Phylogeny and biogeography of the camaenid land snails of eastern Australia

Andrew Forrest Hugall

PhD Thesis

School of Earth and Environmental Sciences

The University of Adelaide

April 2011

Declaration

This work contains no material that previously has been accepted for the award of any degree in any institution by anyone, and is the result of my own endeavours except where duly acknowledged.

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

The author acknowledges that copyright of published works contained within this thesis (as listed below) resides with the copyright holders of those works.

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

Andrew Forrest Hugall

Hugall AF & Stanisic J. 2011. Beyond the prolegomenon: a molecular phylogeny of the Australian camaenid land snail radiation. *Zoological Journal of the Linnean Society* 161: 531-572. Copyright © 2010 The Linnean Society of London, Zoological Journal of the Linnean Society.

Hugall AF, Moritz C, Moussalli A & Stanisic J. 2002. Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences USA* 99: 6112-6117. Copyright © 2002, The National Academy of Sciences

Hugall AF, Stanisic J & Moritz C. 2003. Phylogeography of terrestrial mollusks: the case of the Sphaerospira lineage and history of Queensland rainforests. In: Lydeard C, Lindberg D. eds. *Molecular Systematics and Phylogeography of Mollusks*. Smithsonian Institution Press, Washington DC, pp. 270–301. Copyright © 2003 Smithsonian Institution.

Contents

Statement of Author Contributions

Chapter 2: Hugall AF & Stanisic J. 2011. Beyond the prolegomenon: a molecular phylogeny of the Australian camaenid land snail radiation. *Zoological Journal of the Linnean Society* 161: 531-572. The authors jointly conceived the project. AFH obtained the genetic data, executed all the analyses and wrote the text. JS collected many of the specimens, provided access to the museum collections, was responsible for identifying all specimens and nomenclature, and contributed to the discussion on taxonomic implications.

Chapter 3: Hugall AF, Moritz C, Moussalli A & Stanisic J. 2002. Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences USA* 99: 6112-6117. AFH and CM jointly conceived the project and wrote the text. AFH obtained and executed all the genetic data and analyses. JS provided access to the museum specimens and distribution database, and was responsible for identifying all specimens. AM and AFH jointly developed and conducted the GIS analyses.

Chapter 4: Hugall AF, Stanisic J & Moritz C. 2003. Phylogeography of terrestrial mollusks: the case of the Sphaerospira lineage and history of Queensland rainforests. In: Lydeard C, Lindberg D. eds. *Molecular Systematics and Phylogeography of Mollusks*. Smithsonian Institution Press, Washington DC, pp. 270–301. AFH conceived the project, obtained all the genetic data, conducted all the analyses and wrote the text. JS provided access to the museum specimens and databases, and was responsible for identifying all specimens and nomenclature. CM provided resources, editorial advice and discussion.

Andrew Forrest Hugall

John Stanisic

Craig Moritz

Adnan Moussalli

Summary

Camaenid land snails are the subject of this study, the object of which is the historical biogeography of eastern Australia, with phylogenetic diversity the *modus operandi*. The thesis is structured around two main sources of information: the distribution data and the phylogenetic data, and makes an empirical contribution to both. Beyond this are inferences on the major evolutionary processes governing biodiversity as seen from this large – or macroecological – perspective.

The first chapter introduces three themes: the mosaic of forests along eastern Australia, the camaenid snails and macroecological methods. It also introduces the distribution dataset, issues concerning analysis and provides a picture of the region and the snails' place in this – the *mise en scène*.

Having stated the case for the need for molecular phylogenetic data, the second chapter describes in detail the gathering and assembly of such data into a comprehensive phylogeny for the snails. In doing so it makes a substantial contribution to the higher taxonomy of the whole group, and to the methodology of supermatrix 'tree-of-life' phylogenetics. The results suggest that the camaenids stem from Oligo-Miocene Laurasian immigration. Relictual endemics indicate that many ancestral lineages were in place before the major decline of the mesic forest realm. Polyphyly of many genera highlight the repeated radiation of shell forms and that the higher taxonomy was unacceptable. The results provide the basis of a new generic framework.

The subsequent chapters use the resources now provided to investigate historical and environmental influences on biodiversity at different phylogenetic and spatial scales: within a single species in one region; a clade (genus) spread across regions; the entire biota.

Chapter 3 investigates in detail a single species in a single well studied region, the Wet Tropics rainforests. This multifaceted approach, combining spatially explicit paleoclimatological models and comparative phylogeography, provides a powerful approach to locating historical refugia, highlighting the role of landscape, ecology and history in shaping population structure and hence the foundations of allopatric speciation.

The next chapter expands this to encompass the whole clade – the *Sphaerospira* lineage – spanning the 'mesotherm archipelago' of mesic forest along eastern Australia. A trans-species phylogeny is combined with bioclimatic modelling, spatial mapping of phylogenetic diversity and lineage diversification analyses to reveal the profound link between the intra-specific phylogeography and the underlying inter-specific phylogeny.

The final chapter (5) represents a culmination of the ideas and data introduced in the previous sections to: i) explore methods of incorporating phylogenetic information into biogeography and macroecology; ii) provide a detailed and comprehensive biogeography of eastern Australia; iii) speculate upon major patterns and processes in biodiversity: speciation, accumulation, retention, extinction. However, it does come to some broad conclusions: that diversification proceeds from peripheral isolation, is driven by environmental gradients; that diversity is governed by environment through extinction, with ecosystem turnover due to the late Tertiary aridification.

Chapter 1: Introductory remarks on the research topic

The mosaic of forests along the east coast of Australia provide a rich system for interpreting evolutionary process from biogeographical pattern, with biodiversity arranged in suites of taxa specialised to spatially structured habitats. The history of these habitats is dominated by the long term drying of the climate from the Miocene leading to decline and fragmentation of a formerly vast and rich rainforest habitat (Adam 1992; Truswell 1993; McGowran et al. 2004). In parallel, sclerophyllous, xeric and tropical seasonal habitats have expanded from the mid-Miocene (Bowler 1982; Martin 1994; 2006) with widespread desertification across the continent accelerating from the late Miocene (Fujioka et al. 2005; Byrne et al. 2008). As they currently stand, major rainforest domains are restricted to the eastern areas of the Great Dividing Range (Webb & Tracey 1981), centred around a disjunct series of upland areas: an “archipelago” of cooler – mesothermal – climate (Nix 1982; 1991; Nott 2005). During the Pleistocene these forests have waxed and waned to an uncertain but considerable extent (Kershaw 1994; Kershaw et al. 2005).

This pressure of wholesale environmental change has not only lead to much fragmentation and extinction but also set up a powerful ecological forcing across the environmental gradient between the mesic and expanding drier classes of habitat: vine scrubs and sclerophyll woodlands. What is the balance of these forces in shaping current diversity? This depends on the scale at which these processes operate. For mammals and other large vertebrates extinction due to habitat contraction is a major theme, with most extant species being quite old (Winter 1988; Archer et al. 1991; Moritz et al. 1997; Moritz et al. 2005; Johnson 2006; Hocknull et al. 2007; Meredith et al. 2009). But for the land snail this vast complex mosaic may harbour multiple ongoing processes of vicariant and environmentally driven diversification, with relative patterns mediated by environmental change and ecology (e.g. Moritz et al. 2000; Crisp et al., 2004). As this complexity of diversification likely covers wide and overlapping spatial, temporal and taxonomic scales, a phylogenetic approach will taken to defining and linking the various levels of diversity. Aspects of the tempo and mode of evolution may be reflected in the spatial distribution of this phylogenetic diversity (PD: e.g. Faith 1994; Webb et al. 2002; Davies et al. 2007; Graham et al. 2009). This approach has a number of theoretical advantages that converge with practical issues of describing the well collected but poorly systematised land snail diversity.

Defining the gauge of diversity

Although biological variation exists at many levels, generally the distribution of biodiversity is classified by taxonomic rank starting with the species: '..the fundamental units of biodiversity..' (amongst others Rosenzweig 1995). Biological units below this have, until recently, rarely been considered, while beyond the spatial scales of individual species, higher taxonomic ranks are invoked (Gaston 2000; Lomolino et al. 2006). In remarking that 'the polytypic species and superspecies must be used as faunistic units, not the old monotypic species, in zoogeographical studies' A. J. Cain (1954) recognised the importance of evolutionary processes within spatially structured groups, which has more recently emerged in the concept of phylogeography: "the bridge between population genetics and systematics" (see Avise 2008). These issues converge on the question of taxonomic discrimination in land snails. Based on species distributions Solem (1979a) estimated that the average geographic range of a land snail was less than 50 miles. Hence a large proportion of the taxonomic diversity is distributed allopatrically, and subjective discrimination becomes a significant concern (e.g. Mayr 1976; Isaac et al. 2004; Hey 2009). Combined with the vagaries of distinguishing higher taxonomic ranks,

notoriously unstable for many land snails (see for example Emberton et al. 1990), it appears desirable (even necessary) to include a detailed phylogenetic perspective in assessing patterns of diversity.

