian phylogenetic analysis using the program ExaBayes (Aberer et al., 2014). Using the highly complete nucleotide matrix, partitioned by codon position, I ran four Metropolis-coupled ExaBayes replicates for 2 million generations, each with four chains (three heated), and sampling every 1,000 generations. The GTR+Γ model was assigned to each partition as invariant site estimation is not yet implemented in ExaBayes. Using the 'postProcParam' tool included with the ExaBayes package, I checked for convergence and adequate sampling of the posterior distribution of the parameter values by ensuring that the effective sample sizes (ESS) of all estimated parameters were greater than 200 and that the average standard deviation of split frequencies and potential scale reduction factors across runs were close to zero and one, respectively. A consensus tree was created by combining the trees from the four separate runs, with the first 25% removed as burn in in each case, using the 'consensus' tool included with the ExaBayes package.

The dating analysis was conducted using the Bayesian program BEAST (Drummond et al., 2012). As the oldest known fossil rhytidid only dates from the Pliocene (~5 Ma; Tracey et al., 1993) I used a secondary calibration derived from a previous study (Chapter 4) that produced a fossil calibrated phylogeny for Panpulmonata, which included four representatives of the Rhytididae and *Chlamydephorus gibbonsi*. Specifically, I used the mean node age, 75.31 Ma, and associated confidence intervals estimated for the split between the clade containing the South African lineages *Chlamydephorus gibbonsi* and *Natalina cafra natalensis,* and the Australian taxa. The analysis was conducted using the highly complete matrix, partitioned by codon position (GTR+Γ assigned to each), using the lognormal relaxed clock model and a Birth-Death tree prior. I used the tree produced from the maximum likelihood nucleotide analysis, with each codon assigned GTR + Γ, as a starting tree. I ran two independent beast analyses for 80 million generations each and convergence was assessed by ensuring that the ESS values were >200 for all parameters using the program Tracer v1.6 (http://beast.bio.ed.ac.uk/Tracer).

4.4 RESULTS

4.4.1 Sequence capture

Of the 2,483 targeted exons I captured sequence for 2,414 across the 115 samples. Novel exon boundaries were present in 604 of the captured exons. Splitting up the exons with novel exons boundaries brought the total number of exons captured to 4,126 exons. Of the 325 genes for which I did not delineate exon boundaries, 277 had novel exon boundaries. I was able to successfully capture sequence from old museum preserved specimens (>25 years old, ethanol preserved; Figure 4.2 a, b). Capture success was more dependent on the amount of starting DNA used for the libraries than the age of the specimen (Figure 4.2 a, b). I removed 1,552 individual sequences (0.005%) due to high levels of heterozygosity (>3% of heterozygous sites). The final data matrix contained 4,126 exons, but was relatively sparse (57% complete). Removing exons represented by < 80% of the samples resulted in an alignment which was 84% complete (1,276 exons, 185,388 bp). There was no evidence of sequence saturation for any of the exons; the smallest C-value was 1,477 (median = 9938) with c-values near zero indicating saturation (Struck et al., 2008). There was also no evidence of significant biases in sequence homogeneity across clades, for each exon, based on the χ^2 test of sequence homogeneity.

4.4.2 Phylogenetic analysis

The deep relationships within the Australian rhytidids are characterised by short internal branch lengths and uncertainty. The Bayesian analysis shows strong support for most of the relationships between the Australian lineages (Figure 4.4). However, the nucleotide maximum likelihood analysis, while topologically consistent, shows no resolution of the relationships between the Australian lineages (Figure 4.4). The amino acid analyses also show a lack of resolution for these relationships although there is some support for *Torresiropa* being basal (BS = 90-93; Appendix 4.3, 4.4). Here I only consider relationships with at least 75% bootstrap support and a posterior probability of >0.95 as supported. Given the lack of basal resolution and low taxonomic sampling of South African and the New Zealand lineages I cannot confirm that the Australian taxa are not monophyletic. The analyses do, however, show that the African Rhytididae are not monophyletic as the South African genera *Natalina* and *Chlamydephorus* form a clade that is sister to the clade comprising all the Australian lineages, the NZ samples, and the South African genus *Nata*.

Despite this early period of relatively rapid cladogenesis, the majority of the Australian lineages form four well supported clades that are consistent across analyses. To aid discussion these clades are labelled in Figure 4.4. Clade 1 is the largest, comprising nine genera of larger rhytidids (Figure 4.4), and is distributed along the eastern coast of Australia from the Atherton tablelands in far north Queensland to Tasmania and the Flinders Ranges (Figure 4.3). Clade 2 contains six genera of smaller rhytidids (Figure 4.4) and is distributed across the eastern coast of Queensland (Figure 4.3). Clade 3 also comprises smaller rhytidids (two genera) but is confined to north eastern NSW. Clade 4 contains three genera including both large and small rhytidids that are distributed across the eastern coast of Queensland. The remaining Australian lineages are phylogenetically diverse but not highly speciose, and are concentrated in southern Australia (apart from *Torresiropa* which is found on the tip of Cape York). I refer to these southern lineages as the ‘southern temperate group’ although the relationships between these lineages are unresolved.

The relationships within the four monophyletic clades are consistent across all analyses, with most nodes fully supported by both the Bayesian and maximum likelihood (ML) analyses. Only a few relationships within these clades have complete posterior probability support but are not supported in the ML analyses, namely: the placement of *Austrorhytida warrumbunglensis*, and the relationships within the genus *Pseudechotrida,* and within the species *Montidelos exiguus*, and *Terrycarlessia bullacea*. Several currently recognised genera are well supported as paraphyletic: *Austrorhytida*, *Briansmithia*, *Montidelos*, *Murphitella* and

Terrycarlessia, or polyphyletic: *Annabellia*, *Griffithsina*, *Prolesophanta*, *Protorugosa*, *Scagacola*, *Saladelos*, *Victaphanta*, and *Vitellidelos*, with several genera split across the four clades.

The relationships within the Australian Rhytididae estimated by the BEAST analysis are broadly consistent with the phylogenetic estimate obtained in Exabayes and RAxML (Figure 4.5). Constraining the date for the split between the South African *Natalina* and *Chlamydephorus*, and the rest of the Rhytididae at 75 million years resulted in estimates of mean substitution rates for each codon position of 0.56%, 0.38%, 2.05% per lineage per million years respectively. The South African *Nata*, the New Zealand and Australian lineages emerged in an apparent pulse of diversification between 45-30 Ma (Figure 4.5). The crown age of the four monophyletic Australian clades ranges from Clade 4 approximately 30 Ma, to Clade 1 which has a crown age of approx. 20 Ma.

4.5 DISCUSSION

Here I present the first detailed molecular phylogenetic study of the Australian Rhytididae. The phylogenetic reconstruction shows an early period of rapid cladogenesis, with most basal relationships remaining unresolved. Nevertheless, the majority of Australian rhytidids fall within four distinct well supported clades, which all appear to have emerged during the late Eocene. As suggested by shell morphology, I confirm that most of the larger Rhytididae are closely related (Clade 1), and that the *Victaphanta* belong to a separate lineage. However, the analyses support a number of unpredicted relationships. While the *Murphitella,* a genus of large snails from Far North Queensland, have been considered distinct (Smith, 1979; Solem, 1959), they were still thought to be related to the other large snails found in Queensland and New South Wales, including *Strangesta* and *Austrorhytida* (Solem 1959; Smith 1979). However, I show that *Murphitella* belongs to a different highly divergent clade (Clade 5) and is more closely related to the smaller *Echotrida* and *Saladelos commixta*. Despite the recent description of many new genera of small Rhytididae by Stanisic et al. (2010) the results also show that there is additional unrecognised diversity in the smaller Australian Rhytididae (Stanisic et al., 2010). Three genera of smaller snails, *Saladelos*, *Montidelos*, and *Echotrida*, have been regarded as belonging to the one genus in the past (Smith, 1992). The present results, however, show that these three genera belong to four separate major clades. In addition to these broad differences several currently recognised genera (Stanisic et al., 2010) are not monophyletic and I discuss the taxonomic implications of these results below. I also preliminarily discuss the biogeography of the Australian clades and the timing of diversification in relation to the breakup of Gondwana.

4.5.1 Taxonomic implications

4.5.1.1 Clade 1

Thirteen of the currently recognised Australian genera are either paraphyletic or polyphyletic. Most of the relationships that differ from the current taxonomy are not reflected in the shell morphology, on which the current taxonomy is based. Clade 1 contains three paraphyletic genera: *Austrorhytida* Smith, 1987, *Terrycarlessia* Stanisic, 2010, and *Briansmithia* Stanisic, 2010. The genus *Austrorhytida* is paraphyletic as it includes two species currently designated to the genera *Annabellia* Shea & Griffiths, 2010, and *Protorugosa*, Shea & Griffiths, 2010, respectively. *Annabellia occidentalis*, which occurs west of the Blue Mountains, is nested within *Austrorhytida capillacea* and is closely related to *A. capillacea* samples from the Blue Mountains area. While still in Clade 1, *Annabellia sensu stricto* (type species is *Annabellia bingara*), is highly distinct from the *Austrorhytida* and is closely related to the genera *Emmalena* Stanisic, 2017, from the Flinders ranges, and *Strangesta* Iredale, 1933. The two *A. occidentalis* specimens sequenced in this study were collected from the type locality, west of the Blue Mountains near Black Spring, however, I could not distinguish them from *A. capillacea* based on shell morphology. It is therefore possible that *Annabellia occidentalis* exists; however, more detailed sampling across the range of *A. occidentalis* is needed to confirm whether *A. occidentalis* is a population of *A. capillacea*.

Within the *Austrorhytida*, the species *A. glaciamans* appears to be polyphyletic. We sequenced eight specimens of *A. glaciamans*, one from the type locality at Wilsons Valley near Mt Kosciusko, three from eastern and alpine Victoria, and four from the Otways in Western Victoria. The main differences in morphology used to distinguish *A. glaciamans* from *A. capillacea* are smaller size and a body with orange and cream sides (Stanisic et al., 2010). These characteristics were exhibited by the *A. glaciamans* specimen collected from the type locality, however, this specimen is nested within the *A. capillacea* and is most closely related to the *A. capillacea* samples from the Blue Mountains. The three samples from eastern and alpine Victoria, which included a sample from a locality in the Victorian alps just 70km south east of Wilsons Valley, formed a distinct clade more closely related to the Victorian *Austrorhytida*. These results suggest that *A. glaciamans* is a synonym of *A. capillacea*, and that the eastern Victorian specimens represent a new species, however, I would advocate more detailed sampling of the alpine *Austrorhytida* to assess the range extent of the two species. In addition, *Austrorhytida glaciamans* specimens from the Otways in south western Victoria are highly divergent from the specimens collected from eastern Victoria, potentially representing an additional species.

Protorugosa Shea & Griffiths, 2010, is also polyphyletic as the species *Protorugosa burraga* is nested within the *Austrorhytida*. The two samples I sequenced are from localities on opposite sides of the Barrington Tops National Park: Burraga Swamp and Gloucester Tops. The monophyletic *P. burraga* are most closely related to *Austrorhytida barringtonia*, the sample of which was also collected from Burraga swamp. While closely related, the two species are morphologically distinct: *P. burraga* has a dark body and a shell with coarse radial ribs whereas *A. barringtonia* has a pale body and a shell with finer radial ribs and distinctive red radial streaks. It is therefore possible that these two lineages represent distinct species that have come back into contact after diverging in allopatry. More sampling across the distributions of these taxa is needed to determine whether they are separate species or an example of morphological polymorphism. The only other species in the genus *Protorugosa*, *P. alpica*, is reciprocally monophyletic and is shown to have a sister relationship with the *Austrorhytida* clade. As *P. burraga* is the type species of the genus, *P. alpica* should either be included in *Austrorhytida* or represent a new genus, depending on detailed morphological analyses.

The second paraphyletic genus in Clade 1 is *Terrycarlessia* Stanisic, 2010. Due to differences in shell size and shape *Griffithsina brisbanica* was only tentatively placed within the genus *Griffithsina* Stanisic, 2010 (Stanisic et al., 2010). My results show that *G. brisbanica* is in fact nested within the *Terrycarlessia*. *Scagacola eddiei*, the most southerly distributed of the species within *Scagacola* Stanisic, 2010, is also nested within *Terrycarlessia*. The other representatives of *Scagacola* and *Griffithsina* are found within the third paraphyletic genus within Clade 1, *Briansmithia* Stanisic, 2010. There is a general lack of resolution within the '*Briansmithia*' clade, possibly representing a species complex, but a more detailed phylogeographic study is need to investigate the relationships further. The species *Terrycarlessia bullacea* is also shown to be polyphyletic. The *T. bullacea* sample from Queensland is highly divergent from a monophyletic clade which contains the NSW *T. bullacea* samples. As the Queensland specimen is near the type locality of the species it is likely that the NSW lineage represents a new species.

4.5.1.2 Clade 3

Clade 3 comprises two polyphyletic genera, *Montidelos* Iredale, 1943 and *Vitellidelos* Stanisic, 2010. Both genera contain small snails but they have been differentiated by several shell characters including the strength of the radial and spiral sculpture on the shells, and the width of the umbilicus (Stanisic et al., 2010). My results, however, show that Clade 3 contains three highly divergent lineages, two of which contain representatives of both *Vitellidelos* and *Montidelos*. Complicating the situation further, two additional species of *Vitellidelos*, namely *V. helmsiana* and a yet undescribed species from Tasmania, *V. sp* L178, represent highly distinct and divergent lineages. Accordingly, the 'Vitellidelos form' appears to have evolved multiple times, and may be ecologically driven as many currently designated species occur either at high altitude or in cool temperate environments. As the type species for these two genera are *M. orcadis* Iredale 1943 and *V. dulcis* (Iredale 1943) respectively, we tentatively recommend both *V. kaputarensis* and *V. dorrigoensis* be placed in *Montidelos*. Further anatomical work and more comprehensive sampling is necessary to determine whether *Montidelos* should be further split, and to formally recognise *V. helmsiana* and *V. sp* L178 as distinct new genera.

4.5.1.3 Clade 2 and 4

Clade 4 is an interesting group because it comprises both small and large rhytidids, ranging from the very small *Echotrida globosa* (7mm), to the large *Murphitella franklandiensis* (33mm), with *Echotrida* Iredale, 1933, nested within a paraphyletic *Murphitella* Iredale, 1933. This clade therefore represents an important study system for future investigations of morphological evolution in the Rhytididae. The only taxonomic implication arising from this clade is that *Saladelos* Iredale, 1933, is polyphyletic. There are currently two nominal species, *S. commixta*, which is found throughout far north Queensland, and *S. lacertina*, which is restricted to Lizard Island. My results show that *S. commixta* is most closely related to *Echotrida* and *Murphitella* in Clade 4, whereas *S. lacertina* is more closely related to the genus *Umbilidelos,* in Clade 2. As the type species is S. *commixta*, morphological analyses need to be conducted to determine whether *S. lacertina* belongs to the genus *Umbilidelos* or represents a separate genus. Clade 2 is also found in Queensland and includes many highly restricted lineages, but it only contains small rhytidids.

4.5.1.4 The southern temperate group

The ‘southern temperate group’ represents a collection of morphologically and phylogenetically distinct lineages which are generally species poor. While phylogenetic relationships among these deep lineages remain unresolved, most likely due to the early rapid cladogenesis, southern and particularly south-eastern Australia appears to be an important centre of phylogenetic diversity and endemism. Ranging from the minute *Prolesophanta* (2.5 mm) to the relatively large cool temperate rainforest restricted genus *Victaphanta,* most of these southern lineages geographically restricted. An exception is *Prolesophanta dyeri,* however, which is relatively widespread in wet forests of south-eastern Australia, albeit at very low density. The molecular phylogeny presented here further adds to this diversity by identifying *Prolesophanta nelsonensis* and *Victaphanta lampra* as being highly distinct and divergent form other species in their respective genera. The Victorian *Victaphanta* Iredale 1933, *V. atramentaria* and *V. compacta,* and the Tasmanian *V. lampra* form separate highly divergent monophyletic clades. As I have not sequenced *V. milligani*, which is Tasmanian but morphologically more similar to the Victorian *Victaphanta*, it is unclear whether the Tasmanian *Victaphanta* are monophyletic. Given *V. atramentaria* is the type species of the genus, *V. lampra* likely represents a new genus, although detailed morphological comparisons are needed.

The Tasmanian *Prolesophanta nelsonensis*, another of the highly divergent southern lineages, represents a completely different lineage to *Prolesophanta s.s.* (type species is *P. dyeri*). This finding is supported by morphology as *P. nelsonensis* has a widely open umbilicus whereas *Prolesophanta s.s.* is the only clade of Rhytididae which does not have an umbilicus. *Prolesophanta s.s.* occurs on Tasmania in the form of *Prolesophanta dyeri*, however, my results show that the Victorian *P. dyeri* is more closely related to *P. occlusa* which was collected from the Blue Mountains in NSW. As the type species for *Prolesophanta* is *P. dyeri*, and the type specimen for *P. dyeri* is from Tasmania, it is possible that the Victorian *Prolesophanta* represents a new species. It is also possible that the Victorian *P. dyeri* may represent populations of *P. occlusa*. Both *Prolesophanta* lineages are rare and likely harbour additional undetected diversity.

4.5.2 Biogeography

The Australian Rhytididae are represented by multiple distinct but often sympatric lineages. The Australian Rhytididae are distributed throughout the mesic environments of eastern of Australia including Tasmania, and semi-arid Flinders ranges in South Australia.

The only exception is the undescribed rhytidid from the Stirling Ranges, in the south western corner of Australia, which is the sole representative of Clade 6. A second species from western Australia, *Occirhenea georgiana*, is extremely rare and is presumed to be extinct: the last specimen collected in 1955 (Kendrick et al. 1971). On the eastern coast there are several key areas of high diversity. While it is common to find at least two different lineages at the one locality (typically a small and large rhytidid), a particularly hotspot for sympatric taxonomic diversity is northern NSW and southern Queensland. All four of the monophyletic clades occur in this region, as well as at least four distinct lineages within Clade 1 alone. It is likely that differences in body size, which determine potential prey, allows for this sympatry.

While not taxonomically diverse, south eastern Australia (i.e. the southern temperate group) and far north Queensland also contain unrecognised phylogenetic diversity. There are very few examples of deep relictual diversity in south eastern Australia, however, another example is seen in putatively Gondwanan assassin spiders (Rix and Harvey, 2012). *Vitellidelos* in particular represents a high degree of unrecognised diversity that was not evident in the shell morphology. Deep cryptic diversity has been shown within *Nata*, the south African genus of dwarf carnivorous rhytidids, which also show little morphological variation (Moussalli and Herbert, 2016). Lineages with unrecognised deep diversity also occur in far north Queensland, including 1) the *Torresiropa* (Clade 8), which is found on the tip of Cape York and may be related to the *Ouagapia* in Papua New Guinea, 2) *Saladelos lacertina* and the genus *Umbilidelos* in Clade 2, and 3) the *Murphitella* and *Saladelos commixta* from Clade 5. Several of these deep lineages are only represented by one or two species, many with narrow distributions, thus conservation risk assessment may find that a number of taxa are at risk from threats such as land clearing. There are only three extant Australian Rhytididae currently listed on the ICUN red list. Two of these are southern temperate species, *Victaphanta compacta* and *V. atramentaria*, and the third is *Austrorhytida lamproides* (Clade 1).

4.5.3 Gondwanan connections

All major lineages of the Australian Rhytididae emerged in an apparent pulse of diversification approximately 45-30 Ma. This period, representing the latter half of the Eocene, was a time of climatic change in Australia's history. The last connection between Australia and Antarctica, via Tasmania, flooded approximately 45 million years ago (Lawver and Gahagan, 2003). The opening of the southern ocean led to cooling of the region that in turn led to major changes in climate and the contraction of rainforests on Australia (Byrne et al., 2011). The diversification of the Australian lineages coincides with the separation of Australia and Antarctica, and the establishment of the Antarctic Circumpolar Current. However, the drivers of this diversification are not clear. Given the timeframe vicariance resulting from the separation of Antarctica and Australia is a possible hypothesis. However, as there are no extant Rhytididae on Antarctica so this hypothesis cannot be tested unless fossils are found. What appears to be an apparent increase in diversification may actually represent the signature of a mass extinction event that occurred in the Eocene. Diversification theory and simulations suggest that so called ‘broom handle’ phylogenies can be explained by mass extinction without needing to invoke adaptive radiation (Crisp and Cook, 2009). Similar patterns of diversification (see Appendix 4.5), which suggest an Eocene mass extinction event in the Australian region, are seen in geckos (Oliver and Sanders, 2009), bees (Schwarz et al., 2006), and legumes (Crisp and Cook, 2009).

While I have limited representation of non-Australian lineages, the apparent radiation of the Australian Rhytididae also includes representatives of the New Zealand genera *Delos* and *Schizoglossa*, and the South African *Nata*. Both land masses broke away from Gondwana much earlier than Australia: New Zealand ~ 80 Ma, and Africa ~120 Ma (Chatterjee and Scotese, 1999). Given my dating estimates, these dates imply that cross-water dispersal has occurred in the Rhytididae. A number of studies have shown that cross-water dispersal is possible for land snails (Cowie and Holland, 2006; Gittenberger et al., 2006) and the presence of land snails on volcanic islands suggests that cross-water dispersal leading to colonisation has occurred multiple times (Cowie and Holland, 2006). In addition, prior to the establishment of the southern ocean current there were ocean currents which travelled from the pacific past the north of Australia to southern Africa (Lawver and Gahagan, 2003). Detailed sequencing of similar multi-locus phylogenetic datasets for the South African and New Zealand taxa is needed to put the Australian Rhytididae in context. The role of the persistence of the continental pacific islands also needs to be considered as many of the Pacific island Rhytididae are morphologically distinct and may represent ancestral populations from which many Australian and New Zealand lineages are derived (Climo, 1977).

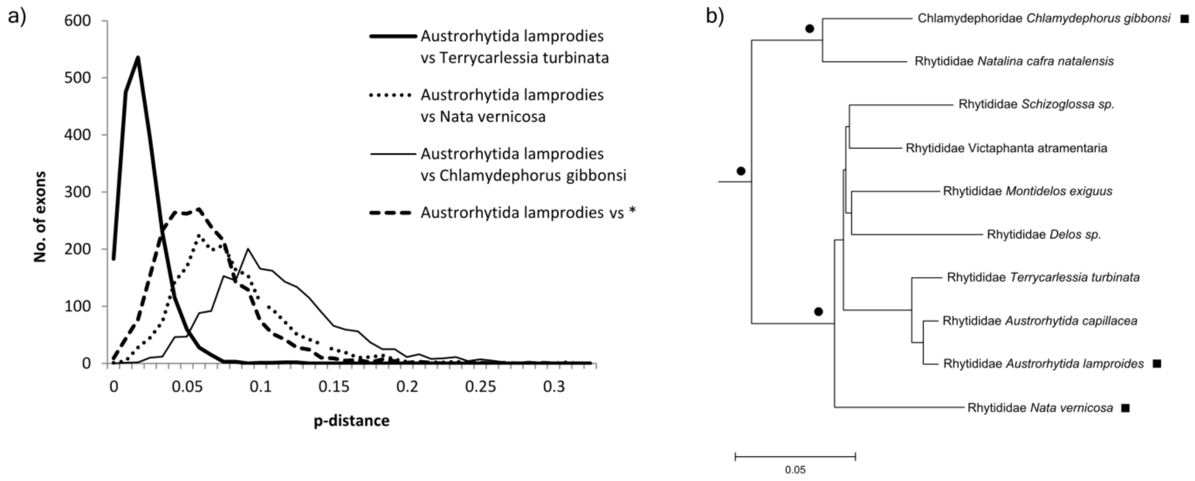


Figure 4.1. Exon capture probe design: a) shows the p-distances between the representative rhytidid transcriptomes used as a reference to construct an exon capture probe design targeting exons from 825 genes, and b) shows the phylogenetic relationships between the 10 reference rhytidid transcriptomes used in the probe design. Six sets of sequences were included in the probe design, the sequences from three species (represented by squares) and three ancestral state sequences (represented by circles).

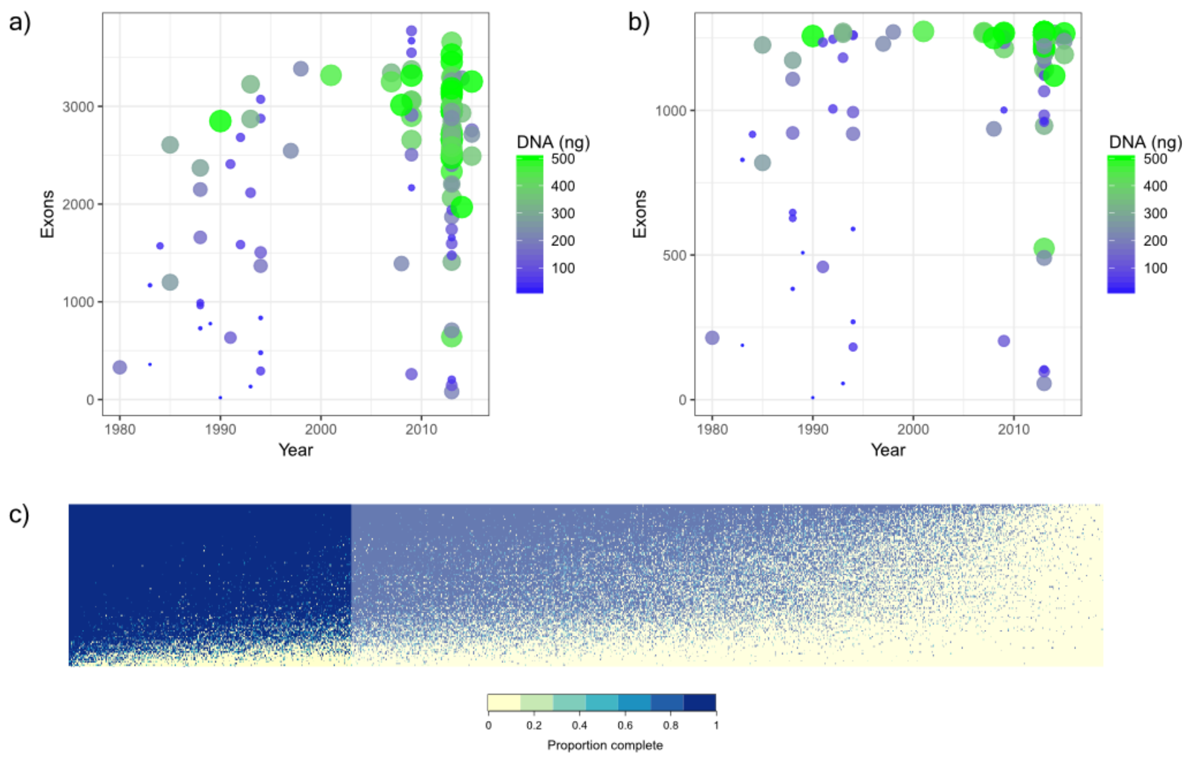


Figure 4.2. Exon capture success rate. a) shows the relationship between the number of exons captured per sample and the year the sample was collected and preserved. The size and colour of each point represents the amount of initial DNA (ng) used in the respective library preparation. In b), the same relationship is shown but for the 1,276 exons with at least 80% of the taxa. The heatmap c) shows the proportion of each exon (left to right) captured per sample (top to bottom). The section of the heat map that is not faded represents the 1,276 with at least 80% of the taxa.

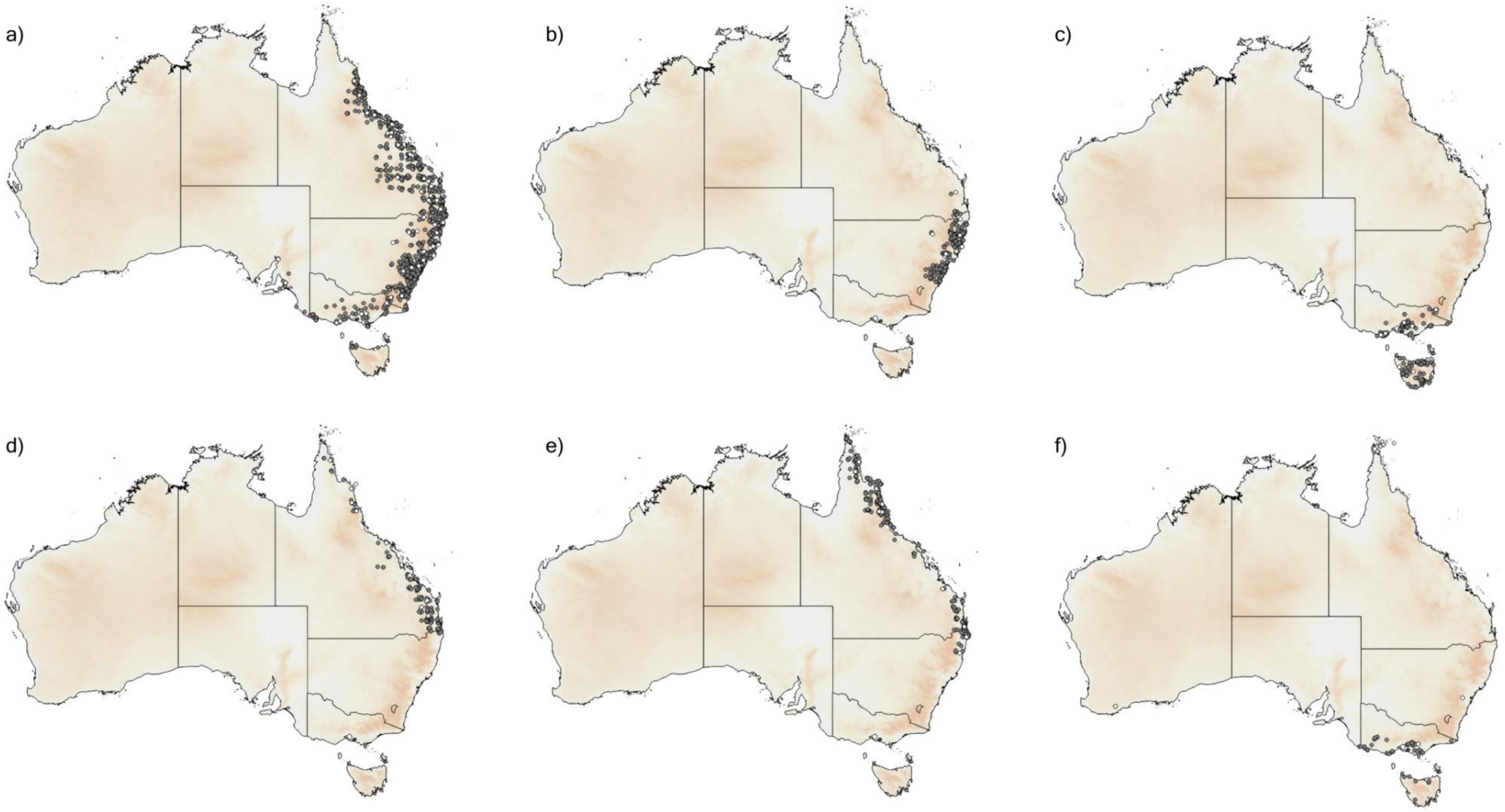
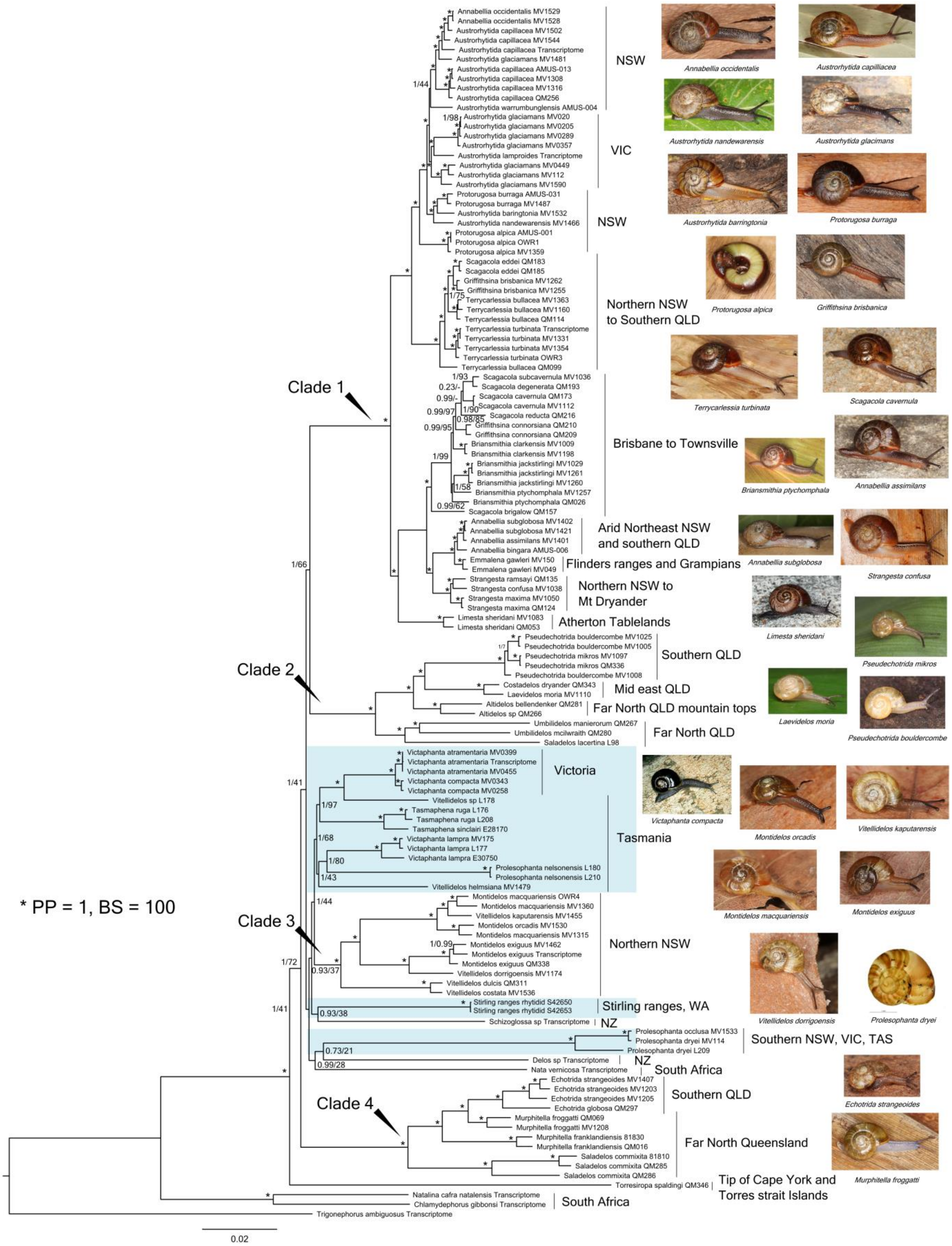


Figure 4.3. Current distribution of the major Australian lineages: a) Clade 1, b) Clade 3, c) Southern temperate lineages including Victaphanta, Tasmaphena, and Vitellidelos, d) Clade 2, e) Clade 4, and f) additional southern temperate lineages including Prolesophanta and the western Australian rhytidid, and Torresiropa, which is found on the tip of Cape York. The grey circles represent museum records and the white circles represent specimens sequenced in this study.