Why these snails?

To recap, investigating differential patterns of diversification requires a group: i) diverse enough to provide multiple comparisons, ii) spread across spatially structured habitats, iii) with a fine scale geographical pattern, iv) a good taxonomy, and v) a phylogeny with time scale, so that we can tell where and when things happened. No group has all these but camaenid snails have a number of useful features making them worth investigation:

1) The camaenids are the most speciose family of snail in Australia with possibly several hundred species encompassing suites of sympatric/overlapping taxa across the biogeographical domains in eastern Australia.

2) Small ecological scale results in numerous populations at various levels of geographical isolation.

3) Explicit expectations of a strong geographical structuring because of low vagility and/or strong habitat specialization.

4) They are available for DNA analysis because of the extensive collection held in the Queensland Museum stored in alcohol.

5) As shells last several years or more, post-mortem collecting means distribution knowledge is relatively good.

Camaenid land snails are widespread across the continent including a rich diversity spread in a broad band down eastern Australia from Cape York to Cape Otway (Figures 4, 11; see also Stanisic 1994). While diversity hugs the coast, there is a consistent presence further west in the drier vine scrubs and open forests. Based on patterns of species diversity and some fossil evidence Solem (1979a,b; 1992; 1997) proposed that the Australian Camaenidae are Miocene and post Miocene immigrants from a large pool of Laurasian taxa. Eastern Australian camaenids have been collected since Férussac (1823), and combined museum records amount to a detailed and fine scale distribution database that is used throughout this study. Further details on taxonomy and systematics are covered in Chapter 2, with the remainder of this part of the introduction devoted to noting some aspects of the use of phylogenetic information in macroecology.

Phylogenetic diversity in biogeography and macroecology

Analysis of spatial and environmental patterns of species diversity is widely used in biogeography to identify to delimit biogeographical regions and endemicity (e.g. Lomolino et al. 2006; Crisp et al. 2001), and also in macroecology (Brown 1995; Gaston 2000) to explore species area relationships and environmental correlates of diversity, most famously the latitudinal gradient in diversity. Traditional use of phylogenetic information in biogeography and macroecology has been limited to topology or Linnean ranks only (Nelson & Platnik 1981; Brooks & Mclennan 1991; Rosenzweig 1995; Heard & Cox 2007), however with the advent of molecular phylogenies, branch length information can meaningfully be incorporated as an index of time, adding considerable information (Ricklefs & Schluter 1993; Webb et al. 2002; Wiens & Donoghue 2004). Such phylogenies supersede Linnean

ranks and provide the most complete natural hierarchical classification. The small geographical range of many snail species means that similarity and turnover functions at medium scale (of the order of the bioregions across eastern Australia) return extreme values – everything is different – and hence tend to be uninformative (Koleff et al. 2003; Legendre et al. 2005). A traditional approach is to use higher Linnean categories for larger scales; the phylogenetic approach takes this to a logical conclusion.

While traditional taxonomy strove for an exact identification of diversity by Linnean classification, putting aside philosophical questions of the vagueness of such concepts (Mayr 1976), proximate estimates of diversity may be adequate for investigation of the major historical and environmental factors correlated with diversity and diversification (e.g. Barraclough & Nee 2001; Ricklefs 2007a). Therefore phylogenetic diversity that incorporates lineages of uncertain taxonomic rank can embody sufficient detail for evolutionary inference, while not necessarily being the definitive measure of diversity. In a manner of speaking, 'it makes up with numbers what it lacks in perfection'.

In the first instance, phylogeny can provide information on sister groups allowing comparison between groups of the same age (e.g. Cardillo 1999) thereby eliminating the confounding factor of time, or account for different ages by estimating net diversification from age and numbers of taxa (e.g. Magallon & Sanderson 2001; Wiens et al. 2006). Many analyses use these approaches, which (tacitly or otherwise) assume non-equilibrium diversity, and so in effect, fail to properly consider all of the alternative hypotheses (Ricklefs & Bermingham 2001; Weir & Schluter 2007; Rabosky 2009a,b). In particular it has the limitation of considering only net diversification - the overall difference between speciation and extinction resulting in a certain number of extant taxa after a certain amount of time. For example, referring to Figure 1, after the same amount of time group x has more species than group y and so the net diversification rate of y is greater than that of x but this should not be confused with y have a higher speciation rate. The danger here is that using net diversification rate as the measure *ipso facto* eliminates various other processes as explanations of the different levels of diversity. For example, much ecological theory considers the concept of carrying capacity – niches fill and diversity reaches a dynamic equilibrium (e.g. Ricklefs 2007; Rabosky 2009b). Such situations may be characterized by density dependent patterns of cladogenesis (e.g. Rabosky & Lovette 2008), so that for example (referring again to Figure 1A) if x occupies a smaller island than y and both have the same speciation rate x will reach capacity earlier with speciation rate slowing; the end result will be of clades showing the same differential net diversity. However, this has nothing to do with speciation rate. In short, the limitations of using net diversification are that 1) it assumes equilibrium diversification rate but 2) assumes diversity itself is not at equilibrium, that there are no environmental constraints on diversity. Neither of these are warranted (especially the latter), and more to the point, using more complete phylogenetic information can ameliorate these shortcomings. Net diversification models have a further potential bias due to the selection of clades to be considered, compared with analyses using a complete phylogeny of all extant lineages. Other distinctions worth noting are: the difference between equilibrium diversity and a steady state with birth-death turnover, where depending on the amount of time passed and the turnover the same equilibrium diversity can give rise to different phylogenetic shapes (Figure 1B); and lineage specific differential diversification reflecting ecosystem change where the total diversity is relatively constant but the proportions contributed by different groups changes (Figure 1C). Both of these may be relevant to interpreting long term trends in Australian biodiversity.

There are two broad approaches to utilizing phylogenetic diversity: 1) Lineage-through-time (LTT) patterns of lineage accumulation in monophyletic groups (Hey 1992; Harvey et al. 1994; Nee 2001; Kubo & Iwasa 1995; Rabosky 2006), where relative rates of speciation and extinction, and deviation

from equilibrium may be inferred; 2) Patterns of species diversity (SD) and phylogenetic diversity (PD), where relationships among SD, PD and environmental indices may inform on environmental determinants of biodiversity (e.g. Faith 1994; Sechrest et al. 2002; Bickford et al. 2004; Davies et al. 2007; Graham et al. 2009). Here the diversity can be polyphyletic assemblages such as local or regional diversity or ecological categories, and a particular version of this gaining widespread interest is the use of relative phylogenetic clustering (e.g. Tofts & Silverton 1999; Webb et al. 2002; Davies et al. 2007). This approach is not without complications, particularly of defining an appropriate null pool of diversity and hence interpretation of the result (Hardy & Sentierre 2007; Vamosi et al. 2008; Graham et al. 2009; see also Chapter 5).

Analysis of SD/PD does not use all of the possible implied information in a phylogeny as can a birth-death *(b/d)* process model but can summarize average PD to SD patterns capturing some of the possible differences in lineage-through-time patterns. In principle *b/d* can estimate the extinction component (not withstanding inherent ambiguities: Nee et al. 1994; Raboksy 2009c) and estimate non-equilibrium models; spatial patterns can't do either but using PD rather than just age can help differentiate patterns such as in Figure 2, which net diversification methods cannot.

A requirement (and strength) of LTT approaches is the need for near complete phylogenies of all extant lineages, despite some efforts to account for incomplete sampling (Pybus & Harvey 2000). It has further become apparent that multiple birth/death processes can theoretically yield the same or effectively the same LTT pattern (Harvey et al. 1994; Ricklefs 2007; Rabosky & Lovette 2008). Particularly, extinction may be very difficult to identify (Nee 2001; Purvis et al. 2009; Crisp & Cook 2009; Rabosky 2010), hence LTT alone can not adjudicate among processes and needs to be augmented with other sources of inference; to be treated as just one 'tool in the box'. For example, outcomes may be constrained by testing *a priori* scenarios drawn from biogeographical or paleoclimate inferences, or strengthened by comparative analysis of independent lineages (Egan & Crandall 2008; McPeek 2008).

Currently *b/d* modelling is restricted (like sister group comparisons) to clades whereas spatial or community PD approaches can pool unrelated taxa. While PD approaches skirt this formal precept required of LTT, the problem with restricting the analysis to whole clades is that usually the factors we are interested in are not clades but other categories (community, ecology etc). Rending the diversity into a set of suitable monophyletic units, other than resorting to clade average values, most often means leaving out a lot of the actual diversity itself, and worse, may do so in a way that selectively bias against lineages that diversify across the categories of interest. Some attempt has been made to apply diversification models to more complex phylogenetic (polyphyletic) assemblages such as the BiSSE method (Maddison et al. 2007; Fitzjohn et al. 2009) that estimates the effect of a binary character (such as mesic/not mesic) on speciation and extinction. While this currently only considers equilibrium *b/d* models, the concept is being extended to include changes through time (Rabosky & Glor 2010).

The interpretation of patterns of local and regional diversity has increasingly focussed upon the distinction of ecological and historical scale, where local pattern is described as an ecological scale sorting of the historically determined regional pool (Ricklefs & Schluter 1993; Rosenzweig 1995; Losos & Schluter 2001; Webb et al. 2002; Ricklefs 2008). Within this framework of local and regional scales, diversity is categorized as alpha, beta or gamma (α, β, γ):α being the site (local) diversity, γ being the total regional diversity, and β the change in diversity from site to site (Whittaker 1960; Rosenzweig 1995; Tuomisto 2010a). Here the phylogenetic approach more naturally defines the scale

for each lineage, and while of limited import to alpha diversity, it has useful contributions to untangling beta and gamma diversities (Wiens & Donoghue 2004; Hardy & Sentierre 2007; Graham & Fine 2008).

How much of the diversity (species/phylogenetic) and process (diversification/extinction) reflects an equilibrium of the current environment, long term trends of ecosystem change, and historical contingency or recent past disruptions? An overarching macroecological gambit is that key trends dominate and can be identified in summary patterns of diversity, where noise and idiosyncrasy are neutralized. The danger in this is that perhaps nature really is a 'tangled bank' (with apologies to Darwin 1859) and a summary trend merely obscures that which can only be appreciated by elucidation of individual cases. The approach taken here of analysing the entire biota, all lineages across the entire range, could fairly be described as unbiased sampling but must be an amalgam of multiple processes, at multiple scales. Given zero-sum game concepts of much diversity (equilibrium states, neutral models: MacArthur & Wilson 1967; Rosenzweig 1995; Hubbell 2001) can it be possible to see 1) see the balance of processes in 2) the dominant patterns?

The phylogeny developed in Chapter 2 provides an almost 'tree-of-life' complete description of lineage and species diversity across the entire east coast distribution from Torres to Bass Straits. Combined with fine scale distribution data (associated with the museum collections) this will enable detailed biogeographical and macroecological analyses of an entire biota. Part II introduces the alpha taxonomy and distribution database and the major geographic and environmental patterns of taxonomic diversity, setting the *mise en scéne*.

Introduction part II: Camaenid species diversity patterns

Distribution data

All specimens collated in the Queensland and Australian Museum collections have a matching computer based record of location in latitude and longitude. Most records also have a location record that has specific information on that site for habitat characteristics. Together with some other records, a distribution database has been assembled covering Eastern Australia. This database comprises: Queensland Museum – 5294; Australian Museum – 3302; Museum of Victoria – 749; Solem monographs – 916.

This includes a total of 9230 records east of 141 ºE longitude which I will define as eastern Australia for the purposes of this study. Each of these records has individual catalogue numbers, taxonomic identifiers, locality information and latitude and longitude data to a precision of three decimal places. Of the 339 Museum of Victoria records half lack species names. Combined with the confusion between *Austrochloritis* and *Chloritobadistes* (see Chapter 2) and inconsistent nomenclature, effectively none of the Victorian Camaenid records can be relied upon at this stage. The details in the gazetteer of the Solem 1992 monograph were accepted as is but all the records of the QM and AM were check for location and taxonomic accuracy. This proceeded in parallel to the gathering of the genetic data allowing for some rationalization of the taxonomy. The nomenclature is elaborated detail in Chapter 2, and follows the Zoological Cataglogue (Smith 1992; AFD website) plus many informal museum codes, some of which are in the process of being described (Stanisic et al. in prep.). The eastern Australia distribution dataset used in this study comprises 321 taxa (133 named 288 codes) with nomenclature

fixed as of November 2005. While there has (and no doubt will be) some changes these are mostly changes of name and not definition.

Methods

The general strategy for these analyses is: 1) Database management in Microsoft Excel; 2) Mapping and bioclimatic statistics using ESRI ArcView 3.2a; 3) statistical analyses in either Excel or JMP 3.1.5 (SAS Institute Inc.). The distribution records were first assembled and checked to form the primary database, which can then be used to feed into other spreadsheets for specific diversity analyses. From these, data tables can then be imported into ArcView 3.2 containing the digital elevation data (9 second resolution) and ANUCLIM digital bioclimatic data layers (Houlder et al. 2000). Various characteristics of points or spatial units can then be calculated and mapped, and exported back to spreadsheets for further analysis. For example, for the calculation of species diversity (SD) for a grid scale, cell centre points are pre-defined and database records assigned to the nearest grid cell centre points to provide SD. These point values of SD are then imported into ArcView where mean values are calculated for a rectangle of ArcView cells that match the grid scale, to expand the point SD onto a grid map for presentation. Grids do not have to be rectangular, and a hexagonal grid has a number of attractive properties (for example for use in mapping turnover; Birch et al. 2007) however a rectangular grid has been used for convenience, and is the most commonly used in mapping (e.g. for Australian examples see Crisp et al. 2001; Bickford et al. 2004; Slatyer et al. 2007). Using this system allows development of environmental indices based on a few key factors that represent the gradient from mesic closed forest, to open dry forest and the related temperature gradient associated with the mesotherm "archipelago" and latitude range (tropical-temperate). This involves using ANUCLIM climate layers, especially the Nix Plant Growth Indices (Mesothermal, Megathermal and Microthermal), including and paleoclimate scenarios of Nix (1982; 1991) and Williams (1991).

Diversity Discovery Curves

Species accumulation or discovery curves can be used to gain an overview of how much of the diversity has been discovered and how much might remain to be discovered (Rosenzweig 1995; Brown 1995). For the discovery curve presented here (Figure 3) the dates are the original date of description of the species, or for informal codes the first date of collection of a specimen that has subsequently been assigned to that code. Of the 321 taxa 133 are formally named species and 188 informal codes. The first camaenid of eastern Australia was described in 1823 (*Hadra bipartita* Ferussac 1823) and the curve shows cumulative number of taxa up to the year 2000. For comparison the cumulative phylogenetic diversity (PD) is also shown; this uses an earlier version of the phylogeny shown in Figure 7 of Chapter 2, and so presages those results but is shown here to allow discussion of a few points. Taxonomic diversity climbs steadily until the 1980s, when extensive geographically organised collecting and collating increased, slowing after about 1990. In particular the rate at which new units are added per museum record slows dramatically over the last 30 years. This suggests that while the taxonomic discovery curve does not have a long asymptote typically taken to imply complete or near complete sampling (Brown 1995; Legendre and Legendre 1998), the combination of extensive collecting and recent levelling of the curve suggests that most of the diversity has now been accounted for. This is accentuated in the PD curve: early sampling of the major morphological types ('families') captures a lot of the deep phylogenetic diversity, hence the curve initially rises relatively faster than SD, while increasingly finer sampling and discrimination in the latter sampling phases splits more terminal level PD hence the PD curve flattens earlier. The sharp rise around 1908 is due to the

discovery of two snails endemic to the rainforest of Far North Queensland (Wet Tropics) '*Austrochloritis' agamemnon* and '*Jacksonena' delicata* (see references in next chapter), which turn out to be ancient biogeographic complexes (see Chapter 2).