Annabellia occidentalis MV1529
Annabellia occidentalis MV1528
Austrorhytida capillacea MV1502
Austrorhytida capillacea MV1544
Austrorhytida capillacea Transcriptome
Austrorhytida glaciamans MV1481
Austrorhytida capillacea AMUS-013
Austrorhytida capillacea MV1308
Austrorhytida capillacea MV1316
Austrorhytida capillacea QM256
Austrorhytida warrumbunglensis AMUS-004
NSW
1/44
1/98
Austrorhytida glaciamans MV020
Austrorhytida glaciamans MV0205
Austrorhytida glaciamans MV0289
Austrorhytida glaciamans MV0357
Austrorhytida lamproides Trancriptome
Austrorhytida glaciamans MV0449
Austrorhytida glaciamans MV112
Austrorhytida glaciamans MV1590
VIC
Protorugosa burraga AMUS-031
Protorugosa burraga MV1487
Austrorhytida baringtonia MV1532
Austrorhytida nandewarensis MV1466
Protorugosa alpica AMUS-001
Protorugosa alpica OWR1
Protorugosa alpica MV1359
NSW
Scagacola eddei QM183
Scagacola eddei QM185
Griffithsina brisbanica MV1262
Griffithsina brisbanica MV1255
1/75
Terrycarlessia bullacea MV1363
Terrycarlessia bullacea MV1160
Terrycarlessia bullacea QM114
Terrycarlessia turbinata Transcriptome
Terrycarlessia turbinata MV1331
Terrycarlessia turbinata MV1354
Terrycarlessia turbinata OWR3
Terrycarlessia bullacea QM099
Northern NSW to Southern QLD
Clade 1
1/93
0.23/-
0.99/-
0.99/97
1/90
0.98/85
0.99/95
Scagacola subcavernula MV1036
Scagacola degenerata QM193
Scagacola cavernula QM173
Scagacola cavernula MV1112
Scagacola reducta QM216
Griffithsina connorsiana QM210
Griffithsina connorsiana QM209
Briansmithia clarkensis MV1009
Briansmithia clarkensis MV1198
1/99
Briansmithia jackstirlingi MV1029
Briansmithia jackstirlingi MV1261
Briansmithia jackstirlingi MV1260
1/58
Briansmithia ptychomphala MV1257
Briansmithia ptychomphala QM026
0.99/62
Scagacola brigalow QM157
Brisbane to Townsville
Annabellia subglobosa MV1402
Annabellia subglobosa MV1421
Annabellia assimilans MV1401
Annabellia bingara AMUS-006
Arid Northeast NSW and southern QLD
Emmalena gawleri MV150
Emmalena gawleri MV049
Flinders ranges and Grampians
1/66
Strangesta ramsayi QM135
Strangesta confusa MV1038
Strangesta maxima MV1050
Strangesta maxima QM124
Northern NSW to Mt Dryander
Limesta sheridani MV1083
Limesta sheridani QM053
Atherton Tablelands
Pseudechotrida bouldercombe MV1025
1/7
Pseudechotrida bouldercombe MV1005
Pseudechotrida mikros MV1097
Pseudechotrida mikros QM336
Pseudechotrida bouldercombe MV1008
Southern QLD
Clade 2
Costadelos dryander QM343
Laevidelos moria MV1110
Mid east QLD
Altidelos bellendenker QM281
Altidelos sp QM266
Far North QLD mountain tops
Umbilidelos manierorum QM267
Umbilidelos mcilwraith QM280
Saladelos lacertina L98
Far North QLD
1/41
Victaphanta atramentaria MV0399
Victaphanta atramentaria Transcriptome
Victaphanta atramentaria MV0455
Victaphanta compacta MV0343
Victaphanta compacta MV0258
Victoria
Vitellidelos sp L178
1/97
Tasmaphena ruga L176
Tasmaphena ruga L208
Tasmaphena sinclairi E28170
1/68
Victaphanta lampra MV175
Victaphanta lampra L177
Victaphanta lampra E30750
1/80
Prolesophanta nelsonensis L180
Prolesophanta nelsonensis L210
1/43
Vitellidelos helmsiana MV1479
Tasmania
* PP = 1, BS = 100
1/44
Montidelos macquariensis OWR4
Montidelos macquariensis MV1360
Vitellidelos kaputarensis MV1455
Montidelos orcadis MV1530
Montidelos macquariensis MV1315
Clade 3
1/0.99
Montidelos exiguus MV1462
Montidelos exiguus Transcriptome
Montidelos exiguus QM338
Vitellidelos dorrigoensis MV1174
1/72
0.93/37
Vitellidelos dulcis QM311
Vitellidelos costata MV1536
Northern NSW
Stirling ranges rhytidid S42650
Stirling ranges rhytidid S42653
Stirling ranges, WA
1/41
0.93/38
Schizoglossa sp Transcriptome
NZ
Prolesophanta occlusa MV1533
Prolesophanta dryei MV114
Prolesophanta dryei L209
Southern NSW, VIC, TAS
0.73/21
Delos sp Transcriptome
NZ
0.99/28
Nata vernicosa Transcriptome
South Africa
Echotrida strangeoides MV1407
Echotrida strangeoides MV1203
Echotrida strangeoides MV1205
Echotrida globosa QM297
Southern QLD
Clade 4
Murphitella froggatti QM069
Murphitella froggatti MV1208
Murphitella franklandiensis 81830
Murphitella franklandiensis QM016
Saladelos commixita 81810
Saladelos commixita QM285
Saladelos commixita QM286
Far North Queensland
Torresiropa spaldingi QM346
Tip of Cape York and Torres strait Islands
Natalina cafra natalensis Transcriptome
Chlamydephorus gibbonsi Transcriptome
South Africa
Trigonephorus ambiguosus Transcriptome
0.02
Annabellia occidentalis
Austrorhytida capilliacea
Austrorhytida nandewarensis
Austrorhytida glacimans
Austrorhytida barringtonia
Protorugosa burraga
Protorugosa alpica
Griffithsina brisbanica
Terrycarlessia turbinata
Scagacola cavernula
Briansmithia ptychomphala
Annabellia assimilans
Annabellia subglobosa
Strangesta confusa
Limesta sheridani
Pseudechotrida mikros
Laevidelos moria
Pseudechotrida bouldercombe
Victaphanta compacta
Montidelos orcadis
Vitellidelos kaputarensis
Montidelos macquariensis
Montidelos exiguus
Vitellidelos dorrigoensis
Prolesophanta dryei
Echotrida strangeoides
Murphitella froggatti

Figure 4.4. Bayesian phylogeny of the Australian Rhytididae. The node labels represent support values from both the Bayesian (PP) and maximum likelihood analyses (BS). The asterisks represent nodes with complete bootstrap and posterior probability support. The blue boxes highlight the 'southern temperature lineages'.

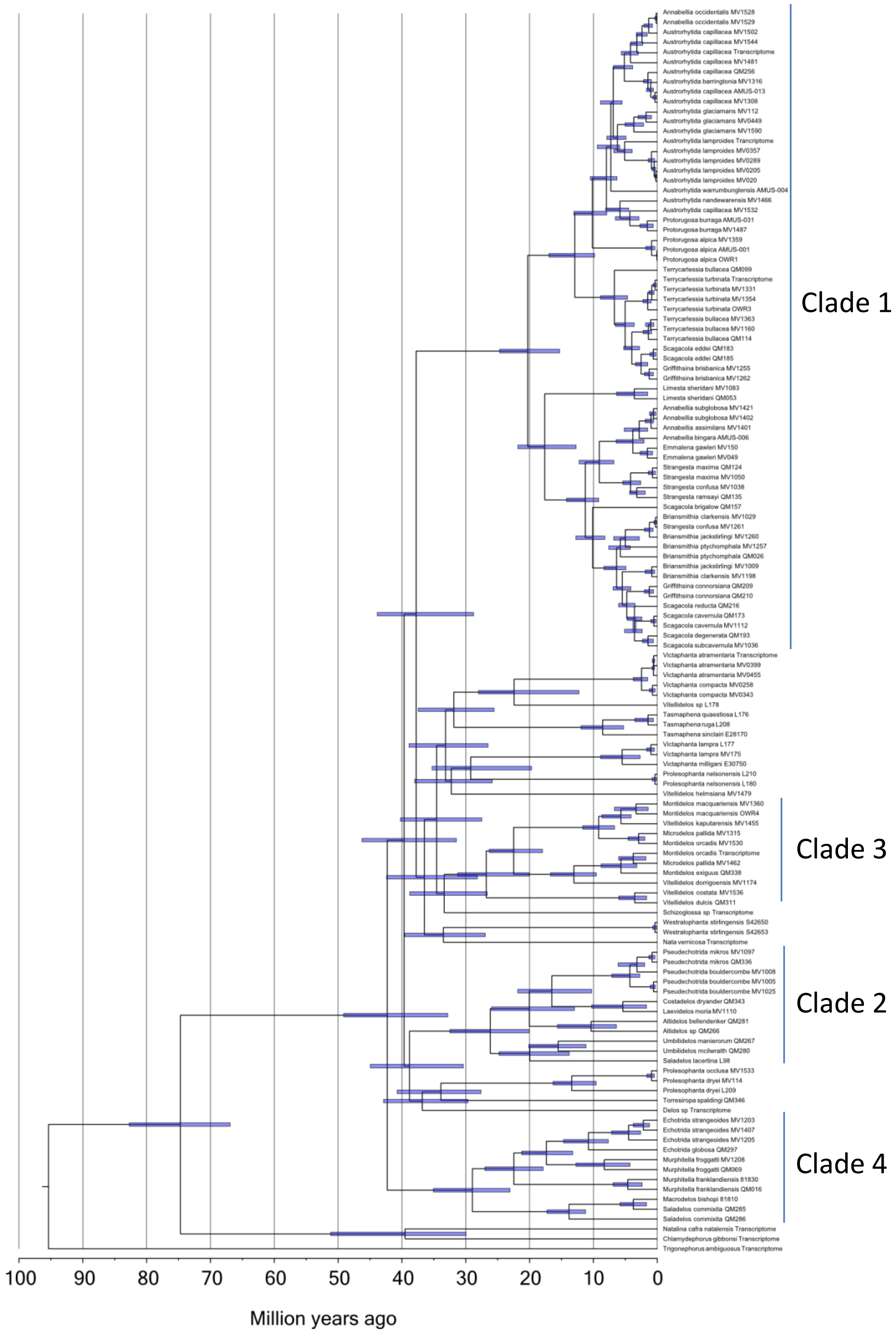

Annabellia occidentalis MV1528
Annabellia occidentalis MV1529
Austrorhytida capillacea MV1502
Austrorhytida capillacea MV1544
Austrorhytida capillacea Transcriptome
Austrorhytida capillacea MV1481
Austrorhytida capillacea QM256
Austrorhytida barringtonia MV1316
Austrorhytida capillacea AMUS-013
Austrorhytida capillacea MV1308
Austrorhytida glaciamans MV112
Austrorhytida glaciamans MV0449
Austrorhytida glaciamans MV1590
Austrorhytida lamproides Trancriptome
Austrorhytida lamproides MV0357
Austrorhytida lamproides MV0289
Austrorhytida lamproides MV0205
Austrorhytida lamproides MV020
Austrorhytida warrumbunglensis AMUS-004
Austrorhytida nandewarensis MV1466
Austrorhytida capillacea MV1532
Protorugosa burraga AMUS-031
Protorugosa burraga MV1487
Protorugosa alpica MV1359
Protorugosa alpica AMUS-001
Protorugosa alpica OWR1
Terrycarlessia bullacea QM099
Terrycarlessia turbinata Transcriptome
Terrycarlessia turbinata MV1331
Terrycarlessia turbinata MV1354
Terrycarlessia turbinata OWR3
Terrycarlessia bullacea MV1363
Terrycarlessia bullacea MV1160
Terrycarlessia bullacea QM114
Scagacola eddei QM183
Scagacola eddei QM185
Griffithsina brisbanica MV1255
Griffithsina brisbanica MV1262
Limesta sheridani MV1083
Limesta sheridani QM053
Annabellia subglobosa MV1421
Annabellia subglobosa MV1402
Annabellia assimilans MV1401
Annabellia bingara AMUS-006
Emmalena gawleri MV150
Emmalena gawleri MV049
Strangesta maxima QM124
Strangesta maxima MV1050
Strangesta confusa MV1038
Strangesta ramsayi QM135
Scagacola brigalow QM157
Briansmithia clarkensis MV1029
Strangesta confusa MV1261
Briansmithia jackstirlingi MV1260
Briansmithia ptychomphala MV1257
Briansmithia ptychomphala QM026
Briansmithia jackstirlingi MV1009
Briansmithia clarkensis MV1198
Griffithsina connorsiana QM209
Griffithsina connorsiana QM210
Scagacola reducta QM216
Scagacola cavernula QM173
Scagacola cavernula MV1112
Scagacola degenerata QM193
Scagacola subcavernula MV1036
Victaphanta atramentaria Transcriptome
Victaphanta atramentaria MV0399
Victaphanta atramentaria MV0455
Victaphanta compacta MV0258
Victaphanta compacta MV0343
Vitellidelos sp L178
Tasmaphena quaestiosa L176
Tasmaphena ruga L208
Tasmaphena sinclairi E28170
Victaphanta lampra L177
Victaphanta lampra MV175
Victaphanta milligani E30750
Prolesophanta nelsonensis L210
Prolesophanta nelsonensis L180
Vitellidelos helmsiana MV1479
Montidelos macquariensis MV1360
Montidelos macquariensis OWR4
Vitellidelos kaputarensis MV1455
Microdelos pallida MV1315
Montidelos orcadis MV1530
Montidelos orcadis Transcriptome
Microdelos pallida MV1462
Montidelos exiguus QM338
Vitellidelos dorrigoensis MV1174
Vitellidelos costata MV1536
Vitellidelos dulcis QM311
Schizoglossa sp Transcriptome
Westralophanta stirlingensis S42650
Westralophanta stirlingensis S42653
Nata vernicosa Transcriptome
Pseudechotrida mikros MV1097
Pseudechotrida mikros QM336
Pseudechotrida bouldercombe MV1008
Pseudechotrida bouldercombe MV1005
Pseudechotrida bouldercombe MV1025
Costadelos dryander QM343
Laevidelos moria MV1110
Altidelos bellendenker QM281
Altidelos sp QM266
Umbilidelos manierorum QM267
Umbilidelos mcilwraith QM280
Saladelos lacertina L98
Prolesophanta occlusa MV1533
Prolesophanta dryei MV114
Prolesophanta dryei L209
Torresiropa spaldingi QM346
Delos sp Transcriptome
Echotrida strangeoides MV1203
Echotrida strangeoides MV1407
Echotrida strangeoides MV1205
Echotrida globosa QM297
Murphitella froggatti MV1208
Murphitella froggatti QM069
Murphitella franklandiensis 81830
Murphitella franklandiensis QM016
Macrodelos bishopi 81810
Saladelos commixita QM285
Saladelos commixita QM286
Natalina cafra natalensis Transcriptome
Chlamydephorus gibbonsi Transcriptome
Trigonephorus ambiguosus Transcriptome
Clade 1
Clade 3
Clade 2
Clade 4
100
90
80
70
60
50
40
30
20
10
0
Million years ago

Figure 4.5. Calibrated phylogeny of the Australian Rhytididae estimated using BEAST.

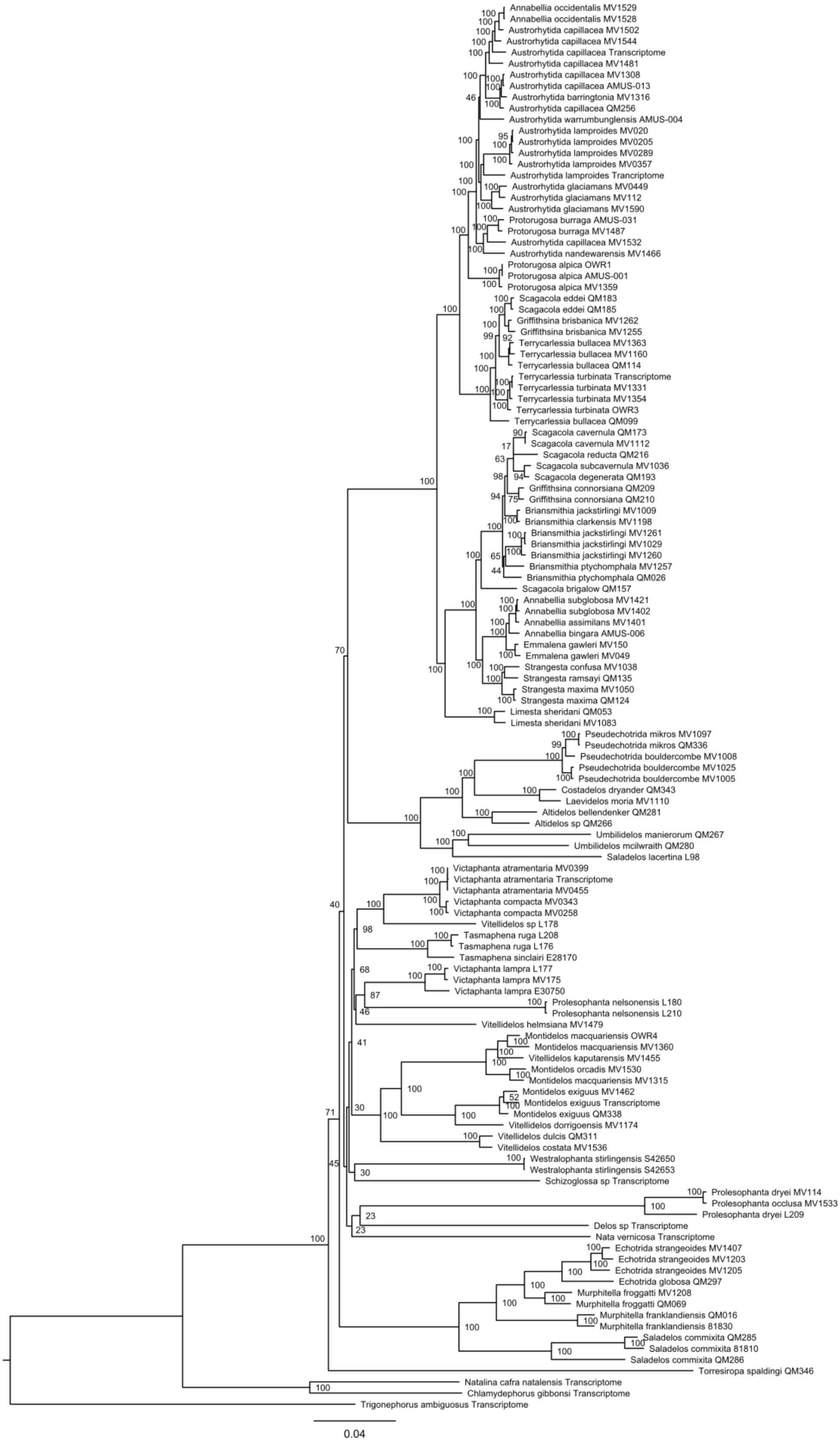
Table 4.1. Specimens used in the exon capture analyses. Localities in italics are at or very near the type localities of the species.