Defining the region: mapping diversity and environmental indices

A map of the eastern Australian region is shown in Figure 4, including species diversity per $^1/_3$ degree grid scale for all records. There are sampling gaps but the map also clearly highlights that the vast majority of diversity follows a relatively narrow transect along eastern Australia, largely demarked by the Great Dividing Range which has been a permanent feature throughout the Tertiary epoch (Nott 2005). It is this region that this study is particularly focussed upon. As evinced by the higher taxonomy (See the Solem 1979-97 monographs), and is clear in the large scale phylogenetic framework of Chapter 2, there is a major phylogenetic distinction between east and not east – the map shows the database distribution of eastern Australian lineages, one of which has a sub-unit in South Australia, and two with distantly related lineages endemic to the Red Centre. Some lineages of camaenids in eastern Australia are distributed further north across Torres Strait and New Guinea; Victoria has low diversity and the Family is absent from Tasmania. This eastern Australian (for brevity will also be referred to as 'east coast') ecogeographical region has been demarked in previous traditional biogeographical classifications encompassing the Torresian and Bassian domains (Spencer 1896; Iredale 1937; Burbridge 1960; Kikkawa & Pearse 1969; Cracraft 1986; Joseph et al. 1993; Crisp et al. 1995).

This domain can reasonably be defined bioclimatically by temperature, rainfall and seasonality. The Plant Growth Indices (PGI) developed by Henry Nix (1982; 1991) provide a versatile synthesis of these factors. To quote Nix (1982) the PGI "…model comprises a set of functions that transform the dynamic, non-linear responses of plants to the major environmental regimes into dimensionless indices on a linear scale of zero to unity". The core functions in the model relate plant response to radiation, temperature and water (see Nix 1982, and shown schematically in Appendix 1A). There are three categories decribed as Megathermal, Mesothermal and Microthermal with optimal temperatures of: Mega(C3) 26-28 ºC; Meso19-22 ºC; Micro10-14 ºC. Combining these functions with the ANUCLIM climate layers of Annual Mean Radiation, Annual Mean Temperature (AMT) and Moisture Index, digital maps of these indices can be calculated. These digital bioclimatic layers were developed by Adnan Moussalli at the Spatial Ecology Laboratory, University of Queensland. The relationships between these indices and AMT and latitude for the snail distribution data points are shown in Figure 5: the Megathermal and Microthermal regimes are truncated by northern and southern range limits, while the Mesothermal domain is fully defined and broadly distributed across the range. Figure 4A shows the spatial distribution of the regions with PGI > 0.40, and Fig. 4B the sum of the three PGI: these mesic domains form a series of subregions along eastern Australia with the Microthermal domain extending into southern Victoria and Tasmania. In particular the "Mesothermal Archipelago" is apparent with three main "islands" FNQ, MEQ and SEQ-SEA, with some northern outliers in Cape York. It also has a major presence in upland New Guinea (Nix 1982). The total land area of the east coast region is approximately 1.1 million km^2, while area with PGI >0.20 = 281,000 km^2, and Mesothermal >0.40 = 194,000 km^2.

The vast majority of camaenid diversity occurs within these domains. From the bioclimatic layers and distribution dataset, of the 321 east coast taxa 74% fall within the >0.20 PGI domains indicated on Figure 4A-B, 65% falling within the >0.40 Mesothermal domain. This can be compared to the distribution of environments within the eastern Australian region. For this purpose it is useful to define

a zone that encompasses the known distribution points bounded by PG indices such that at least one of the three (mega, meso, micro) has a value of >0.20. Excluding most of Victorian due to poor sampling/taxonomy, we get a zone marked on Figure 4C. This does not include some of the inland records and taxa however the zone encompasses 97% of records and 97% of the taxa. The bioclimatic distribution of the snails can then be compared with that of this broad zone, shown in Figure 6. The >0.40 Mesothermal domain occupies only 26% of this area but contains 65% of taxa; the >0.40 PGI domain covers 40% of the area but contains 74% of taxa. Similar numbers obtain if we define the zone as the total area of the half degree grids cells that contain any record. The snail taxa can be split into groups with different bioclimatic envelopes; for example clade 4 (as defined in the phylogeny Chapter 2) is particularly concentrated in the Mesothermal domain (>90% of taxa and records fall within the >0.4 Mesothermal domain).

Distribution ranges and determining spatial sampling and scale

A key aspect in mapping the distribution of diversity is spatial scale: one approach is to analyse at several scales (Crisp et al. 2001; Bickford et al. 2004) but in the case here, I will focus on one optimal scale based on the combination of genetic and distribution data coverage, and endemicity (Gaston 2000; Koleff et al. 2003). The chosen optimal scale of 50km grid cells provides for a near continuous coverage spanning Eastern Australia from Cape York to the Victorian border for which there is near complete species and phylogenetic diversity data.

The distribution dataset has records in decimal degrees, however this can be converted to equal area scale and great circle distances using the Haversine Transform (Sinnott 1984). Across the latitudinal range of the east coast "transect" equal degree scale grids will have different areas such that those at the extreme southern end ca. 40 ºS will have only about 80% of the area of grids at the northern end ca. 10 ºS (Appendix 1B). For the purposes of some types of macroecological analyses it may be necessary to use equal area (and absolute distances), and here it is used in providing estimates involving distribution range and areas. However, as there is a negligible effect on the gross diversity patterns depicted, for many pictorial purposes the 50km and ½ degree grids are interchangeable.

Estimating distribution range of any species or taxonomic unit runs into some complications but given a detailed distribution dataset reasonable approximations can be made (Gaston 1994; Gaston & Fuller 2009). Here two measures are estimated 1) maximum linear range, and 2) an estimate of area occupied as determined from grid cells at a nested series of scales. Both these use of all records, with units expressed in kilometres. Taxa with records from one site only are set at a minimum of $^{1}/_{10}$th degree grid area.

Choosing a suitable scale to describe diversity (such as SD) balances completeness of taxon sampling and of coverage. In estimating things such as range or fit to environment of any particular species, all records can be used but to describe diversity by unit area it is necessary to consider how well the region has been sampled, to determine if all the taxa actually present have been recorded (e.g. Brown 1995). A finer scale can produce a more detailed map correlating diversity to environment but at the potential expense of more and more areas with inadequate sampling or no sampling at all, degrading the coverage. Numerical procedures can also be used to assess sampling and coverage completeness (Gaston 1994; Graham & Hijmans 2006; see Appendix 1C). In the museum collections there is naturally a correlation between number of records and number of species, therefore rather than factoring by records evidence of saturation can be found in the relationship between records and SD: at

the ½ degree scale overall there is a rapid decline in SD-record relationship beyond 10 records. Coverage can be measured as proportion of total area actually covered by grid cells, for a range of grid scales (from 0.1-2 degrees): there is a rapid decline for scales less than ½ degree. Relationship between grid scale (from 0.1-2 degrees) and proportion of grids with at least 1, 2, 3 and 4 records can also inform on coverage: there is a rapid decline for grid scales less that ½ degree scale.

As a preliminary view, Figure 8 shows the number of species per degree and ½ degree latitude bands for eastern Australia from Torres Strait into Victoria (9-39 ºS latitude), and the turnover (as a proportion of species shared between degree latitude blocks) in species across the same range. Broadly, this is a latitudinal transect but with some distortion due to differences in area among degree bands (c.f. mid-domain like effects: Colwell & Lees 2000). There is a high degree of turnover with less than 30% of diversity shared among degree latitude blocks. This fine scale endemism is further depicted in the range and area distributions (Figure 9) with some 50% of species having a maximum linear range of 100 km or less (which amounts to spanning at most three ½ degree grid cells). There are some 50 species with ranges of 10km or less (many island, some mountain top endemics). There are 18 species with maximum ranges exceeding 500km, with two species (*Xanthomelon pachystylum* and *Trachiopsis mucosa)* greater than 1,000km, both across northern-central Qld. Distribution range can be extended from species to genera, phylogenetic clades and groups defined by phylogenetic divergence level (see Chapters 4 & 5).

Species diversity can be estimated at various spatial scales ranging from 5-100km equal area grids, excluding inadequately sampled grid cells (Figure 10). Determining adequacy of sampling considered: SD/records; the number sites visited within each cell; the number of times collections were recorded; considered (expert) opinion for each grid cell. Cell were excluded because 1) the specimen sampling was not adequate and there was reason to suspect that other known taxa would ultimately be recorded within that grid cell; 2) because genetic sequence data collected was inadequate to represent phylogenetic diversity (Chapter 2). In determining the completeness of the phylogenetic diversity, it is necessary to use the taxonomy to extrapolate the results of a limited number of specimens to the entire range of the taxon as indicated by the distribution dataset. In some cases this may overlook phylogeographic diversity, however the generally fine scale distribution of taxa and the high proportion that has been genetically sampled, suggests that most of the major patterns in diversity are represented.

Near point (community) estimates of SD (records falling within a 1 km radius) are typically low (median 2) but show maximum values of 12 on the Carbine Tablelands (Wet Tropics) and 10 on the Lamington Plateaux (Border Ranges). Maximum SD remains similar (8-12) across 5-20km grid scales, rising to 18 and 27 at the 50 and 100km grid scales respectively. For the ½ degree grid, species diversity (SD) ranges from 1-18 (max. 15-20 depending on the exact centring of the ½ degree grids; see Appendix 1D) with a median of around 5. Figure 11 maps SD for the chosen ½ degree grid scale, indicating all grids with any records (11B) and those selected as being sufficiently well sampled to most reasonably be considered complete (11C,D). Thus of the total of 388 east coast ½ degree grid cells with any records 122 were excluded leaving the remaining 266, which contain 92% of records and 94% of species, and covers an actual land area of approximately 568,000 square kilometres (i.e. excluding cell areas below sea level). Considering SD sampling adequacy only, 284 grids can be included covering 618,000 square km, 95% records and 98% of species.

Most notably regions west of the Great Dividing Range across southern Qld-NSW are excluded due to insufficient collecting, and/or lack of genetic data (only shell collections); similarly Victoria. Exclusion

of this diversity should not be taken to imply that it is irrelevant, just that it is not possible to incorporate into the study other than noting that SD is low. After excluding cells the remaining included set still provides a reasonably continuous coverage encompassing the vast majority of diversity at a sufficiently fine scale relative to the geography of Eastern Australia to discriminate subregions.

Endemicity and scale

In addition to defining SD, a map of endemicity can be defined based on the number of grid cells occupied, for example the inverse function of the number of cells (weighted endemicity: Dony & Denholm 1985; Crisp et al. 2001). The inverse function is directly influenced by grid scale as the most rapid decline in relative endemicity is among the fewest number of cells (Figure 12). Therefore there can be an apparently large difference in endemicity depending on how the scale subdivides regions. For example there could be two mountains each with same number of endemic species but if the first fell into one grid cell but the second into two the second would only have half the weighted endemicity. This is of course always a problem for fixed scale and can be addressed by careful choice of scale or use of multiple scales (Laffan & Crisp 2003). Another approach is to allow a similar level of weighting for small numbers of cells with a decline in weighting after that. A suitable function for this is a sigmoidal relationship:

sigED = $1/(1+ \exp(n))$, then normalized (so that for n=1 sigED = 1)

= $[1/(1+\exp(an-b))] * [1/(1+\exp(a-b))]$

where n = number of grid cells, and a and b are parameters that affect width and steepness. Figure 12 shows the relationship of endemicity to grid cells for the inverse function, and the sigmoidal function using several values of a and b. The endemic diversity of any region or spatial unit is then the sum of the endemic diversity of each species. Applying this to the spatial grid, endemic diversity can be mapped as is shown in Figure 11E,F. The proportion of endemic diversity can be expressed as endemicSD/SD; this is the corrected weighted endemism of Crisp et al. (2001).

Species and environmental turnover

Species turnover can be investigated further using a standard statistic such as βt (Wilson and Shmida 1984; Koleff et al. 2003), which calculates the relative turnover between two sets of taxa: in this case pairwise comparisons between 50km scale grids with SD>2 (239 grids; Figure 13). Species turnover can be complimented with an environmental change metric (Brown 1995; Legendre et al. 2005) such as the Euclidian distance between two grids (x and y) across n bioclimatic parameters m:

$\Delta E = \sqrt{[\sum (m_x - m_y)^2]}$

This index of bioclimatic (or environmental) difference could also be expressed on a scale of 0 to 1 as a proportion of the maximum value ($p\Delta E$). Here the three PG indices (mega-, meso- and microthermal) are used to provide a measure of ΔE. This is not a comprehensive measure of environment (what is?) but as the PGI correlate well with snail diversity and are standardized measures, they form a useful metric to match SD patterns across the same grid cells.

The following analysis is largely descriptive and there is no attempt to correct for spatial auto-correlation but use of all comparisons across a reasonably uniform transect of grids should ameliorate some biases (Legendre 1993; Legendre et al. 2005). Much has been made of the latitudinal gradient in species diversity, and many studies have searched for turnover trends following the idea that this reflects finer scale speciation and endemism in the tropics (Rosenzweig 1995; Gaston 2000; Koleff et al. 2003; Qian et al. 2009). It is possible to infer a correlation between SD and latitude (p=0.0001) but weak or negligible correlation between endemic SD and latitude (inverse function SD, p=0.114; sigmoidal SD p=0.576). However, latitude is a weak predictor of SD compared to environment (R^2 of 0.06 versus R^2 of 0.31 for mesothermal), and in a multivariate framework factoring for environment, it becomes a minor effect. While overall there is a trend, the maps highlight that the major patterns of variation in SD are at sub-regional scales (Figure 11).

Figure 13 shows some characteristics of species turnover and environmental difference. Using all pairwise comparisons among the 239 grids βt shows a quasi-logarithmic relationship with geographic distance while ΔE shows a more biphasic relationship, with the steeper gradient out to a distance of about 4 grid cells, probably reflecting the typical east-west width of the grids. As noted previously, turnover is generally very high, often too high to allow useful analysis (βt>0.8; Koleff et al. 2003; Legendre & Legendre 1998), with βt ranging from 0 to 1 even for adjacent grids, and by a distance of 4 grids (ca. 200km) more than 50% of comparisons have βt >0.8. There is also a tail of long range values <1 due to the few very wide-ranging species. In both these situations a phylogenetic perspective should be informative. For a pictorial representation, species turnover can be mapped (as mid-point of adjacent grid cell βt) and compared to a similar map of ΔE (Figure 13 lower panel). This highlights the high turnover across FNQ/Wet Tropics, the Border Ranges and south of the Hunter Valley around the Sydney Basin area, and association of βt with ΔE. Also note the relatively low turnover across SE-central Qld despite reasonably high SD.

One can elaborate on the relationship between species and bioclimatic turnover (βt and ΔE) at different latitudes and at different spatial scales, i.e. the response of species to environment. As expected there is a declining slope of βt/ΔE with increasing distance due to saturation effects in βt because of the short ranges of most species. Therefore to eliminate this and intrinsic effects of geographic distance, comparisons are restricted to adjacent grids only (rather than by using residuals). Given that among these comparisons mean βt = 0.471 with range 0-1 and mean pΔE =0.10 (range 0.02-0.44), this scale encompasses much of the relevant patterns and processes (Brown 1995; Gaston 2004).

Considering only adjacent grids (minimum geographic distance n= 607), east-west (mean βt= 0.485) is slightly higher than north-south (mean βt= 0.459). Similarly longitudinal ΔE is larger than latitudinal ΔE (10.2 vs 6.7; pΔE = 0.12 vs 0.08), reflecting the consistent rainfall gradient from east to west along the Great Divide. Latitudinal patterns of turnover can be investigated using adjacent grids, either individually, or by pooling into latitudinal bands. Species turnover βt is broadly similar across latitude, slightly higher in the south than the north (non-significant trend slope -0.002 p=0.103) but the dominant theme is variation around the core biogeographical domains. Similarly for ΔE and latitude, although there may be a stronger trend with bioclimatic change in the south than north (slope = -0.135 p=0.004). There is a strong correlation between βt and ΔE such that βt = 0.40+0.011ΔE (p<0.001), with a slightly greater but non-significant difference between east-west and north-south (p~0.1) Thus factoring out bioclimatic differences, at the 50km grid scale there is a baseline average trend 40% turnover. Species turnover then tends to double across the strongest bioclimatic gradients (ΔE ~30, pΔE ~0.4). Binning the data into degree latitude bands, it appears that this relationship declines with

increasing latitude, however, if this is done by plotting the residuals of the βt on ΔE regression against mid-latitude there is no relationship (p=0.571). Accounting for ΔE in the relationship of βt and latitude (by least squares regression, or using residuals of the βt on ΔE) eliminates any trend (*p*=0.569), with ΔE far more influential (~100 times so, by F ratio). The implication here is that any large scale latitudinal pattern is a by-product of much more fundamental small scale patterns.

Species diversity of eastern Australian camaenid land snails

To this point I have mostly been concerned with various methodological details but return to a broad narrative on east coast camaenid diversity with only brief reference to eastern Australian biogeography, details on which there is an abundance of literature (see Chapters 2 & 4) but rarely done, at the same time, with such large scale and fine detail. A number of features and nomenclature can been seen in the Interim Biogeographic Regionalization of Australia (IBRA 6.1 website).