Clade	Genus	Species	Tissue no.	Locality	Year	DNA (ng)	No. exons	No. subset exons
2	*Altidelos*	*bellendenker*	QM281	*Mt Bellenden-ker, QLD*	1983	2	359	188
2	*Altidelos*	*sp.*	QM266	Bakers Blue Mt, QLD	1989	3	776	508
1	*Annabellia*	*assimilans*	MV1401	Jackadgery, NSW	2013	500	2948	1260
1	*Annabellia*	*bingara*	AMUS-006	Bundarra, NSW	1990	500	2849	1257
1	*Annabellia*	*occidentalis*	MV1528	*Black Spring, NSW*	2013	500	3454	1271
1	*Annabellia*	*occidentalis*	MV1529	*Black Spring, NSW*	2013	424	3658	1272
1	*Annabellia*	*subglobosa*	MV1421	Sundown NP, NSW	2013	417	2774	1256
1	*Annabellia*	*subglobosa*	MV1402	*Glenlyon, NSW*	2013	476	2334	1216
1	*Austrorhytida*	*barringtonia*	MV1532	*Burraga swamp, NSW*	2013	260	3296	1268
1	*Austrorhytida*	*capillacea*	MV1502	Mount Tomah, NSW	2013	419	3286	1270
1	*Austrorhytida*	*capillacea*	MV1544	Countegany, NSW	2013	427	3303	1268
1	*Austrorhytida*	*glaciamans*	MV1481	Wilsons Valley, NSW	2013	207	1869	1135
1	*Austrorhytida*	*capillacea*	QM256	*Sydney, NSW*	1993	87	2116	1182
1	*Austrorhytida*	*capillacea*	AMUS-013	Mt. Sugarloaf, NSW	1998	218	3385	1271
1	*Austrorhytida*	*capillacea*	MV1308	Mt. Sugarloaf, NSW	2013	310	3221	1271
1	*Austrorhytida*	*capillacea*	MV1316	Seal Rocks, NSW	2013	319	2766	1261
1	*Austrorhytida*	*glaciamans*	MV1481	*Wilsons Valley, NSW*	2013	207	1869	1135
1	*Austrorhytida*	*glaciamans*	MV112	Powelltown, VIC	2009	347	3375	1270
1	*Austrorhytida*	*glaciamans*	MV0449	Saxton, VIC	2009	33	2167	1001
1	*Austrorhytida*	*glaciamans*	MV1590	Native Dog Flat, Vic	2014	229	3287	1269
1	*Austrorhytida*	*glaciamans*	MV0357	Mount Cowley, VIC	2009	94	3773	1275
1	*Austrorhytida*	*glaciamans*	MV0289	Johanna, VIC	2009	39	3671	1272
1	*Austrorhytida*	*glaciamans*	MV0205	Triplet Falls, VIC	2009	83	3549	1271
1	*Austrorhytida*	*glaciamans*	MV020	Triplet Falls, VIC	2007	438	3251	1269
1	*Austrorhytida*	*nandewarensis*	MV1466	*Mt Kaputar NP, NSW*	2013	500	3530	1271
1	*Austrorhytida*	*warrumbunglensis*	AMUS-004	*Warrumbungle NP, NSW*	2001	458	3318	1273
1	*Briansmithia*	*clarkensis*	MV1198	*Eungella NP, QLD*	2013	308	2947	1264
1	*Briansmithia*	*clarkensis*	MV1009	*Eungella NP, QLD*	2013	118	2665	1243
1	*Briansmithia*	*jackstirlingi*	MV1029	Brandy creek, QLD	2013	319	2849	1259
1	*Briansmithia*	*jackstirlingi*	MV1261	Brandy creek, QLD	2013	347	2691	1246
1	*Briansmithia*	*jackstirlingi*	MV1260	*Gregory river, QLD*	2013	320	1410	947
1	*Briansmithia*	*ptychomphala*	MV1257	Bowling Green NP, QLD	2013	465	3095	1265
1	*Briansmithia*	*ptychomphala*	QM026	Hervey Ra, QLD	1994	58	292	182
2	*Costadelos*	*dryander*	QM343	*Mt Dryander, QLD*	1983	8	1169	829
5	*Echotrida*	*globosa*	QM297	*Kalpowar SF, QLD*	1984	36	1572	917
5	*Echotrida*	*strangeoides*	MV1205	Tamborine NP, QLD	2013	500	2504	1213
5	*Echotrida*	*strangeoides*	MV1203	Dorrigo NP, NSW	2013	457	2550	1207
5	*Echotrida*	*strangeoides*	MV1407	Davis Scrub NR, NSW	2013	383	2064	1142
1	*Emmalena*	*gawleri*	MV150	Mt Remarkable, SA	2009	345	3063	1270
1	*Emmalena*	*gawleri*	MV049	*Mt Lofty, SA*	2009	122	260	203
1	*Griffithsina*	*brisbanica*	MV1255	Benarkin SF, QLD	2013	500	2668	1259

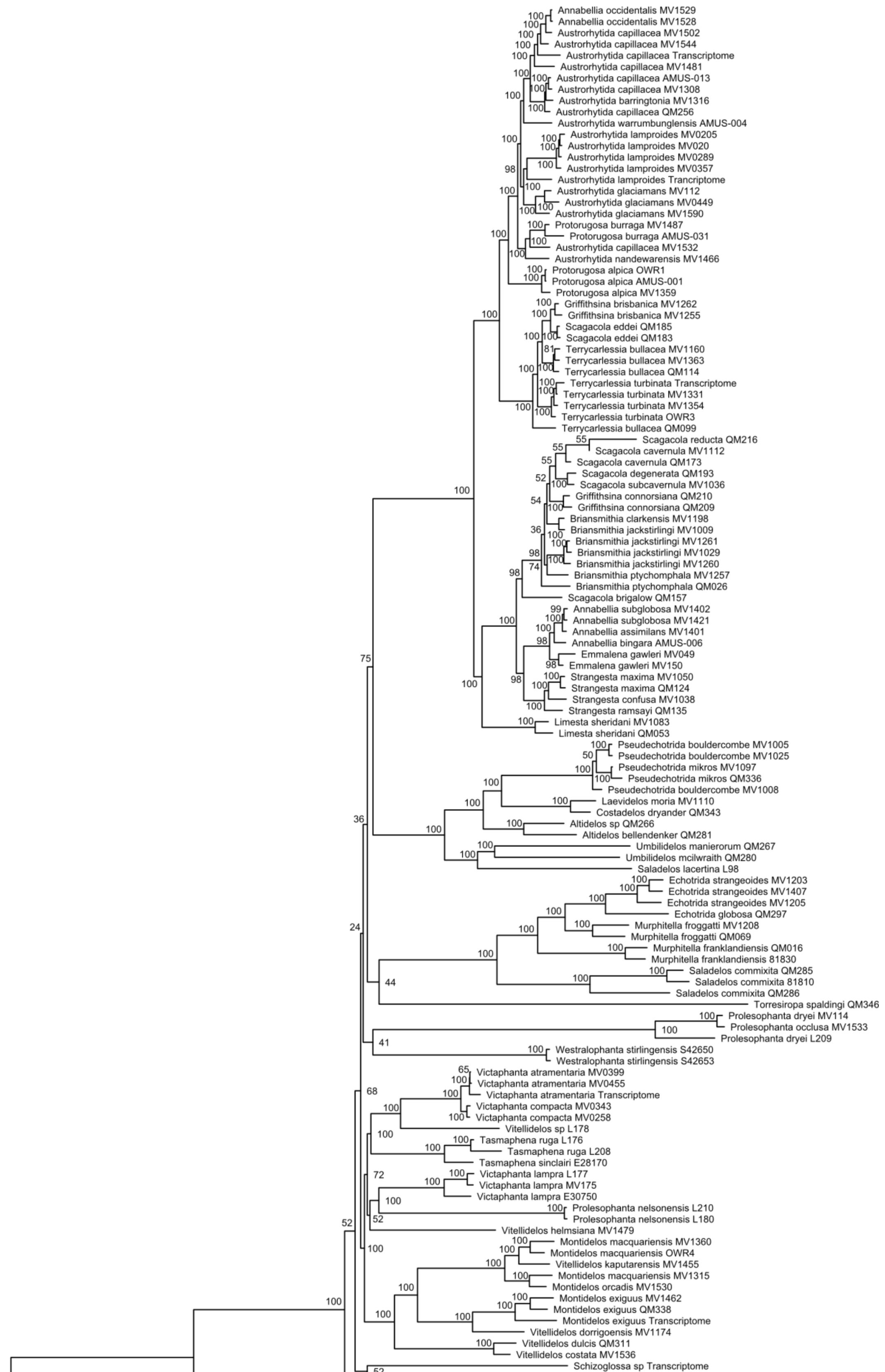
Clade	Genus	Species	Tissue no.	Locality	Year	DNA (ng)	No. exons	No. subset exons
1	*Griffithsina*	*brisbanica*	MV1262	Mount Glorious, QLD	2013	435	2587	1254
1	*Griffithsina*	*connorsiana*	QM209	Dipperu NP, QLD	1994	127	1506	994
1	*Griffithsina*	*connorsiana*	QM210	*Connors Ra, QLD*	1994	60	3072	1261
2	*Laevidelos*	*moria*	MV1110	*Mt Etna NP, QLD*	2013	442	642	523
1	*Limesta*	*sheridani*	MV1083	Maalan, QLD	2013	500	2652	1238
1	*Limesta*	*sheridani*	QM053	Mt Spurgeon, QLD	1991	130	633	459
3	*Montidelos*	*exiguus*	QM338	Kenilworth SF, QLD	1980	179	329	214
3	*Montidelos*	*exiguus*	MV1462	Mallanganee, NSW	2013	209	80	56
3	*Montidelos*	*macquariensis*	MV1360	*Port Macquarie, NSW*	2013	319	2970	1264
3	*Montidelos*	*macquariensis*	OWR4	Oxley Wild Rivers NP, NSW	2015	367	2491	1193
3	*Montidelos*	*macquariensis*	MV1315	Wingham Brush, NSW	2013	109	1596	983
3	*Montidelos*	*orcadis*	MV1530	*Burraga Swamp, NSW*	2013	348	2648	1242
5	*Murphitella*	*franklandiensis*	81830	Cape York, QLD	2013	191	2202	1194
5	*Murphitella*	*franklandiensis*	QM016	Cairns, QLD	1997	230	2545	1230
5	*Murphitella*	*froggatti*	MV1208	East Barron, QLD	2013	229	708	490
5	*Murphitella*	*froggatti*	QM069	Wongabel SF, Qld	1988	42	990	627
8	*Prolesophanta*	*dryei*	MV114	Saxton, Vic	2009	398	2656	1215
8	*Prolesophanta*	*dryei*	L209	TAS	2013	81	2613	1219
4	*Prolesophanta*	*nelsonensis*	L210	TAS	2013	215	2919	1252
4	*Prolesophanta*	*nelsonensis*	L180	TAS	2013	256	3076	1254
8	*Prolesophanta*	*occlusa*	MV1533	Mount Tomah, NSW	2013	176	2401	1169
1	*Protorugosa*	*alpica*	MV1359	Dorrigo NP, NSW	2013	74	1475	960
1	*Protorugosa*	*alpica*	AMUS-001	Tapin Tops NP, NSW	2007	333	3344	1270
1	*Protorugosa*	*alpica*	OWR1	Oxley Wild Rivers NP, NSW	2015	500	3254	1266
1	*Protorugosa*	*burraga*	AMUS-031	Gloucester Tops, NSW	1993	4	133	56
1	*Protorugosa*	*burraga*	MV1487	Burraga Swamp, NSW	2013	85	1939	1120
2	*Pseudechotrida*	*bouldercombe*	MV1008	Woowoonga NP, QLD	2013	284	2205	1190
2	*Pseudechotrida*	*bouldercombe*	MV1025	Mt Biggenden, QLD	2013	500	2472	1221
2	*Pseudechotrida*	*bouldercombe*	MV1005	Mt Biggenden, QLD	2013	236	2593	1231
2	*Pseudechotrida*	*mikros*	QM336	Murgon, QLD	1994	9	835	590
2	*Pseudechotrida*	*mikros*	MV1097	Benarkin SF, QLD	2013	485	2711	1253
5	*Saladelos*	*commixita*	QM285	Coen, QLD	1988	191	2149	1108
5	*Saladelos*	*commixita*	QM286	McIvor River, QLD	1988	285	2370	1173
5	*Saladelos*	*commixita*	81810	Cape York, QLD	2013	500	2718	1211
2	*Saladelos*	*lacertina*	L98	*Lizard ls, QLD*	2014	500	1969	1120
1	*Scagacola*	*brigalow*	QM157	Robinson Gorge NP, QLD	1992	61	2682	1246
1	*Scagacola*	*cavernula*	QM173	Johannsens Caves, QLD	1994	67	2876	1259
1	*Scagacola*	*cavernula*	MV1112	*Mt Etna NP, QLD*	2013	110	146	97
1	*Scagacola*	*degenerata*	QM193	*Biggenden, QLD*	1992	65	1585	1005
1	*Scagacola*	*eddei*	QM183	Tabletop Mt, QLD	1993	325	2871	1265
1	*Scagacola*	*eddei*	QM185	Gatton, QLD	1993	328	3226	1271
1	*Scagacola*	*reducta*	QM216	Eungella, QLD	1990	2	19	7
1	*Scagacola*	*subcavernula*	MV1036	Mt Mudlo NP, QLD	2013	245	2596	1246
1	*Strangesta*	*confusa*	MV1038	Eungella NP, QLD	2013	327	2752	1259

Clade	Genus	Species	Tissue no.	Locality	Year	DNA (ng)	No. exons	No. subset exons
1	*Strangesta*	*maxima*	QM124	Goomeri, QLD	1994	179	1368	919
1	*Strangesta*	*maxima*	MV1050	Mt Biggenden, QLD	2013	500	2985	1262
1	*Strangesta*	*ramsayi*	QM135	Lamington NP, QLD	1985	290	2606	1226
4	*Tasmaphena*	*ruga*	L176	TAS	2013	500	3125	1258
4	*Tasmaphena*	*ruga*	L208	TAS	2013	48	203	105
4	*Tasmaphena*	*sinclairi*	E28170	Viormy, TAS	2014	351	2932	1262
1	*Terrycarlessia*	*bullacea*	QM099	*Bunya Mts NP, QLD*	1985	265	1200	819
1	*Terrycarlessia*	*bullacea*	MV1363	Iluka, NSW	2013	274	3261	1267
1	*Terrycarlessia*	*bullacea*	MV1160	Dorrigo NP, NSW	2013	408	3018	1264
1	*Terrycarlessia*	*bullacea*	QM114	Yabbra SF, NSW	1991	73	2409	1235
1	*Terrycarlessia*	*turbinata*	MV1331	*Boorganna NR, QLD*	2013	500	3152	1271
1	*Terrycarlessia*	*turbinata*	MV1354	Port Macquarie, NSW	2013	500	2945	1264
1	*Terrycarlessia*	*turbinata*	OWR3	Oxley Wild Rivers NP, NSW	2015	254	2712	1239
7	*Torresiropa*	*spaldingi*	QM346	*Punsand Bay, NSW*	1988	39	962	647
2	*Umbilidelos*	*manierorum*	QM267	McIvor River, QLD	1988	158	1659	922
2	*Umbilidelos*	*mcilwraith*	QM280	Pascoe River, QLD	1988	7	729	383
4	*Victaphanta*	*atramentaria*	MV0455	Baw Baw, VIC	2009	405	3057	1269
4	*Victaphanta*	*atramentaria*	MV0399	Toolangi, VIC	2009	433	2897	1264
4	*Victaphanta*	*compacta*	MV0258	*Melba Gully, VIC*	2009	168	2912	1253
4	*Victaphanta*	*compacta*	MV0343	*Mount Cowley, VIC*	2009	160	2505	1236
4	*Victaphanta*	*lampra*	MV175	TAS	2009	500	3315	1268
4	*Victaphanta*	*lampra*	L177	TAS	2013	41	1657	963
4	*Victaphanta*	*lampra*	E30750	Pieman River SR, TAS	2015	168	2755	1255
3	*Vitellidelos*	*costata*	MV1536	Masseys Creek SF	2013	441	2510	1251
3	*Vitellidelos*	*dorrigoensis*	MV1174	*Dorrigo NP, NSW*	2013	285	2947	1248
3	*Vitellidelos*	*dulcis*	QM311	Maitland, NSW	1994	11	479	269
4	*Vitellidelos*	*helmsiana*	MV1479	*Wilsons Valley, NSW*	2013	500	3184	1270
3	*Vitellidelos*	*kaputarensis*	MV1455	*Mt Kaputar NP, NSW*	2013	123	1741	1066
4	*Vitellidelos*	*sp.*	L178	TAS	2013	238	2879	1222
6	Stirling ranges	rhytidid	S42653	Stirling ranges, WA	2008	220	1391	936
6	Stirling ranges	rhytidid	S42650	Stirling ranges, WA	2008	500	3014	1250

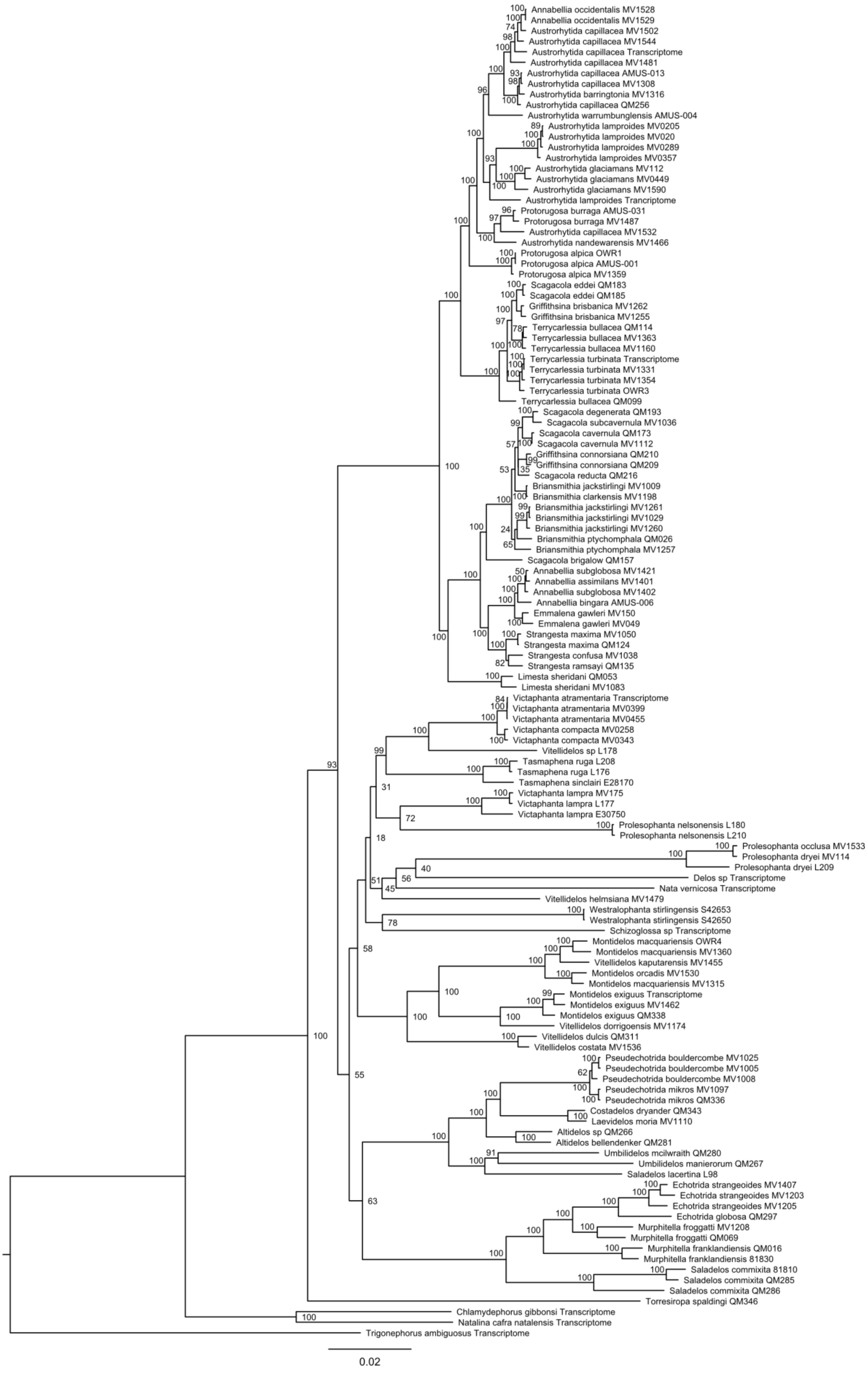
4.6 APPENDICES


Annabellia occidentalis MV1529
Annabellia occidentalis MV1528
Austrorhytida capillacea MV1502
Austrorhytida capillacea MV1544
Austrorhytida capillacea Transcriptome
Austrorhytida capillacea MV1481
Austrorhytida capillacea MV1308
Austrorhytida capillacea AMUS-013
Austrorhytida barringtonia MV1316
Austrorhytida capillacea QM256
Austrorhytida warrumbunglensis AMUS-004
Austrorhytida lamproides MV020
Austrorhytida lamproides MV0205
Austrorhytida lamproides MV0289
Austrorhytida lamproides MV0357
Austrorhytida lamproides Trancriptome
Austrorhytida glaciamans MV0449
Austrorhytida glaciamans MV112
Austrorhytida glaciamans MV1590
Protorugosa burraga AMUS-031
Protorugosa burraga MV1487
Austrorhytida capillacea MV1532
Austrorhytida nandewarensis MV1466
Protorugosa alpica OWR1
Protorugosa alpica AMUS-001
Protorugosa alpica MV1359
Scagacola eddei QM183
Scagacola eddei QM185
Griffithsina brisbanica MV1262
Griffithsina brisbanica MV1255
Terrycarlessia bullacea MV1363
Terrycarlessia bullacea MV1160
Terrycarlessia bullacea QM114
Terrycarlessia turbinata Transcriptome
Terrycarlessia turbinata MV1331
Terrycarlessia turbinata MV1354
Terrycarlessia turbinata OWR3
Terrycarlessia bullacea QM099
Scagacola cavernula QM173
Scagacola cavernula MV1112
Scagacola reducta QM216
Scagacola subcavernula MV1036
Scagacola degenerata QM193
Griffithsina connorsiana QM209
Griffithsina connorsiana QM210
Briansmithia jackstirlingi MV1009
Briansmithia clarkensis MV1198
Briansmithia jackstirlingi MV1261
Briansmithia jackstirlingi MV1029
Briansmithia jackstirlingi MV1260
Briansmithia ptychomphala MV1257
Briansmithia ptychomphala QM026
Scagacola brigalow QM157
Annabellia subglobosa MV1421
Annabellia subglobosa MV1402
Annabellia assimilans MV1401
Annabellia bingara AMUS-006
Emmalena gawleri MV150
Emmalena gawleri MV049
Strangesta confusa MV1038
Strangesta ramsayi QM135
Strangesta maxima MV1050
Strangesta maxima QM124
Limesta sheridani QM053
Limesta sheridani MV1083
Pseudechotrida mikros MV1097
Pseudechotrida mikros QM336
Pseudechotrida bouldercombe MV1008
Pseudechotrida bouldercombe MV1025
Pseudechotrida bouldercombe MV1005
Costadelos dryander QM343
Laevidelos moria MV1110
Altidelos bellendenker QM281
Altidelos sp QM266
Umbilidelos manierorum QM267
Umbilidelos mcilwraith QM280
Saladelos lacertina L98
Victaphanta atramentaria MV0399
Victaphanta atramentaria Transcriptome
Victaphanta atramentaria MV0455
Victaphanta compacta MV0343
Victaphanta compacta MV0258
Vitellidelos sp L178
Tasmaphena ruga L208
Tasmaphena ruga L176
Tasmaphena sinclairi E28170
Victaphanta lampra L177
Victaphanta lampra MV175
Victaphanta lampra E30750
Prolesophanta nelsonensis L180
Prolesophanta nelsonensis L210
Vitellidelos helmsiana MV1479
Montidelos macquariensis OWR4
Montidelos macquariensis MV1360
Vitellidelos kaputarensis MV1455
Montidelos orcadis MV1530
Montidelos macquariensis MV1315
Montidelos exiguus MV1462
Montidelos exiguus Transcriptome
Montidelos exiguus QM338
Vitellidelos dorrigoensis MV1174
Vitellidelos dulcis QM311
Vitellidelos costata MV1536
Westralophanta stirlingensis S42650
Westralophanta stirlingensis S42653
Schizoglossa sp Transcriptome
Prolesophanta dryei MV114
Prolesophanta occlusa MV1533
Prolesophanta dryei L209
Delos sp Transcriptome
Nata vernicosa Transcriptome
Echotrida strangeoides MV1407
Echotrida strangeoides MV1203
Echotrida strangeoides MV1205
Echotrida globosa QM297
Murphitella froggatti MV1208
Murphitella froggatti QM069
Murphitella franklandiensis QM016
Murphitella franklandiensis 81830
Saladelos commixita QM285
Saladelos commixita 81810
Saladelos commixita QM286
Torresiropa spaldingi QM346
Natalina cafra natalensis Transcriptome
Chlamydephorus gibbonsi Transcriptome
Trigonephorus ambiguosus Transcriptome
0.04

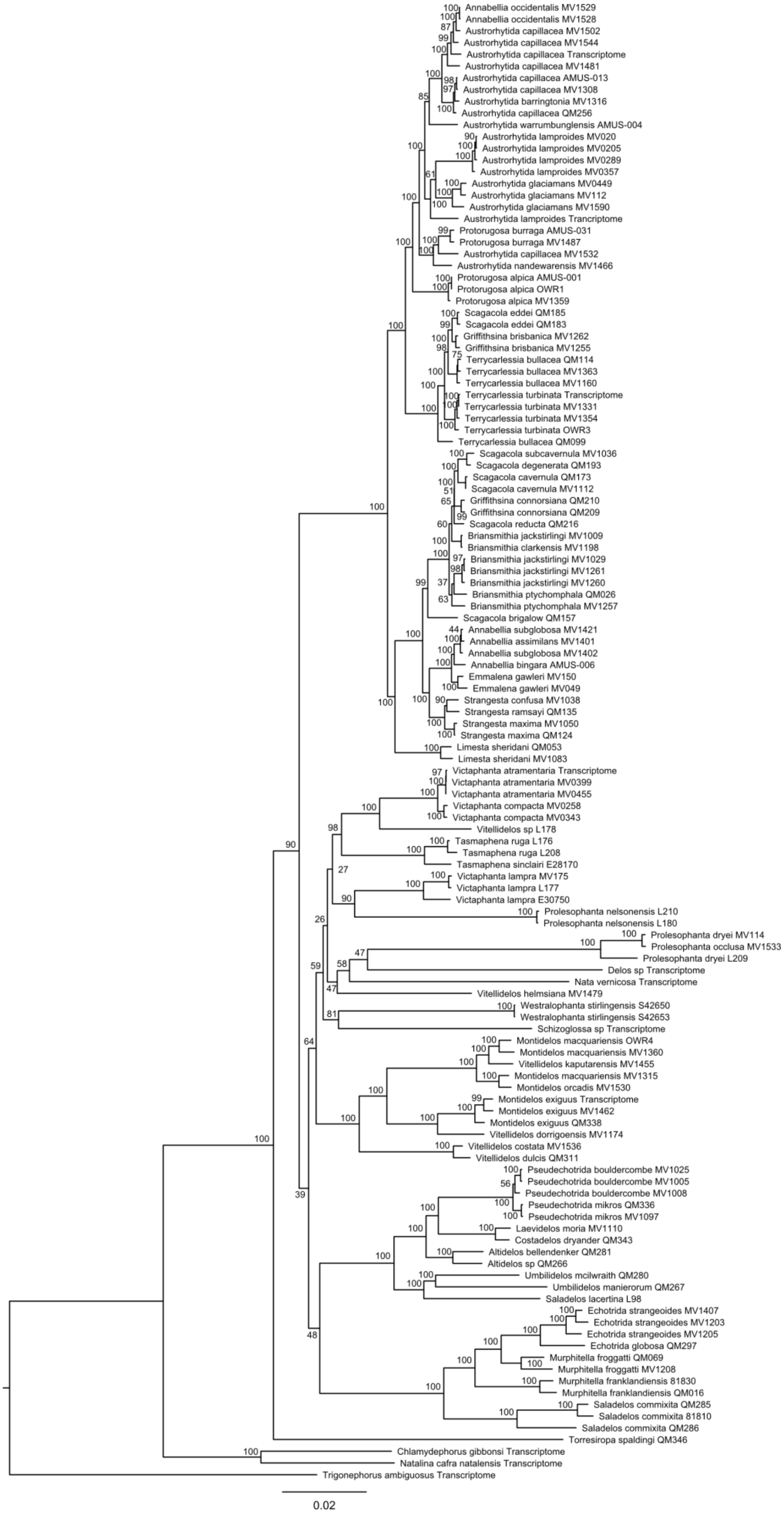
Appendix 4.1. Maximum likelihood analysis with GTR + I + Γ.


Annabellia occidentalis MV1529
Annabellia occidentalis MV1528
Austrorhytida capillacea MV1502
Austrorhytida capillacea MV1544
Austrorhytida capillacea Transcriptome
Austrorhytida capillacea MV1481
Austrorhytida capillacea AMUS-013
Austrorhytida capillacea MV1308
Austrorhytida barringtonia MV1316
Austrorhytida capillacea QM256
Austrorhytida warrumbunglensis AMUS-004
Austrorhytida lamproides MV0205
Austrorhytida lamproides MV020
Austrorhytida lamproides MV0289
Austrorhytida lamproides MV0357
Austrorhytida lamproides Trancriptome
Austrorhytida glaciamans MV112
Austrorhytida glaciamans MV0449
Austrorhytida glaciamans MV1590
Protorugosa burraga MV1487
Protorugosa burraga AMUS-031
Austrorhytida capillacea MV1532
Austrorhytida nandewarensis MV1466
Protorugosa alpica OWR1
Protorugosa alpica AMUS-001
Protorugosa alpica MV1359
Griffithsina brisbanica MV1262
Griffithsina brisbanica MV1255
Scagacola eddei QM185
Scagacola eddei QM183
Terrycarlessia bullacea MV1160
Terrycarlessia bullacea MV1363
Terrycarlessia bullacea QM114
Terrycarlessia turbinata Transcriptome
Terrycarlessia turbinata MV1331
Terrycarlessia turbinata MV1354
Terrycarlessia turbinata OWR3
Terrycarlessia bullacea QM099
Scagacola reducta QM216
Scagacola cavernula MV1112
Scagacola cavernula QM173
Scagacola degenerata QM193
Scagacola subcavernula MV1036
Griffithsina connorsiana QM210
Griffithsina connorsiana QM209
Briansmithia clarkensis MV1198
Briansmithia jackstirlingi MV1009
Briansmithia jackstirlingi MV1261
Briansmithia jackstirlingi MV1029
Briansmithia jackstirlingi MV1260
Briansmithia ptychomphala MV1257
Briansmithia ptychomphala QM026
Scagacola brigalow QM157
Annabellia subglobosa MV1402
Annabellia subglobosa MV1421
Annabellia assimilans MV1401
Annabellia bingara AMUS-006
Emmalena gawleri MV049
Emmalena gawleri MV150
Strangesta maxima MV1050
Strangesta maxima QM124
Strangesta confusa MV1038
Strangesta ramsayi QM135
Limesta sheridani MV1083
Limesta sheridani QM053
Pseudechotrida bouldercombe MV1005
Pseudechotrida bouldercombe MV1025
Pseudechotrida mikros MV1097
Pseudechotrida mikros QM336
Pseudechotrida bouldercombe MV1008
Laevidelos moria MV1110
Costadelos dryander QM343
Altidelos sp QM266
Altidelos bellendenker QM281
Umbilidelos manierorum QM267
Umbilidelos mcilwraith QM280
Saladelos lacertina L98
Echotrida strangeoides MV1203
Echotrida strangeoides MV1407
Echotrida strangeoides MV1205
Echotrida globosa QM297
Murphitella froggatti MV1208
Murphitella froggatti QM069
Murphitella franklandiensis QM016
Murphitella franklandiensis 81830
Saladelos commixita QM285
Saladelos commixita 81810
Saladelos commixita QM286
Torresiropa spaldingi QM346
Prolesophanta dryei MV114
Prolesophanta occlusa MV1533
Prolesophanta dryei L209
Westralophanta stirlingensis S42650
Westralophanta stirlingensis S42653
Victaphanta atramentaria MV0399
Victaphanta atramentaria MV0455
Victaphanta atramentaria Transcriptome
Victaphanta compacta MV0343
Victaphanta compacta MV0258
Vitellidelos sp L178
Tasmaphena ruga L176
Tasmaphena ruga L208
Tasmaphena sinclairi E28170
Victaphanta lampra L177
Victaphanta lampra MV175
Victaphanta lampra E30750
Prolesophanta nelsonensis L210
Prolesophanta nelsonensis L180
Vitellidelos helmsiana MV1479
Montidelos macquariensis MV1360
Montidelos macquariensis OWR4
Vitellidelos kaputarensis MV1455
Montidelos macquariensis MV1315
Montidelos orcadis MV1530
Montidelos exiguus MV1462
Montidelos exiguus QM338
Montidelos exiguus Transcriptome
Vitellidelos dorrigoensis MV1174
Vitellidelos dulcis QM311
Vitellidelos costata MV1536
Schizoglossa sp Transcriptome
Delos sp Transcriptome
Nata vernicosa Transcriptome
Natalina cafra natalensis Transcriptome
Chlamydephorus gibbonsi Transcriptome
Trigonephorus ambiguosus Transcriptome
0.03

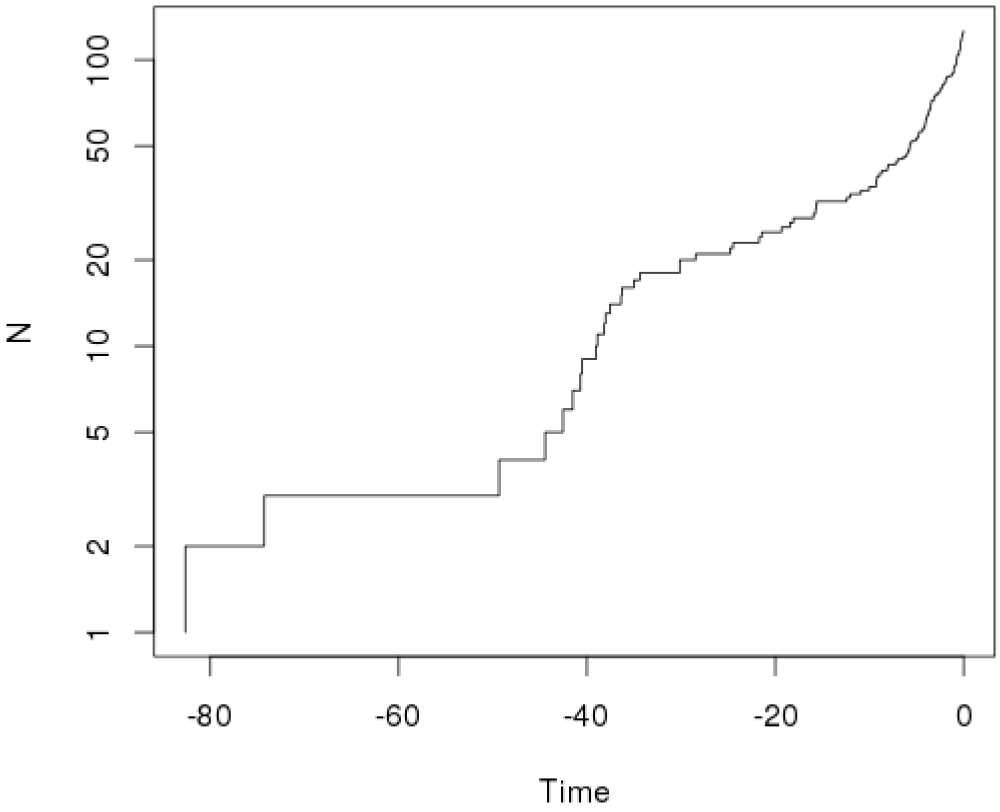
Appendix 4.2. Maximum likelihood analysis with GTR + Γ with the full but sparse matrix