The biogeography of eastern Australia is dominated by the Great Dividing Range, which extends almost continuously from Victoria to Cape York. It stems from the Cretaceous rifting forming the Tasman Sea and has had a continuous existence throughout the Tertiary (Wellman 1979; Nott 2005; Twidale 2007). During this time the Australian plate has migrated northward as much as 20 degrees of latitude but steepening of the southern hemisphere latitude-temperature gradient has buffered climate change and allowed a long continuous history of the mesothermal forest realm (Bowler 1982; Nix 1982; Truswell 1993; McGowran et al. 2004; Greenwood & Christophel 2005). The formal description of eastern Australian biogeography begins with the natural history of Spencer (1896), through pioneering statistical analyses of Kikkawa & Pearse (1969) and continues with the molecular phylogeographies of today (e.g. Moussalli et al. 2009), providing a reference for most of the big features mentioned below: the tropical versus temperate domains and the MacPherson Macleay overlap (= Torresian and Bassian of Spencer 1896); the Wet Tropics (Burbidge 1960) and Burdekin Gap (Kikkawa & Pearce 1969); the Hunter Valley (Spencer 1896; White & Key 1957).

The region, and the snail diversity within it, forms a large transect from tropical 9 ºS to temperate 39 ºS spanning some 3,300 kilometers with species diversity for 50km (or near equivalent ½ degree) grids ranging from one to twenty species. Most diversity is east of the Great Dividing Range. The highest diversities are in the mesic forest domains such as the Iron and McIlwraith Ranges of Cape York, the Wet Tropics, mid-east Queensland (Clarke and Conway Ranges), south-east Queensland (Connodale to D'Aguilar Ranges), and Border Ranges and north-east NSW. These areas of overall biotic richness and endemicity are considered core refugial regions (Webb & Tracey 1981; Truswell 1993; Moritz et al. 2005). East to west of each, there are various gradients of diversity: to the west of the Wet Tropics out to the Einasleigh uplands; an extension of diversity to the west in the central Queensland highlands where the Great Dividing Range turns inland to the Carnarvon Ranges; a steep gradient of diversity immediately to the west of SEQ-Border Ranges (across the Darling Downs) and a lesser gradient across the north-west slopes of the Great Dividing Range in northern NSW. Diversity declines abruptly at the edges of recognized biogeographic boundaries or breaks: the Laura Gap south of the McIlwraith Ranges, the Burdekin Gap south of the Wet Tropics, the Hunter valley south of Barrington-Gloucester Tops.

Clearly diversity is strongly correlated with moisture (the mesic indices Figure 4A,B) and temperature, with diversity declining towards the southern end of the range but also attenuated at higher elevation in the middle of the range (e.g. the New England Tablelands). Figure 7 highlights these correlations by

plotting species diversity against mega- and meso- thermal indices (although this figure does not attempt to correct for phylogenetic and spatial auto-correlation). The relationship between SD and AMT (annual mean temperature) can be modelled (Appendix 1E) suggesting a minimum bound of around 10 ºC. Camaenids do make it into microthermal and sub-alpine environments (such as the high New England and the Snowy Mountains) but only a handful of species (SD 1-2 and only of the *Austrochloritis* type), and are absent from Tasmania. The correlation with temperature and moisture across bioregions strongly suggests a causal relationship (and not just historical contingency or co-incidence) and indeed bioclimatic models can account for a large (~50%) proportion of the variation in SD at these spatial scales (see Chapter 5 for a multivariate analysis of diversity and diversification).

These SD patterns can be complimented and contrasted with endemic SD patterns (Figure 11D-F): this highlights the diversity in the northern Wet Tropics, MEQ (especially the northern parts) and diversity either side of the Hunter Valley, in the southern parts of NENSW and the greater Sydney Basin region. By contrast there is a relative lack of endemism across SEQ-Border Ranges region despite high SD (and moderate levels of turnover: Figure 13). Across central to south-east Qld there are small islands of endemicity, e.g. Carnarvon Ranges and northern parts of SE Qld such as Kroombit Tops/Bulburin.

We now have a series of maps depicting bioclimatic trends and patterns of species diversity across eastern Australia. Each map has different qualities: SD pools all diversity widespread or not, endemic SD measures total endemic diversity irrespective of the proportion of endemic species while βt emphasizes change. To some extent differences in endemic SD can be affected by scale, as is inherent in the difference between the inverse and sigmoid function endemic SD maps (Figure 11E,F) for Cape York (Iron-McIlwraith Range) versus Harvey Rang/Mt Elliot rainforest endemics (just north of the Burdekin Gap). They are both discrete and endemic but the smaller scale for the latter (one grid cell) especially emphasizes inverse function endemicity. While endemic SD detects total range and βt the edges of ranges, in some ways they provide complimentary perspectives, but given the small range of most east coast camaenid species, perhaps turnover is the more encompassing view?

No one map embodies all aspects; it is comparison and contrast among them that highlight some basic questions about regional diversity, and where phylogenetic diversity should be informative. First are questions on biographical scale; the whole region is large with respect to individual species ranges, and it is apparent from SD, endemic SD and turnover (Figures 11 & 13) that the whole can be subdivided into sub-regions (domains) that generally mirror bioclimatic structuring depicted by the mesic index (Figure 5; and mesic index turnover Figure 13). While our eye is drawn to these domains, it is worth noting that the spaces in-between are also relevant if not so easy to circumscribe, as they too likely reflect processes governing biodiversity. However for the time being we will allow ourselves to focus on some obvious domains: Cape York, Wet Tropics, MEQ, SEQ-Border ranges, Sydney Basin: each has a different combination of SD endemic SD and turnover.

Cape York appears a discrete domain of SD (separated to the south by the Laura Gap) less defined by endemic diversity but much more so with turnover. The Wet Tropics domain has high measures of everything: SD, endemic SD, and species and bioclimatic turnover. Not only is it strongly demarked on all sides it has high levels of turnover within, reflecting fine scale regional biogeography (Williams et al. 1996; and see Chapter 3). Mid-east Queensland is similar but is more strongly demarked by endemicity and turnover than SD (more noticeable at the smaller spatial scale in Figure 4C), and has less within region turnover. South of this, from the St Lawrence Gap to the Hunter Valley is a broad zone of SD peaking mid-way around SEQ-Border Ranges. However, endemic SD shows quite a

different picture (with minor sensitivity to scale) with only small patches of diversity in the north and middle (high SD) areas rising to a zone of high endemic SD toward the southern end. This is paralleled by turnover, shallow in the northern half (both species and bioclimatic), rising across the Border Ranges with distinct east-west component along the Great Divide. This reflects the Torresian-Bassian overlap (Spencer 1896; first noted in snails by Iredale 1937) and the strong east-west environmental gradient across the Great Dividing Range. Particularly in SEQ (adjacent to Fraser Island-Morton Islands on the maps) there is high SD, only minor endemic SD and moderate turnover, mostly east-west in species. South of the Hunter Valley SD there is a domain of moderate SD but relatively higher endemic SD (somewhat sensitive to scale) with matching large amount of turnover.

Can we describe diversity patterns in further detail and hence identify different processes? Are these domains of high SD centres of retention of diversity, with many old endemic species, are they centres of recent diversification, are they zones of admixture or overlap? For areas of overlap (accumulation), are these expansions from core areas (akin to secondary contact) or a continuum? And what of the gradients at the edges of core diversity domains, are they sinks or sources? In short, what environments are creating, accumulating, retaining, or losing diversity? The contrasts of SD, endemic SD and species turnover can differentiate some of this (for example overlap in the case of SEQ-Border Ranges) but adding information on phylogenetic diversity greatly expands the inference space, and this theme will be taken up at various spatial and phylogenetic scales in subsequent Chapters 3, 4 and 5. But first that data must be gathered.