100 Annabellia occidentalis MV1528
100 Annabellia occidentalis MV1529
74 Austrorhytida capillacea MV1502
98 Austrorhytida capillacea MV1544
100 Austrorhytida capillacea Transcriptome
Austrorhytida capillacea MV1481
100
93 Austrorhytida capillacea AMUS-013
98 Austrorhytida capillacea MV1308
96
Austrorhytida barringtonia MV1316
100 Austrorhytida capillacea QM256
Austrorhytida warrumbunglensis AMUS-004
89 Austrorhytida lamproides MV0205
100
100 Austrorhytida lamproides MV020
100 Austrorhytida lamproides MV0289
Austrorhytida lamproides MV0357
93
100 Austrorhytida glaciamans MV112
100 Austrorhytida glaciamans MV0449
100
100 Austrorhytida glaciamans MV1590
Austrorhytida lamproides Trancriptome
96 Protorugosa burraga AMUS-031
97 Protorugosa burraga MV1487
100
Austrorhytida capillacea MV1532
100 Austrorhytida nandewarensis MV1466
100 Protorugosa alpica OWR1
100 Protorugosa alpica AMUS-001
Protorugosa alpica MV1359
100 Scagacola eddei QM183
100 Scagacola eddei QM185
100
100 Griffithsina brisbanica MV1262
Griffithsina brisbanica MV1255
97
78 Terrycarlessia bullacea QM114
Terrycarlessia bullacea MV1363
100
100 Terrycarlessia bullacea MV1160
100 Terrycarlessia turbinata Transcriptome
100 Terrycarlessia turbinata MV1331
100
100 Terrycarlessia turbinata MV1354
Terrycarlessia turbinata OWR3
Terrycarlessia bullacea QM099
100 Scagacola degenerata QM193
99 Scagacola subcavernula MV1036
Scagacola cavernula QM173
57
100 Scagacola cavernula MV1112
Griffithsina connorsiana QM210
100
99 Griffithsina connorsiana QM209
53
35 Scagacola reducta QM216
Briansmithia jackstirlingi MV1009
100 Briansmithia clarkensis MV1198
100
99 Briansmithia jackstirlingi MV1261
99 Briansmithia jackstirlingi MV1029
100
24 Briansmithia jackstirlingi MV1260
Briansmithia ptychomphala QM026
65 Briansmithia ptychomphala MV1257
Scagacola brigalow QM157
100
50 Annabellia subglobosa MV1421
100 Annabellia assimilans MV1401
100 Annabellia subglobosa MV1402
100 Annabellia bingara AMUS-006
Emmalena gawleri MV150
100 Emmalena gawleri MV049
100
100
100 Strangesta maxima MV1050
100 Strangesta maxima QM124
Strangesta confusa MV1038
82 Strangesta ramsayi QM135
100 Limesta sheridani QM053
Limesta sheridani MV1083
84 Victaphanta atramentaria Transcriptome
Victaphanta atramentaria MV0399
100 Victaphanta atramentaria MV0455
100 Victaphanta compacta MV0258
100 Victaphanta compacta MV0343
99 Vitellidelos sp L178
100 Tasmaphena ruga L208
93
100 Tasmaphena ruga L176
Tasmaphena sinclairi E28170
31
100 Victaphanta lampra MV175
100 Victaphanta lampra L177
Victaphanta lampra E30750
72
100 Prolesophanta nelsonensis L180
Prolesophanta nelsonensis L210
18
100 Prolesophanta occlusa MV1533
100 Prolesophanta dryei MV114
Prolesophanta dryei L209
40
56 Delos sp Transcriptome
51
45 Nata vernicosa Transcriptome
Vitellidelos helmsiana MV1479
100 Westralophanta stirlingensis S42653
Westralophanta stirlingensis S42650
78 Schizoglossa sp Transcriptome
100 Montidelos macquariensis OWR4
100 Montidelos macquariensis MV1360
58
100 Vitellidelos kaputarensis MV1455
100 Montidelos orcadis MV1530
Montidelos macquariensis MV1315
100
99 Montidelos exiguus Transcriptome
100 Montidelos exiguus MV1462
100
100 Montidelos exiguus QM338
Vitellidelos dorrigoensis MV1174
100 Vitellidelos dulcis QM311
100
Vitellidelos costata MV1536
100 Pseudechotrida bouldercombe MV1025
Pseudechotrida bouldercombe MV1005
62 Pseudechotrida bouldercombe MV1008
55
100 Pseudechotrida mikros MV1097
100 Pseudechotrida mikros QM336
100
100 Costadelos dryander QM343
100
100 Laevidelos moria MV1110
Altidelos sp QM266
100
100 Altidelos bellendenker QM281
91 Umbilidelos mcilwraith QM280
100 Umbilidelos manierorum QM267
Saladelos lacertina L98
100 Echotrida strangeoides MV1407
100 Echotrida strangeoides MV1203
63
100 Echotrida strangeoides MV1205
100 Echotrida globosa QM297
100 Murphitella froggatti MV1208
100 Murphitella froggatti QM069
100
100 Murphitella franklandiensis QM016
Murphitella franklandiensis 81830
100 Saladelos commixita 81810
100 Saladelos commixita QM285
Saladelos commixita QM286
Torresiropa spaldingi QM346
Chlamydephorus gibbonsi Transcriptome
100 Natalina cafra natalensis Transcriptome
Trigonephorus ambiguosus Transcriptome
0.02

Appendix 4.3. Maximum likelihood analysis with amino acids JTT + Γ one partition


Annabellia occidentalis MV1529
Annabellia occidentalis MV1528
Austrorhytida capillacea MV1502
Austrorhytida capillacea MV1544
Austrorhytida capillacea Transcriptome
Austrorhytida capillacea MV1481
Austrorhytida capillacea AMUS-013
Austrorhytida capillacea MV1308
Austrorhytida barringtonia MV1316
Austrorhytida capillacea QM256
Austrorhytida warrumbunglensis AMUS-004
Austrorhytida lamproides MV020
Austrorhytida lamproides MV0205
Austrorhytida lamproides MV0289
Austrorhytida lamproides MV0357
Austrorhytida glaciamans MV0449
Austrorhytida glaciamans MV112
Austrorhytida glaciamans MV1590
Austrorhytida lamproides Trancriptome
Protorugosa burraga AMUS-031
Protorugosa burraga MV1487
Austrorhytida capillacea MV1532
Austrorhytida nandewarensis MV1466
Protorugosa alpica AMUS-001
Protorugosa alpica OWR1
Protorugosa alpica MV1359
Scagacola eddei QM185
Scagacola eddei QM183
Griffithsina brisbanica MV1262
Griffithsina brisbanica MV1255
Terrycarlessia bullacea QM114
Terrycarlessia bullacea MV1363
Terrycarlessia bullacea MV1160
Terrycarlessia turbinata Transcriptome
Terrycarlessia turbinata MV1331
Terrycarlessia turbinata MV1354
Terrycarlessia turbinata OWR3
Terrycarlessia bullacea QM099
Scagacola subcavernula MV1036
Scagacola degenerata QM193
Scagacola cavernula QM173
Scagacola cavernula MV1112
Griffithsina connorsiana QM210
Griffithsina connorsiana QM209
Scagacola reducta QM216
Briansmithia jackstirlingi MV1009
Briansmithia clarkensis MV1198
Briansmithia jackstirlingi MV1029
Briansmithia jackstirlingi MV1261
Briansmithia jackstirlingi MV1260
Briansmithia ptychomphala QM026
Briansmithia ptychomphala MV1257
Scagacola brigalow QM157
Annabellia subglobosa MV1421
Annabellia assimilans MV1401
Annabellia subglobosa MV1402
Annabellia bingara AMUS-006
Emmalena gawleri MV150
Emmalena gawleri MV049
Strangesta confusa MV1038
Strangesta ramsayi QM135
Strangesta maxima MV1050
Strangesta maxima QM124
Limesta sheridani QM053
Limesta sheridani MV1083
Victaphanta atramentaria Transcriptome
Victaphanta atramentaria MV0399
Victaphanta atramentaria MV0455
Victaphanta compacta MV0258
Victaphanta compacta MV0343
Vitellidelos sp L178
Tasmaphena ruga L176
Tasmaphena ruga L208
Tasmaphena sinclairi E28170
Victaphanta lampra MV175
Victaphanta lampra L177
Victaphanta lampra E30750
Prolesophanta nelsonensis L210
Prolesophanta nelsonensis L180
Prolesophanta dryei MV114
Prolesophanta occlusa MV1533
Prolesophanta dryei L209
Delos sp Transcriptome
Nata vernicosa Transcriptome
Vitellidelos helmsiana MV1479
Westralophanta stirlingensis S42650
Westralophanta stirlingensis S42653
Schizoglossa sp Transcriptome
Montidelos macquariensis OWR4
Montidelos macquariensis MV1360
Vitellidelos kaputarensis MV1455
Montidelos macquariensis MV1315
Montidelos orcadis MV1530
Montidelos exiguus Transcriptome
Montidelos exiguus MV1462
Montidelos exiguus QM338
Vitellidelos dorrigoensis MV1174
Vitellidelos costata MV1536
Vitellidelos dulcis QM311
Pseudechotrida bouldercombe MV1025
Pseudechotrida bouldercombe MV1005
Pseudechotrida bouldercombe MV1008
Pseudechotrida mikros QM336
Pseudechotrida mikros MV1097
Laevidelos moria MV1110
Costadelos dryander QM343
Altidelos bellendenker QM281
Altidelos sp QM266
Umbilidelos mcilwraith QM280
Umbilidelos manierorum QM267
Saladelos lacertina L98
Echotrida strangeoides MV1407
Echotrida strangeoides MV1203
Echotrida strangeoides MV1205
Echotrida globosa QM297
Murphitella froggatti QM069
Murphitella froggatti MV1208
Murphitella franklandiensis 81830
Murphitella franklandiensis QM016
Saladelos commixita QM285
Saladelos commixita 81810
Saladelos commixita QM286
Torresiropa spaldingi QM346
Chlamydephorus gibbonsi Transcriptome
Natalina cafra natalensis Transcriptome
Trigonephorus ambiguosus Transcriptome
0.02

Appendix 4.4. Maximum likelihood analysis with amino acids and seven partitions based on exon clustering by amino acid composition JTT + Γ.



Appendix 4.5. The lineage through time plot for the Rhytididae based on the taxonomic sampling and BEAST analysis presented in this study. N is the number of lineages and time is in units of millions of years before present.

CHAPTER 5:

General discussion

In this thesis I have used genomic-scale datasets to investigate the pattern and timing of previously uncertain evolutionary relationships within the pulmonate snails and slugs. I have addressed relationships at multiple scales including deep relationships between the major pulmonate lineages and the genus level relationships within the carnivorous land snail family the Rhytididae. This research has paved the way for multiple avenues of future research which I discuss below. In particular, I highlight that the pulmonates are a promising system for future studies investigating evolutionary processes such as genome evolution and adaptation.

I have addressed long-standing evolutionary questions regarding the pattern of pulmonate evolution using orthologous genes identified from transcriptome datasets (Chapter 2). The air-breathing snails and slugs (Pulmonata) were traditionally regarded as monophyletic but morphological and molecular studies have questioned this monophyly. Panpulmonata, which unites traditionally pulmonate and non-air-breathing lineages, was recently proposed by Jörger et al. (2010). Despite a number of molecular studies, the relationships between the major lineages within Panpulmonata, however, remained uncertain (Chapter 3: Figure 3.1). Combined with previously sequenced datasets, I sequenced new transcriptomes for a number of pulmonate lineages to address the relationships within Panpulmonata and found strong support for several major clades that have not been strongly supported in previous molecular studies. While suggested in previous studies (Dinapoli and Klussmann-Kolb, 2010; Jörger et al., 2010), this is the first molecular study to show unequivocal support for the clade here named Pylopulmonata, which includes the non-pulmonate Glacidorbidae and Pyramidellidae, and the traditionally pulmonate Amphiboloidea. This result has important implications for my understanding of pulmonate evolution as it confirms that Pulmonata is not monophyletic and that air-breathing has evolved multiple times (within Panpulmonata). Pylopulmonata unites the three panpulmonate lineages that retain an operculum as adults. It is unclear, however, whether the adult operculum – an operculum is found in the larvae of several other panpulmonate lineages – evolved secondarily in the Pylopulmonata or has been lost multiple times across the Heterobranchia.

While supported in a number of molecular studies (Dinapoli and Klussmann-Kolb, 2010; Jörger et al., 2010; Klussmann-Kolb et al., 2008), the Eupulmonata was not monophyletic in the two most recent studies to address these relationships (Dayrat et al., 2011; Romero et al., 2016). The Eupulmonata comprises intertidal and terrestrial pulmonates that share characteristics of the nervous system and a contractile pneumostome. The pneumostome is the opening of the 'pulmonate' lung and a contractile pneumostome allows more control over the lung for respiration and water storage. It is likely one of the key adaptations which allow terrestriality; therefore a non-monophyletic Eupulmonata changes our interpretation of when and how the major habitat transitions within Panpulmonata occurred. The molecular analyses in Chapter 3, however, show unequivocal support for a monophyletic Eupulmonata. These results show that a contractile pneumostome has only evolved three times: once in the common ancestor of the mostly terrestrial Eupulmonata, once within the Hygrophila, and once in the Siphonariidae in which both contractile and non-contractile forms are found. The results presented in Chapter 3 also represent the first molecular evidence to support the Geophila hypothesis. Geophila, which was first proposed by Férussac (1819), comprises the Systellommatophora and Stylommatophora and is supported by morphological evidence (Barker, 2001; Ponder and Lindberg, 2008).

Previous studies have suggested that the evolutionary relationships of the Stylommatophora remain largely unresolved due to a relatively rapid diversification event early in its history, based on a limited number of loci and the fossil record (Tillier et al., 1996; Wade et al., 2006, 2001). The analyses in Chapter 3 support the idea of a relatively rapid diversification within Stylommatophora as there is a lack of strong resolution, short internodes, and incongruence between different subsets of the data. The evolution of characters such as complex courtship behaviour and a calcareous egg may have allowed the Stylommatophora to radiate into a wider range of terrestrial habitats compared to other terrestrial lineages (Little, 1983). More detailed taxonomic sampling, however, is necessary to determine rates of diversification and test whether an adaptive radiation is likely to have occurred. The dating analyses presented in Chapter 3 also provide a framework for interpreting the timing of evolution within clades which have minimal representation in the fossil record.

With the continuing development and wide spread use of new sequencing technologies, it is more feasible than ever to conduct large-scale transcriptome-based phylogenomic studies. Increased taxonomic sampling, particularly of the Siphonariidae and the Stylommatophora, will allow us to evaluate these results and further investigate relationships with poor resolution. Denser taxonomic sampling will also allow us to assess the role of extinction verses adaptive radiation in the diversification patterns of this group (Crisp and Cook, 2009). By subsetting the alignment I was able to investigate the support for these results across different gene classes and different sized datasets. Additional analyses, however, may still provide a clearer picture, particularly within the Stylommatophora. While nearly 80% of the 500 orthologous genes were saturated across Panpulmonata it is likely that a large proportion of these genes are not saturated across the shallower Stylommatophora. It may therefore be informative to address the phylogenetic relationships within the Stylommatophora using nucleotide- or codon- based analyses. Many models of sequence evolution, such as GTR, assume homogeneity, stationarity, and time-reversibility. While deviation from homogeneity was not detected in my dataset, deviations from stationarity and time-reversibility may be present. Substitution models that do not make these assumptions are being developed, and hold promise for the field of molecular phylogenetics, particularly where these assumptions are likely incorrect (Kaehler et al., 2015).

The Rhytididae are a family of carnivorous land snails with a Gondwanan distribution (Chapter 4). Australia has the highest diversity of the Rhytididae; however, the relationships within this group had never been examined with detailed molecular evidence. Using the orthologous genes identified herein (Teasdale et al. 2016), I designed an exon capture probe set to capture sequences from genomic DNA to investigate the major relationships within the Australian Rhytididae. My analyses show strong support for four major divergent clades, most of which were not predicted by shell based taxonomy. I also showed that many of the Australian genera are polyphyletic or paraphyletic and that there is evidence for several new species. These findings will facilitate a future formal systematic revision of the group which will incorporate detailed morphological comparisons. To better place the Australian Rhytididae in context, higher taxonomic representation of the New Zealand, South African, and Pacific island Rhytididae is needed. The Pacific island snails have been thought of as potentially basal (Climo, 1977); this idea is not incongruent with my analyses, but would imply cross-water dispersal.

The two exon capture experiments presented in this thesis demonstrate that we can use the multi-locus datasets identified herein (Teasdale et al. 2016; Chapter 4) to design exon capture probe kits to sequence any family represented in my transcriptome datasets. Designing a new exon capture probe set from the alignments presented in this thesis is straight forward as they are already delimitated into exons and I have qualified the orthology of the exons at multiple taxonomic scales. While I have only tested exon capture designs for specific families, I did include multiple samples per probe design to capture broader sequence diversity. Studies such as Hugall et al. (2016) have extended this concept to produce probe designs which successfully capture sequences across a Class, namely the Ophiuroidea. By targeting a smaller number of exons, but including copies of each exon from many different families, it may be possible to produce a similar exon capture probe design which would capture sequence across, for example, all of the Stylommatophora. Such a design would greatly facilitate systematic research across the pulmonates.

Exon capture data can also be used to address shallow relationships within genera and species complexes. I had sufficient sequence variability to address the shallow phylogenetic relationships in the Australian Rhytididae (Chapter 4); however, an additional source of sequence variability lies in the flanking regions of the exons. A significant amount of the flanking regions are sequenced even though they are not specifically targeted. The Anchored Enrichment (Lemmon et al., 2012) and Ultraconserved Element (Faircloth et al., 2012) approaches to targeted enrichment essentially employ this approach by designing probes for highly conserved regions of the genome and analysing the adjacent, more variable sequence. This is similar to targeting relatively conserved exons within a genus or species complex and analysing the co-captured and more variable flanking sequences.

I was also able to capture sequence from genomic DNA from museum specimens. We only used samples preserved in ethanol in this research; however, there is also the potential to capture sequence from the large collections of formalin preserved specimens that exist for the pulmonates. While it is difficult to extract enough DNA from formalin preserved specimens for next generation sequencing, a number of studies have demonstrated success with formalin preserved specimens, which could be applied to exon capture (Carrick et al., 2015; Eijkelenboom et al., 2016).

A major theme throughout this thesis has been the detection and minimisation of paralogy within phylogenomic datasets. Undetected paralogy can mislead phylogenetic analysis and result in incorrect inference (Struck, 2013). Using a relatively manual approach of orthology determination, in conjunction with gene tree screening, I was able to identify a large set of nuclear genes from transcriptome sequences that are orthologous across the taxa sequenced (Teasdale et al. 2016). I further qualified orthology by sequencing this gene set from genomic

DNA. There was minimal evidence of unexpressed paralogs or pseudogenes within the samples sequenced in the two exon capture experiments. In contrast, some studies which only conducted a thorough search for paralogous sequences after exon capture sequencing had to discard a large proportion of the loci with undetected paralogs (Eytan et al., 2015). Whenever new taxa are sequenced, undetected lineage-specific duplications may be present. Duplications can happen at any taxonomic scale and thus I advocate that checks for paralogous sequences need to be conducted both before and after an exon capture experiment. Potentially paralogous sequences can be identified either by examining the alignments, screening for sequences with high levels of heterozygosity, or using an automated orthology determination method. Automated methods are essential, given the scale of next generation sequencing datasets, but algorithmic improvements are still needed. In Teasdale et al. (2016), I compared a manual approach to orthology determination with the automated pipeline Agalma (Dunn et al., 2013). The comparison highlighted that automated orthology detection algorithms that screen gene trees for orthologous subclades are susceptible to transcriptome assembly errors, which are common, especially if sequence coverage is not extremely high.