Figure 1. Net diversification versus birth and death (speciation and extinction) rates: equilibrium versus non-equilibrium effects

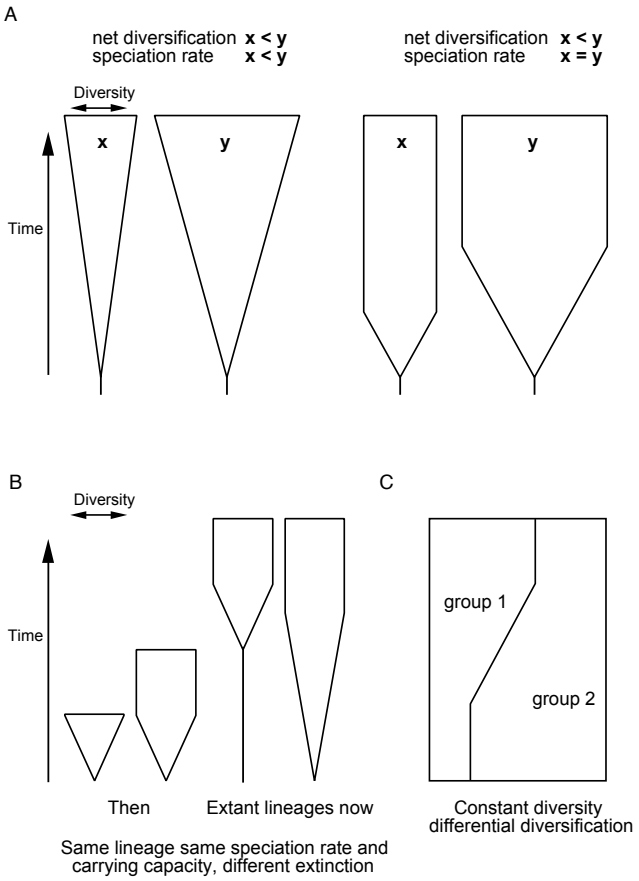


Figure 2. Tree shape and Lineage -Through-Time and Phylogenetic Diversity patterns. Age (=2) and taxonomic diversity (=4) are the same in each but PD is different: i) 6, ii) 7.6, iii) 4.4.

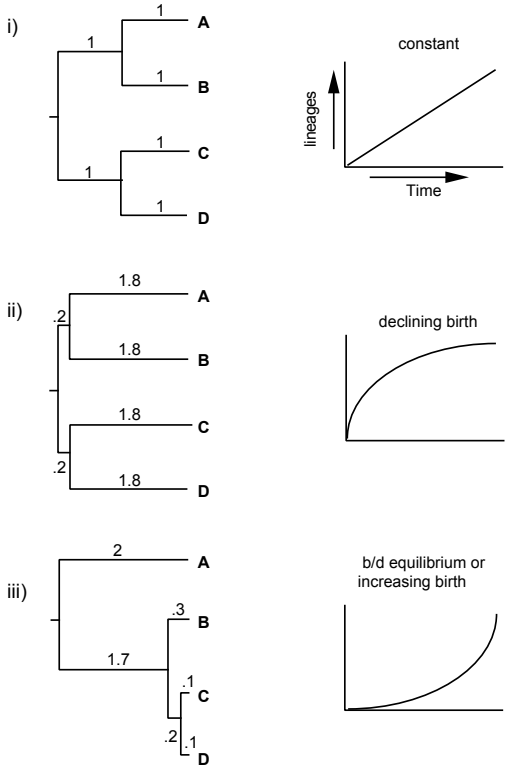
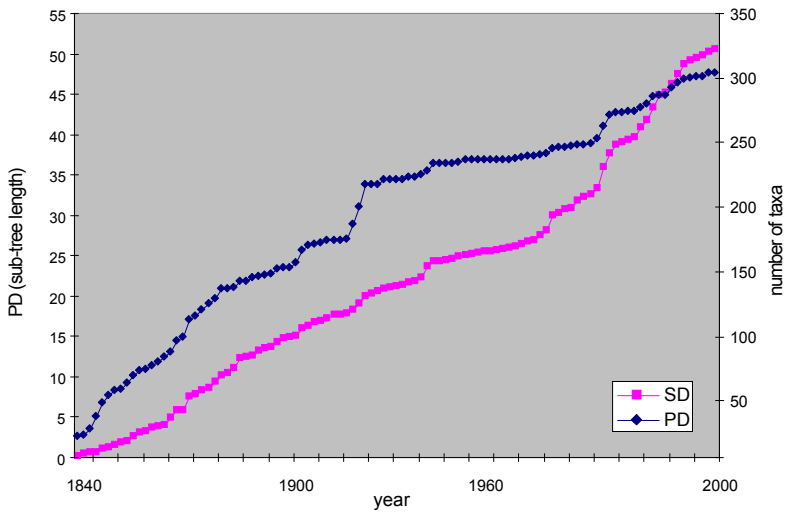
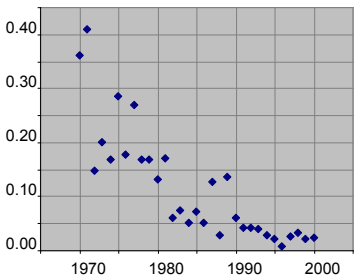


Figure 3. Diversity discovery curve: date at which species either named or specimen collected. PD from mtDNA phylogeny divergence branch lengths (Chapter 2, Figure 7).



New taxa per museum record, last 30 years



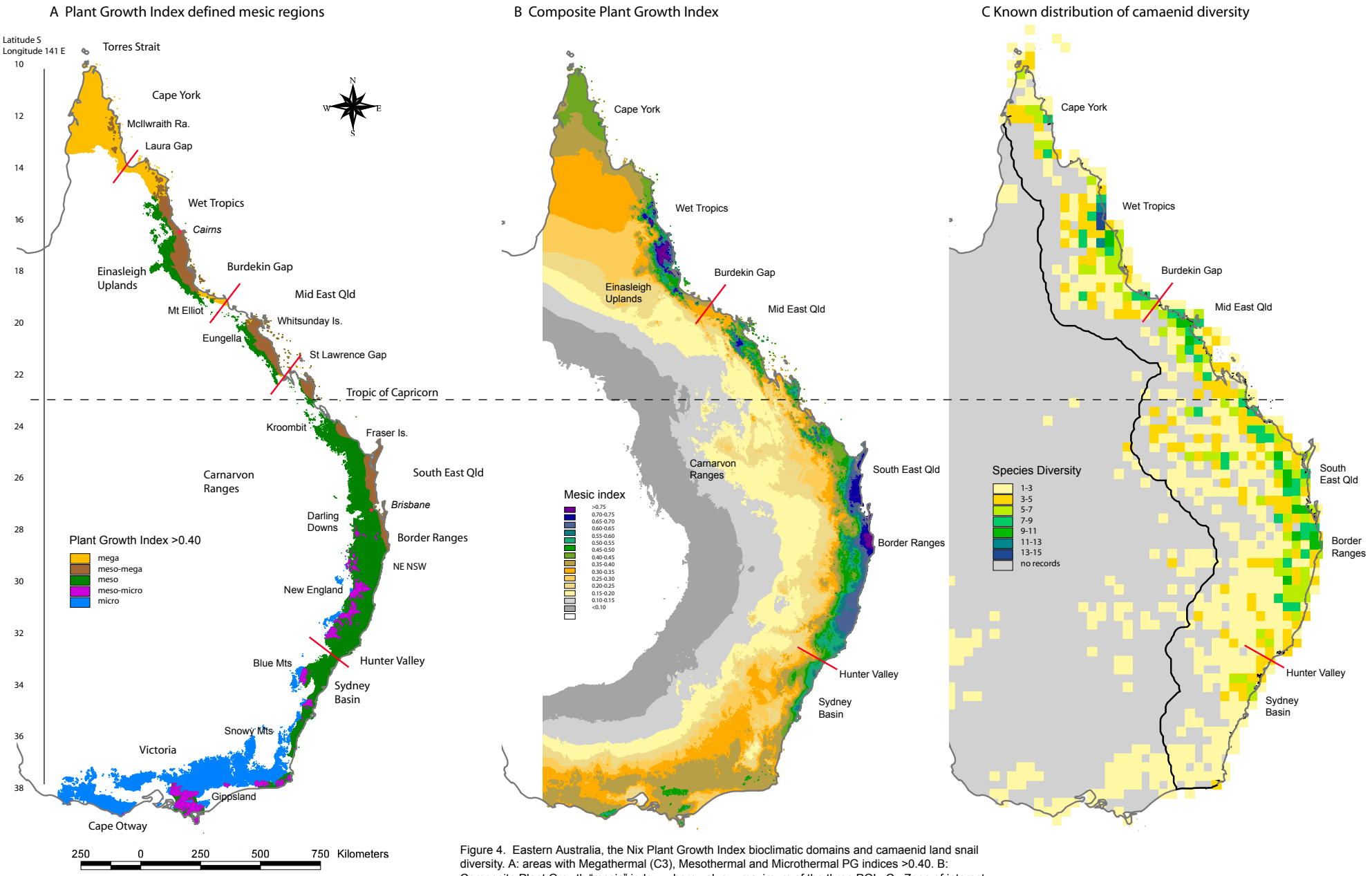
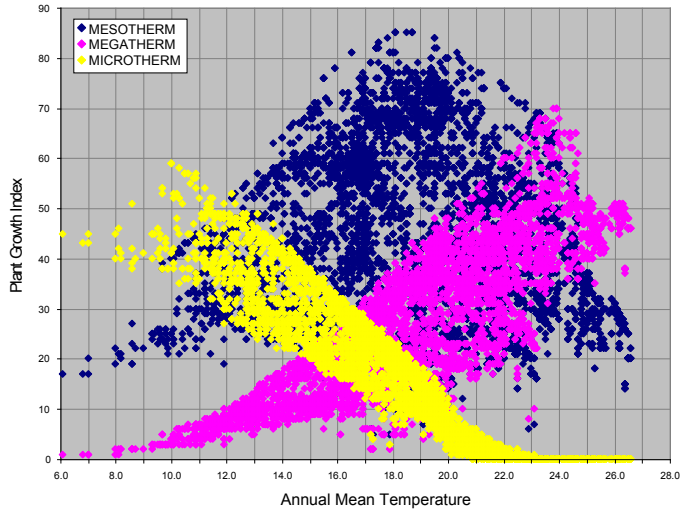