While the main focus of the research presented in this thesis was to identify orthologous genes for phylogenetics, I also identified many paralogous genes with duplications within the pulmonates (Teasdale et al., 2016). These genes are not useful for phylogenetic analyses but could potential be used as additional evidence for understanding the evolutionary relationships between the pulmonates. These paralogous genes may be evidence of duplications within the genome. Karyotype analyses have suggested that a genome duplication took place in the common ancestor of the Stylommatophora (Hallinan and Lindberg, 2011). I did not find evidence of polyploidy in my dataset as the number of unexpressed paralogs and pseudogenes in the exon capture sequencing was very low. It is still possible, however, that a partial duplication took place and further investigation of the paralogous genes may show evidence of such an event. An alternative explanation for larger genome size in the Stylommatophora, supported by the exon capture sequencing, is an expansion of the non-coding genomic regions. Both the Camaenidae and Rhytididae showed many additional exons relative to the non-pulmonate *Lottia gigantea*. More exon capture or whole genome sequencing would be needed to explore the pattern of exon boundary evolution across Panpulmonata. Additionally, many of the paralogous genes are involved in fundamental physiological pathways (e.g. most of the nuclear ribosomal genes are paralogous in the Eupulmonata; Chapter 2). Given that fundamental physiological changes are linked to different habitats and life histories, further exploration of important gene families across Panpulmonata could also provide insight into the evolution of morphological and physiological adaptation to different environments.

Literature Cited

Aberer, A.J., Kobert, K., Stamatakis, A., 2014. ExaBayes: massively parallel bayesian tree inference for the whole-genome era. Mol. Biol. Evol. 31, 2553–6. doi:10.1093/molbev/msu236

Ahlberg, P.E., Milner, A.R., 1994. The origin and early diversification of tetrapods. Nature 368, 507–514. doi:10.1038/368507a0

Altenhoff, A.M., Skunca, N., Glover, N., Sueki, A., Pilizota, I., Gori, K., Tomiczek, B., Muller, S., Redestig, H., Gonnet, G.H., Dessimoz, C., 2015. The OMA orthology database in 2015 : function predictions, better plant support, synteny view and other improvements. Nucleic Acids Res. 43, D240–D249. doi:10.1093/nar/gku1158

Ashkenazy, H., Penn, O., Doron-Faigenboim, A., Cohen, O., Cannarozzi, G., Zomer, O., Pupko, T., 2012. FastML: A web server for probabilistic reconstruction of ancestral sequences. Nucleic Acids Res. 40, 580–584. doi:10.1093/nar/gks498

Baker, H.B., 1955. Heterurethrous and Aulacopod. Nautilus (Philadelphia). 58, 109–112.

Barker, G.M., 2001. Gastropods on land: phylogeny, diversity and adaptive morphology, in: The Biology of Terrestrial Molluscs. pp. 1–146.

Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C., Good, J.M., 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics 13, 403. doi:10.1186/1471-2164-13-403

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 1–7. doi:10.1093/bioinformatics/btu170

Borowiec, M.L., 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4, e1660. doi:10.7717/peerj.1660

Bouchet, P., Rocroi, J.-P., 2005. Classification and Nomenclator of Gastropod Families. Malacologia 47, 1–397.

Bragg, J.G., Potter, S., Bi, K., Moritz, C., 2016. Exon capture phylogenomics: efficacy across scales of divergence. Mol. Ecol. Resour. 16, 1059–68. doi:10.1111/1755-0998.12449

Bruggen, A.C. Van, 1980. Gondwanaland connections in the terrestrial molluscs of Africa and Australia. J. Malacol. Soc. Aust. 4, 215–222. doi:10.1080/00852988.1980.10673930

Byrne, M., Steane, D.A., Joseph, L., Yeates, D.K., Jordan, G.J., Crayn, D., Aplin, K., Cantrill, D.J., Cook, L.G., Crisp, M.D., Keogh, J.S., Melville, J., Moritz, C., Porch, N., Sniderman, J.M.K., Sunnucks, P., Weston, P.H., 2011. Decline of a biome: Evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. J. Biogeogr. 38, 1635–1656. doi:10.1111/j.1365-2699.2011.02535.x

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

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. doi:10.1093/bioinformatics/btp348

Carrick, D.M., Mehaffey, M.G., Sachs, M.C., Altekruse, S., Camalier, C., Chuaqui, R., Cozen, W., Das, B., Hernandez, B.Y., Lih, C.J., Lynch, C.F., Makhlouf, H., McGregor, P., McShane, L.M., Rohan, J.P., Walsh, W.D., Williams, P.M., Gillanders, E.M., Mechanic, L.E., Schully, S.D., 2015. Robustness of next generation sequencing on older formalin-fixed paraffin-embedded tissue. PLoS One 10, 3–10. doi:10.1371/journal.pone.0127353

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552. doi:10.1093/oxfordjournals.molbev.a026334

Chatterjee, S., Scotese, C.R., 1999. The breakup of Gondwana and the evolution and biogeography of the Indian plate. Proc. Indian Acad. Sci. 65, 397–425.

Clarkson, M.O., Kasemann, S.A., Wood, R.A., Lenton, T.M., Daines, S.J., Richoz, S., Ohnemueller, F., Meixner, A., Poulton, S.W., Tipper, E.T., 2015. Ocean acidification and the Permo-Triassic mass extinction. Science (80-.). 348, 229–233.

Climo, F.M., 1977. A new higher classification of New Zealand Rhytididae (Mollusca: Pulmonata). J. R. Soc. New Zeal. 7, 59–65.

Cowie, R.H., Holland, B.S., 2006. Dispersal is fundamental to biogeography and the evolution of biodiversity on oceanic islands. J. Biogeogr. 33, 193–198. doi:10.1111/j.1365-2699.2005.01383.x

Crisp, M.D., Cook, L.G., 2009. Explosive radiation or cryptic mass extinction? interpreting signatures in molecular phylogenies. Evolution (N. Y). 63, 2257–2265. doi:10.1111/j.1558-5646.2009.00728.x

Cuvier, G., 1817. Le Régne Animal. Libraire Déterville, Paris.

Dávalos, L.M., Cirranello, A.L., Geisler, J.H., Simmons, N.B., 2012. Understanding phylogenetic incongruence: lessons from phyllostomid bats., Biological reviews of the Cambridge Philosophical Society. doi:10.1111/j.1469-185X.2012.00240.x

Dayrat, B., Conrad, M., Balayan, S., White, T.R., Albrecht, C., Golding, R., Gomes, S.R., Harasewych, M.G., Martins, A.M.D.F., 2011. Phylogenetic relationships and evolution of pulmonate gastropods (Mollusca): new insights from increased taxon sampling. Mol. Phylogenet. Evol. 59, 425–37. doi:10.1016/j.ympev.2011.02.014

Dayrat, B., Tillier, S., 2002. Evolutionary relationships of euthyneuran gastropods (Mollusca): a cladistic re-evaluation of morphological characters. Zool. J. Linn. Soc. 135, 403–470.

Dinapoli, A., Klussmann-Kolb, A., 2010. The long way to diversity - phylogeny and evolution of the Heterobranchia (Mollusca: Gastropoda). Mol. Phylogenet. Evol. 55, 60–76. doi:10.1016/j.ympev.2009.09.019

Dinapoli, A., Zinssmeister, C., Klussmann-Kolb, A., 2011. New insights into the phylogeny of the pyramidellidae (Gastropoda). J. Molluscan Stud. 77, 1–7. doi:10.1093/mollus/eyq027

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973. doi:10.1093/molbev/mss075

Dunn, C.W., Howison, M., Zapata, F., 2013. Agalma: an automated phylogenomics workflow. BMC Bioinformatics 14, 330. doi:10.1186/1471-2105-14-330

Ebersberger, I., Strauss, S., von Haeseler, A., 2009. HaMStR: profile hidden markov model based search for orthologs in ESTs. BMC Evol. Biol. 9, 157. doi:10.1186/1471-2148-9-157

Edwards, S. V., 2016. Phylogenomic subsampling: a brief review. Zool. Scr. 45, 63–74. doi:10.1111/zsc.12210

Efford, M., Howitt, R., Gleeson, D., 2002. Phylogenetic relationships of Wainuia (Mollusca: Pulmonata) - biogeography and conservation implications. J. R. Soc. New Zeal. 32, 445–456.

Eijkelenboom, A., Kamping, E., Kastner-van Raaij, A., Hendriks-Cornelissen, S., Neveling, K., Kuiper, R., Hoischen, A., Nelen, M., Ligtenberg, M., Tops, B., 2016. Reliable Next-Generation Sequencing of Formalin-Fixed, Paraffin-Embedded Tissue Using Single Molecule Tags. J. Mol. Diagnostics 18, 851–863. doi:10.1016/j.jmoldx.2016.06.010

Engel, M.S., Grimaldi, D.A., 2004. New light shed on the oldest insect. Nature 427, 627–630. doi:10.1038/nature02334.1.

Enright, A.J., Van Dongen, S., Ouzounis, C.A., 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30, 1575–1584.

Eytan, R.I., Evans, B.R., Dornburg, A., Lemmon, A.R., Lemmon, E.M., Wainwright, P.C., Near, T.J., 2015. Are 100 enough? Inferring acanthomorph teleost phylogeny using Anchored Hybrid Enrichment. BMC Evol. Biol. 15, 113. doi:10.1186/s12862-015-0415-0

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61, 717–726. doi:10.1093/sysbio/sys004

Feldmeyer, B., Wheat, C.W., Krezdorn, N., Rotter, B., Pfenninger, M., 2011. Short read Illumina data for the de novo assembly of a non-model snail species transcriptome (Radix balthica, Basommatophora, Pulmonata), and a comparison of assembler performance. BMC Genomics 12, 317. doi:10.1186/1471-2164-12-317

Férussac, A.E.J., 1819. Histoire naturelle des pulmonés sans opercules. Bertrand, Paris.

Fitch, W.M., 1970. Distinguishing Homologous from Analogous proteins. Syst. Zool. 19, 99–113.

Fitch, W.M., 2000. Homology: a personal view on some of the problems. Trends Genet. 16, 227–231.

García-Sandoval, R., 2014. Why some clades have low bootstrap frequencies and high Bayesian posterior probabilities. Isr. J. Ecol. Evol. 60, 41–44. doi:10.1080/15659801.2014.937900

Gittenberger, E., Groenenberg, D.S.J., Kokshoorn, B., Preece, R.C., 2006. Biogeography: molecular trails from hitch-hiking snails. Nature 439, 409. doi:10.1038/439409a

Golding, R.E., Ponder, W.F., Byrne, M., 2010. Taxonomy and anatomy of Amphiboloidea (Gastropoda: Heterobranchia: Archaeopulmonata). Zootaxa 50, 1–2.

Gontcharov, A.A., Marin, B., Melkonian, M., 2004. Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta). Mol. Biol. Evol. 21, 612–24. doi:10.1093/molbev/msh052

Gordon, M.S., Olson, E.C., 1995. Invasions of the land: the transitions of organisms from aquatic to terrestrial life. Columbia University Press, New York.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Palma, F., Birren, B.W., Nusbaum, C., Lindblad-toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. doi:10.1038/nbt.1883

Grande, C., Templado, J., Cervera, J.L., Zardoya, R., 2004. Molecular phylogeny of Euthyneura (Mollusca: Gastropoda). Mol. Biol. Evol. 21, 303–313. doi:10.1093/molbev/msh016

Grande, C., Templado, J., Zardoya, R., 2008. Evolution of gastropod mitochondrial genome arrangements. BMC Evol. Biol. 15, 1–15. doi:10.1186/1471-2148-8-61

Grigoriev, I. V, Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R.A., Otillar, R., Poliakov, A., Ratnere, I., Riley, R., Smirnova, T., Rokhsar, D., Dubchak, I., 2012. The genome portal of the Department of Energy Joint Genome Institute. Nucleic Acids Res. 40, D26–D32.

doi:10.1093/nar/gkr947

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512. doi:10.1038/nprot.2013.084

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.

Hallinan, N.M., Lindberg, D.R., 2011. Comparative analysis of chromosome counts infers three paleopolyploidies in the Mollusca. Genome Biol. Evol. 3, 1150–1163. doi:10.1093/gbe/evr087

Haszprunar, G., 1985. The Heterobranchia - a new concept of the phylogeny of the higher Gastropoda. J. Zool. Syst. Evol. Res. 23, 15–37.

Haszprunar, G., 1988. On the origin and evolution of major gastropod groups, with special reference to the Streptoneura (Mollusca). J. Molluscan Stud. 54, 367–441.

Haszprunar, G., Huber, G., 1990. On the central nervous system of Smeagolidae and Rhodopidae, two families questionably allied with the Gymnomorpha (Gastropoda: Euthyneura). J. Zool. 220, 185–199. doi:10.1111/j.1469-7998.1990.tb04302.x

Hausdorf, B., 1998. Phylogeny of the Limacoidea Sensu Lato (Gastropoda: Stylommatophora). J. Molluscan Stud. 64, 35–66. doi:10.1093/mollus/64.1.35

Herbert, A.D.G., Moussalli, A., 2010. Revision of the Larger Cannibal Snails (Natalina s. l.) of Southern Africa — Natalina s. s., Afrorhytida and Capitina (Mollusca: Gastropoda: Rhytididae) 51, 1–132.

Herbert, D.G., Moussalli, A., 2016. Revision of the dwarf cannibal snails (Nata s.l.) of southern Africa—Nata s.s. and Natella (Mollusca: Gastropoda: Rhytididae), with description of three new species.

Herbert, D.G., Moussalli, A., Griffiths, O.L., 2015. Rhytididae (Eupulmonata) in Madagascar: reality or conjecture? J. Molluscan Stud. 81, 1–10.

doi:10.1093/mollus/eyu088

Holznagel, W.E., Colgan, D.J., Lydeard, C., 2010. Pulmonate phylogeny based on 28S rRNA gene sequences: a framework for discussing habitat transitions and character transformation. Mol. Phylogenet. Evol. 57, 1017–25. doi:10.1016/j.ympev.2010.09.021

Homer, N., Merriman, B., Nelson, S.F., 2009. BFAST: an alignment tool for large scale genome resequencing. PLoS One 4, e7767. doi:10.1371/journal.pone.0007767

Hubendick, B., 1979. Systematics and comparative morphology of the Basommatophora, in: Pulmonates. pp. 1–48.

Hugall, A.F., O’Hara, T.D., Hunjan, S., Nilsen, R., Moussalli, A., 2016. An Exon-Capture System for the Entire Class Ophiuroidea. Mol. Biol. Evol. 33, 281–94. doi:10.1093/molbev/msv216

Hugall, A.F., Stanisic, J., 2011. Beyond the prolegomenon: a molecular phylogeny of the Australian camaenid land snail radiation. Zool. J. Linn. Soc. 161, 531–572. doi:10.1111/j.1096-3642.2010.00644.x

Iredale, T., 1933. Systematic notes on Australian land shells. Rec. Aust. Museum 19, 37–59. doi:10.3853/j.0067-1975.19.1933.690

Jensen, K.R., 2011. Comparative morphology of the mantle cavity organs of shelled sacoglossa, with a discussion of relationships with other heterobranchia. Thalassas 27, 169–192.

Jörger, K.M., Brenzinger, B., Neusser, T.P., Alexander, V., 2014. Panpulmonate habitat transitions: tracing the evolution of Acochlidia (Heterobranchia , Gastropoda). bioRxiv. doi:https://doi.org/10.1101/010322

Jörger, K.M., Stöger, I., Kano, Y., Fukuda, H., Knebelsberger, T., Schrödl, M., 2010. On the origin of Acochlidia and other enigmatic euthyneuran gastropods, with implications for the systematics of Heterobranchia. BMC Evol. Biol. 10, 323. doi:10.1186/1471-2148-10-323

Kaehler, B.D., Yap, V.B., Zhang, R., Huttley, G. a., 2015. Genetic Distance for a General Non-Stationary Markov Substitution Process. Syst. Biol. 64, 281–293. doi:10.1093/sysbio/syu106

Kaim, A., 2004. The evolution of conch ontogeny in Mesozoic open sea gastropods. Palaentologia Pol. 62, 3–183.

Kainer, D., Lanfear, R., 2015. The effects of partitioning on phylogenetic inference. Mol. Biol. Evol. 32, 1611–1627. doi:10.1093/molbev/msv026

Kano, Y., Neusser, T.P., Fukumori, H., Jörger, K.M., Schrödl, M., 2015. Sea-slug invasion of the land. Biol. J. Linn. Soc. 116, 253–259. doi:10.1111/bij.12578

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. 30, 772–780. doi:10.1093/molbev/mst010

Kelley, L., Gardner, S., Sutcliffe, M., 1996. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally-related subfamilies. Protein Eng 9, 1063–1065.

Klussmann-Kolb, A., Dinapoli, A., Kuhn, K., Streit, B., Albrecht, C., 2008. From sea to land and beyond--new insights into the evolution of euthyneuran Gastropoda (Mollusca). BMC Evol. Biol. 8, 57. doi:10.1186/1471-2148-8-57

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., Mclellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., Wilson, R.K., 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22, 568–576. doi:10.1101/gr.129684.111

Kocot, K.M., Cannon, J.T., Todt, C., Citarella, M.R., Kohn, A.B., Meyer, A., Santos, S.R., Schander, C., Moroz, L.L., Lieb, B., Halanych, K.M., 2011. Phylogenomics reveals deep molluscan relationships. Nature 477, 452–456. doi:10.1038/nature10382

Kocot, K.M., Citarella, M.R., Moroz, L.L., Halanych, K.M., 2013a. PhyloTreePruner: A Phylogenetic Tree-Based Approach for Selection of Orthologous Sequences for Phylogenomics. Evol. Bioinform. Online 9, 429–35. doi:10.4137/EBO.S12813

Kocot, K.M., Halanych, K.M., Krug, P.J., 2013b. Phylogenomics supports Panpulmonata: Opisthobranch paraphyly and key evolutionary steps in a major radiation of gastropod molluscs. Mol. Phylogenet. Evol. 69, 764–771. doi:10.1016/j.ympev.2013.07.001

Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B.M., Wägele, J.W., Misof, B., 2010. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. Front. Zool. 7, 10. doi:10.1186/1742-9994-7-10

Kück, P., Struck, T.H., 2014. BaCoCa - A heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. Mol. Phylogenet. Evol. 70, 94–98. doi:10.1016/j.ympev.2013.09.011

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. Mol. Biol. Evol. 29, 1695–1701. doi:10.1093/molbev/mss020

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., Stamatakis, A., 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14, 82. doi:10.1186/1471-2148-14-82

Lawver, L.A., Gahagan, L.M., 2003. Evolution of cenozoic seaways in the circum-antarctic region. Palaeogeogr. Palaeoclimatol. Palaeoecol. 198, 11–37. doi:10.1016/S0031-0182(03)00392-4

Le Renard, J., 1983. Mise en evidence d’algueraies a caulerpa par les Juliidae (Gasteropodes a 2 valves: Sacoglossa) dans L'eocene du bassin de paris. Geobios 16, 39–51.

Le, S.Q., Dang, C.C., Gascuel, O., 2012. Modeling protein evolution with several amino acid replacement matrices depending on site rates. Mol. Biol. Evol. 29, 2921–2936. doi:10.1093/molbev/mss112

Leaché, A.D., Rannala, B., 2011. The accuracy of species tree estimation under simulation: a comparison of methods. Syst. Biol. 60, 126–37. doi:10.1093/sysbio/syq073

Lee, M.S.Y., Hugall, a. F., 2003. Partitioned Likelihood Support and the Evaluation of Data Set Conflict. Syst. Biol. 52, 15–22. doi:10.1080/10635150390132650

Lemmon, A.R., Emme, S., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61, 727–744.

Lemmon, E.M., Lemmon, A.R., 2013. High-Throughput Genomic Data in Systematics and Phylogenetics. Annu. Rev. Ecol. Evol. Syst. 44, 99–121. doi:10.1146/annurev-ecolsys-110512-135822

Li, C., Hofreiter, M., Straube, N., Corrigan, S., Naylor, G.J.P., 2013. Capturing protein-coding genes across highly divergent species. Biotechniques 54, 321–6. doi:10.2144/000114039

Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993. doi:10.1093/bioinformatics/btr509

Li, L., Stoeckert, C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189. doi:10.1101/gr.1224503

Little, C., 1983. The colonisation of land. Cambridge University Press, Cambridge.

Lohse, M., Bolger, A.M., Nagel, A., Fernie, A.R., Lunn, J.E., Stitt, M., Usadel, B., 2012. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res. 40, W622–W627. doi:10.1093/nar/gks540

Lydeard, C., Cowie, R.H., Ponder, W.F., Bogan, A.E., Bouchet, P., Clark, S.A., Cummings, K.S., Frest, T.J., Gargominy, O., Herbert, D.G., Hershler, R., Perez, K.E., Roth, B., Seddon, M., Strong, E.E., Thompson, F.G., 2010. The Global Decline of Nonmarine Mollusks. Bioscience 54, 321–330.

Maddison, W.P., Maddison, D.R., 2016. Mesquite: a modular system for evolutionary analysis [WWW Document]. URL http://mesquiteproject.org

Martin, A.P., Burg, T.M., 2002. Perils of paralogy: using HSP70 genes for inferring organismal phylogenies. Syst. Biol. 51, 570–587. doi:10.1080/10635150290069995

McCormack, J.E., Harvey, M.G., Faircloth, B.C., Crawford, N.G., Glenn, T.C., Brumfield, R.T., 2013. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS One 8, e54848. doi:10.1371/journal.pone.0054848

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M., 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. doi:10.1101/gr.107524.110.20

Medina, M., Lal, S., Vallès, Y., Takaoka, T.L., Dayrat, B. a, Boore, J.L., Gosliner, T., 2011.

Crawling through time: Transition of snails to slugs dating back to the Paleozoic, based on mitochondrial phylogenomics. Mar. Genomics 4, 51–9. doi:10.1016/j.margen.2010.12.006

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspock, U., Aspock, H., Bartel, D., Blanke, A., Berger, S., Bohm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schutte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K.F., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Xu, X., Yang, H., Wang, J., Kjer, K.M., Zhou, X., 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science (80-.). 346, 763–767. doi:10.1126/science.1257570

Moritz, C., Richardson, K.S., Ferrier, S., Monteith, G.B., Stanisic, J., Williams, S.E., Whiffin, T., 2001. Biogeographical concordance and efficiency of taxon indicators for establishing conservation priority in a tropical rainforest biota. Proc. Biol. Sci. 268, 1875–81. doi:10.1098/rspb.2001.1713

Moussalli, A., Herbert, D.G., 2016. Deep molecular divergence and exceptional morphological stasis in dwarf cannibal snails Nata sensu lato Watson, 1934 (Rhytididae) of southern Africa. Mol. Phylogenet. Evol. 95, 100–115. doi:10.1016/j.ympev.2015.11.003

Moussalli, A., Herbert, D.G., Stuart-Fox, D., 2009. A phylogeny of the cannibal snails of southern Africa, genus Natalina sensu lato (Pulmonata: Rhytididae): assessing concordance between morphology and molecular data. Mol. Phylogenet. Evol. 52, 167–82. doi:10.1016/j.ympev.2009.02.018

O’Hara, T.D., Hugall, A.F., Thuy, B., Moussalli, A., 2014. Phylogenomic resolution of the class Ophiuroidea unlocks a global microfossil record. Curr. Biol. 24, 1874–1879. doi:10.1016/j.cub.2014.06.060

Oakley, T.H., Wolfe, J.M., Lindgren, A.R., Zaharoff, A.K., 2012. Phylotranscriptomics to Bring the Understudied into the Fold: Monophyletic Ostracoda, Fossil Placement, and Pancrustacean Phylogeny. Mol. Biol. Evol. 30, 215–233. doi:10.1093/molbev/mss216

Oliver, P.M., Sanders, K.L., 2009. Molecular evidence for Gondwanan origins of multiple lineages within a diverse Australasian gecko radiation. J. Biogeogr. 36, 2044–2055. doi:10.1111/j.1365-2699.2009.02149.x

Ostlund, G., Schmitt, T., Forslund, K., Köstler, T., Messina, D.N., Roopra, S., Frings, O., Sonnhammer, E.L.L., 2010. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. Nucleic Acids Res. 38, D196–203. doi:10.1093/nar/gkp931

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J., Quackenbush, J., 2003. TIGR gene indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. Bioinformatics 19, 651–652. doi:10.1093/bioinformatics/btg034

Philippe, H., Brinkmann, H., Lavrov, D. V, Littlewood, D.T.J., Manuel, M., Wörheide, G., Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9, e1000602. doi:10.1371/journal.pbio.1000602

Pilsbry, H.A., 1900. On the Zoölogical Position of Partula and Achatinella. Proc. Acad. Nat. Sci. Philadelphia 52, 561–567.

Pirie, M.D., Vargas, M.P.B., Botermans, M., Bakker, F.T., Chatrou, L.W., 2007. Ancient paralogy in the cpDNA trnL-F region in Annonaceae: implications for plant molecular systematics. Am. J. Bot. 94, 1003–1016.

Ponder, W., Lindberg, D.R., 2008. Phylogeny and Evolution of the Mollusca. University of California Press, Berkeley.

Ponder, W.F., 1986. Glacidorbidae (Glacidorbacea: Basommatophora), a new family and superfamily of operculate freshwater gastropods. Zool. J. Linn. Soc. 87, 53–83. doi:10.1111/j.1096-3642.1986.tb01330.x

Ponder, W.F., Avern, G.J., 2000. The Glacidorbidae (Mollusca: Gastropoda: Heterobranchia) of Australia. Rec. Aust. Museum 52, 307–353.

Ponder, W.F., Lindberg, D.R., 1997. Towards a phylogeny of gastropod molluscs: an analysis using morphological characters. Zool. J. Linn. Soc. 119, 83–265.

Qiu, H., Yang, E.C., Bhattacharya, D., Yoon, H.S., 2012. Ancient gene paralogy may mislead inference of plastid phylogeny. Mol. Biol. Evol. 29, 3333–43. doi:10.1093/molbev/mss137

Ranwez, V., Delsuc, F., Ranwez, S., Belkhir, K., Tilak, M.-K., Douzery, E.J., 2007. OrthoMaM: a database of orthologous genomic markers for placental mammal phylogenetics. BMC Evol. Biol. 7, 241. doi:10.1186/1471-2148-7-241

Ranwez, V., Harispe, S., Delsuc, F., Douzery, E.J.P., 2011. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. PLoS One 6, e22594. doi:10.1371/journal.pone.0022594

Remm, M., Storm, C.E. V, Sonnhammer, E.L.L., 2001. Automatic Clustering of Orthologs and In-paralogs from Pairwise Species Comparisons. J. Mol. Biol. 314, 1041–1052. doi:10.1006/jmbi.2001.5197

Rix, M.G., Harvey, M.S., 2012. Phylogeny and historical biogeography of ancient assassin spiders (Araneae: Archaeidae) in the Australian mesic zone: Evidence for Miocene speciation within Tertiary refugia. Mol. Phylogenet. Evol. 62, 375–396. doi:10.1016/j.ympev.2011.10.009

Romero, P.E., Weigand, A.M., Pfenninger, M., 2016. Positive selection on panpulmonate mitogenomes provide new clues on adaptations to terrestrial life. BMC Evol. Biol. 16, 1–13. doi:10.1186/s12862-016-0735-8

Rosenberg, G., 2014. A New Critical Estimate of Named Species-Level Diversity of the Recent Mollusca. Am. Malacol. Bull. 32, 308–322.

Rothfels, C.J., Larsson, A., Li, F.-W., Sigel, E.M., Huiet, L., Burge, D.O., Ruhsam, M., Graham, S.W., Stevenson, D.W., Wong, G.K.-S., Korall, P., Pryer, K.M., 2013. Transcriptome-mining for single-copy nuclear markers in ferns. PLoS One 8, e76957. doi:10.1371/journal.pone.0076957

Sadamoto, H., Takahashi, H., Okada, T., Kenmoku, H., Toyota, M., Asakawa, Y., 2012. De novo sequencing and transcriptome analysis of the central nervous system of mollusc Lymnaea stagnalis by deep RNA sequencing. PLoS One 7, e42546. doi:10.1371/journal.pone.0042546

Salvador, R.B., Simone, R.L. De, 2013. Taxonomic revision of the fossil pulmonate mollusks of Itaboraí basin (paleocene), Brazil, Papéis Avulsos de Zoologia. doi:10.1590/S0031-10492013000200001

Schrödl, M., 2014. Time to say "Bye-bye Pulmonata"? Spixiana 37, 161–164.

Schrödl, M., Km, J., Ng, W., 2011. Bye Bye "Opisthobranchia"! A review on the contribution of mesopsammic sea slugs to euthyneuran systematics. Thalassas 27, 101–112.

Schwarz, M.P., Fuller, S., Tierney, S.M., Cooper, S.J.B., 2006. Molecular phylogenetics of the exoneurine allodapine bees reveal an ancient and puzzling dispersal from Africa to Australia. Syst. Biol. 55, 31–45. doi:10.1080/10635150500431148

Seton, M., Müller, R.D., Zahirovic, S., Gaina, C., Torsvik, T., Shephard, G., Talsma, A., Gurnis, M., Turner, M., Maus, S., Chandler, M., 2012. Global continental and ocean basin reconstructions since 200Ma. Earth-Science Rev. 113, 212–270. doi:10.1016/j.earscirev.2012.03.002

Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–7.

Simakov, O., Marletaz, F., Cho, S.-J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., Kuo, D.-H., Larsson, T., Lv, J., Arendt, D., Savage, R., Osoegawa, K., de Jong, P., Grimwood, J., Chapman, J.A., Shapiro, H., Aerts, A., Otillar, R.P., Terry, A.Y., Boore, J.L., Grigoriev, I. V, Lindberg, D.R., Seaver, E.C., Weisblat, D.A., Putnam, N.H., Rokhsar, D.S., 2013. Insights into bilaterian evolution from three spiralian genomes. Nature 493, 526–531. doi:10.1038/nature11696

Simmons, M.P., Goloboff, P.A., 2014. Dubious resolution and support from published sparse supermatrices: The importance of thorough tree searches. Mol. Phylogenet. Evol. 78, 334–348. doi:10.1016/j.ympev.2014.06.002

Slater, G.S.C., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6, 31. doi:10.1186/1471-2105-6-31

Smith, B.J., 1979. Notes on two species of rhytidid snails from Lizard Island, North Queensland. Rec. Aust. Museum 32, 421–434. doi:10.3853/j.0067-1975.32.1979.469

Smith, B.J., 1992. Non-Marine Mollusca, in: Houston, W.W.K. (Ed.), Zoological Catalogue of Australia. Australian Government Publishing Service, Canberra, p. 405.

Smith, S.A., Wilson, N.G., Goetz, F.E., Feehery, C., Andrade, S.C.S., Rouse, G.W., Giribet, G., Dunn, C.W., 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. Nature 480, 364–367. doi:10.1038/nature10526

Solem, A., 1959. Systematics and zoogeography of the land and fresh-water mollusca of the new hebrides. Fieldiana Zool. 43.

Solem, A., 1979. Classification of the Land Mollusca, in: Fretter, V., Peake, J. (Eds.), Pulmonates. Academic Press, pp. 49–98.

Solem, A., 1992. Camaenid land snails from southern and eastern Australia, excluding Kangaroo Island. Rec. South Aust. Museum, Monogr. Ser. 2, 1–425.

Solem, A., Yochelson, E.L., 1979. North American Paleozoic land snails, with a summary of other Paleozoic non-marine snails. Geol. Surv. Prof. Pap. 1072, 1–42.

Spencer, H.G., Brook, F.J., Kennedy, M., 2006. Phylogeography of Kauri Snails and their allies from Northland, New Zealand (Mollusca: Gastropoda: Rhytididae: Paryphantinae). Mol. Phylogenet. Evol. 38, 835–42. doi:10.1016/j.ympev.2005.10.015

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690. doi:10.1093/bioinformatics/btl446

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

Stanisic, J., Shea, M., Potter, D., Griffiths, O., 2010. Australian Land Snails Volume 1: A field guide to eastern Australian species. Bioculture Press, Mauritius.

Struck, T.H., 2013. The impact of paralogy on phylogenomic studies - a case study on annelid relationships. PLoS One 8, e62892.

Struck, T.H., 2014. TreSpEx—Detection of Misleading Signal in Phylogenetic Reconstructions Based on Tree Information. Evol. Bioinforma. 10, 51–67. doi:10.4137/EBO.S14239.Received

Struck, T.H., Nesnidal, M.P., Purschke, G., Halanych, K.M., 2008. Detecting possibly saturated positions in 18S and 28S sequences and their influence on phylogenetic reconstruction of Annelida (Lophotrochozoa). Mol. Phylogenet. Evol. 48, 628–645. doi:10.1016/j.ympev.2008.05.015

Stworzewicz, E., Szulc, J., Pokryszko, B.M., Stworzewicz, E.W.A., Szulc, J., Pokryszko, B.M., 2009. Late Paleozoic Continental Gastropods from Poland : Systematic , Evolutionary and Paleoecological Approach. J. Paleontol. 83, 938–945.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E. V, Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A. V, Vasudevan, S., Wolf, Y.I., Yin, J.J., Natale, D.A., 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4, 41. doi:10.1186/1471-2105-4-41

Teasdale, L.C., Köhler, F., Murray, K.D., O'Hara, T.D., Moussalli, A., 2016. Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon-capture. Mol. Ecol. Resour. 1–39. doi:10.1111/1755-0998.12552

Thomaz, D., Guiller, a, Clarke, B., 1996. Extreme divergence of mitochondrial DNA within species of pulmonate land snails. Proc. Biol. Sci. 263, 363–8. doi:10.1098/rspb.1996.0056

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

Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192. doi:10.1093/bib/bbs017

Tillier, S., 1984. Relationships of gymnomorph gastropods (Mollusca: gastropoda). Zool. J. Linn. Soc. 82, 345–362. doi:10.1111/j.1096-3642.1984.tb00869.x

Tillier, S., Masselot, M., Tillier, A., 1996. Phylogenetic relationships of the pulmonate gastropods from rRNA sequences, and tempo and age of the stylommatophoran radiation, in: Taylor, J. (Ed.), Origin and Evolutionary Radiation of the Mollusca. Oxford University Press, pp. 267–284.

Tracey, S., Todd, J.A., Erwin, D.H., 1993. Mollusca: Gastropoda, in: Benton, M.J. (Ed.), The Fossil Record, Vol. 2. Chapman and Hall, London, pp. 131–167.

Upchurch, P., 2008. Gondwanan break-up: legacies of a lost world? Trends Ecol. Evol. 23, 229–236. doi:10.1016/j.tree.2007.11.006

Wade, C.M., Hudelot, C., Davison, A., Naggs, F., Mordan, P.B., 2007. Molecular phylogeny of the helicoid land snails (Pulmonata: Stylommatophora: Helicoidea), with special emphasis on the Camaenidae. J. Molluscan Stud. 73, 411–415. doi:10.1093/mollus/eym030

Wade, C.M., Mordan, P.B., Clarke, B., 2001. A phylogeny of the land snails (Gastropoda: Pulmonata). Proc. Biol. Sci. 268, 413–22. doi:10.1098/rspb.2000.1372

Wade, C.M., Mordan, P.B., Naggs, F., 2006. Evolutionary relationships among the Pulmonate land snails and slugs (Pulmonata, Stylommatophora). Biol. J. Linn. Soc. 87, 593–610.

Wägele, H., Deusch, O., Händeler, K., Martin, R., Schmitt, V., Christa, G., Pinzger, B., Gould, S.B., Dagan, T., Klussmann-Kolb, A., Martin, W., 2011. Transcriptomic evidence that longevity of acquired plastids in the photosynthetic slugs Elysia timida and Plakobranchus ocellatus does not entail lateral transfer of algal nuclear genes. Mol. Biol. Evol. 28, 699–706. doi:10.1093/molbev/msq239

Ward, N., Moreno-Hagelsieb, G., 2014. Quickly Finding Orthologs as Reciprocal Best Hits with BLAT, LAST, and UBLAST: How Much Do We Miss? PLoS One 9, e101850. doi:10.1371/journal.pone.0101850

Waterhouse, R.M., Tegenfeldt, F., Li, J., Zdobnov, E.M., Kriventseva, E. V., 2013. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Res. 41, D358–D365. doi:10.1093/nar/gks1116

White, T.R., Conrad, M.M., Tseng, R., Balayan, S., Golding, R., de Frias Martins, A.M., Dayrat, B. a, 2011. Ten new complete mitochondrial genomes of pulmonates (Mollusca: Gastropoda) and their impact on phylogenetic relationships. BMC Evol. Biol. 11, 295. doi:10.1186/1471-2148-11-295

Wortley, A.H., Rudall, P.J., Harris, D.J., Scotland, R.W., 2005. How much data are needed to resolve a difficult phylogeny?: case study in Lamiales. Syst. Biol. 54, 697–709. doi:10.1080/10635150500221028

Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., Chen, J., Chen, S., 2012. FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads. PLoS One 7, e52249. doi:10.1371/journal.pone.0052249

Yang, Z., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591. doi:10.1093/molbev/msm088

Zapata, F., Wilson, N.G., Howison, M., Andrade, S.C.S., Jörger, K.M., Goetz, F.E., Giribet, G., Dunn, C.W., 2014. Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. Proc. R. Soc. B Biol. Sci. 281, 20141739.

Author/s:
Teasdale, Luisa Cinzia

Title:
Phylogenomics of the pulmonate land snails

Date:
2017

Persistent Link:
http://hdl.handle.net/11343/128240

File Description:
Phylogenomics of the pulmonate land snails