Figure 4. Eastern Australia, the Nix Plant Growth Index bioclimatic domains and camaenid land snail diversity. A: areas with Megathermal (C3), Mesothermal and Microthermal PG indices >0.40. B: Composite Plant Growth "mesic" index, where value = maximum of the three PGI. C: Zone of interest outline against 1/3 degree species diversity (all records). The zone is roughly defined by neighbourhood of records, excluding Victoria and some inland areas with maximum mesic index of <0.20.

Figure 5. Relationship between Plant Growth Indices and (top) Annual Mean Temperature and (bottom) latitude in the eastern Australian camaenid snail distribution database.



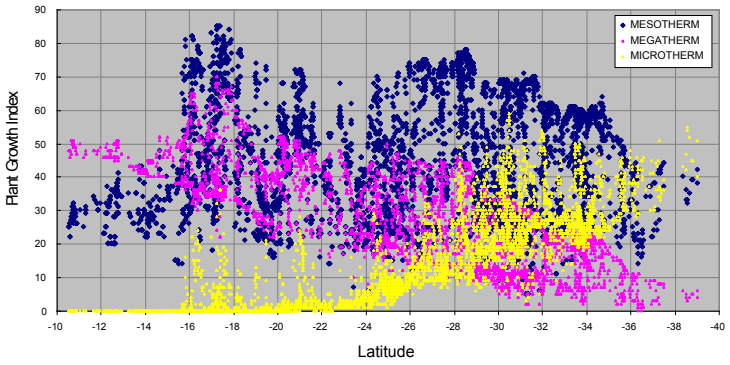


Figure 6. Mesothermal index distribution for region and snails. See Chapter 2 for definition of hadroid lineages

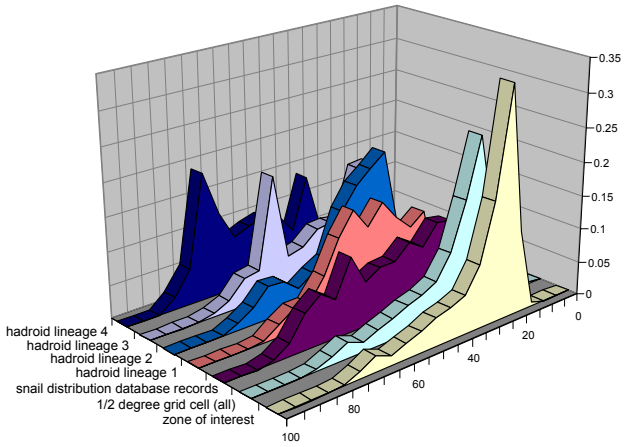


Figure 7. Correlation of species diversity and bioclimatic indices. Average mesothermal and megathermal (c3) PGI for the included set of 1/2 degree grid cells.

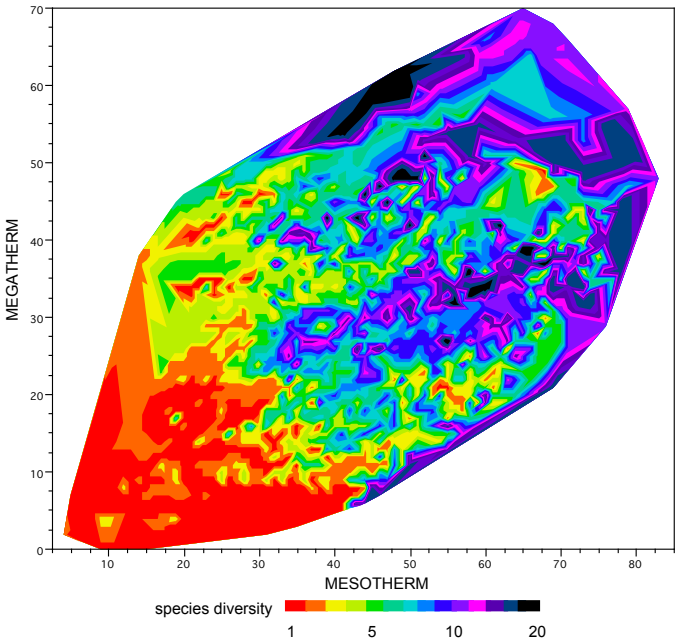


Figure 8. Species diversity by latitude. Top: Species diversity by latitudinal bands of various sizes and centrings. Bottom: Turnover in species per latitude interval (proportion of species found in adjacent intervals)

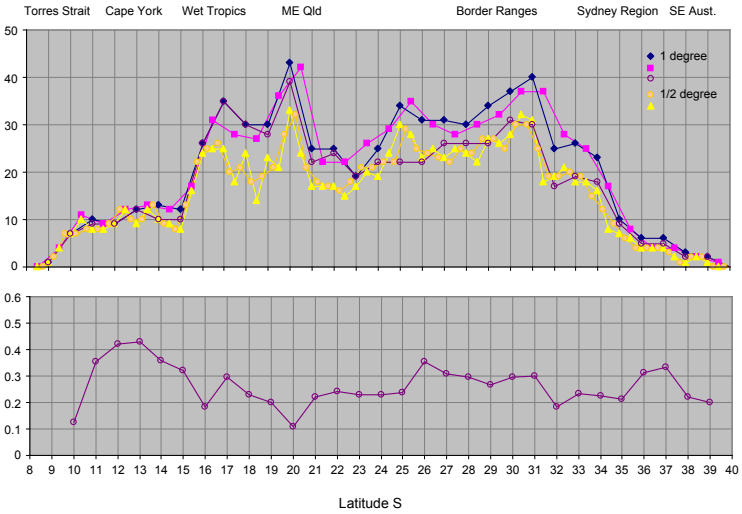


Figure 10. Species Diversity distributions across equal area grid cell scales,. Grids selected for sampling completeness, scale in km. Same graph shown from two perspectives.

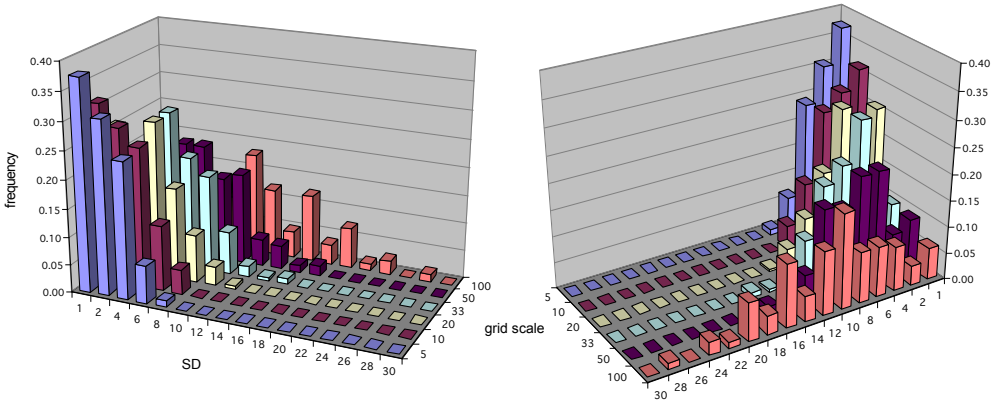


Figure 9. Species range distributions. Left: Maximum linear range (in km). Right: Estimated distribution area (in square km). Matching cumulative frequency distributions below.

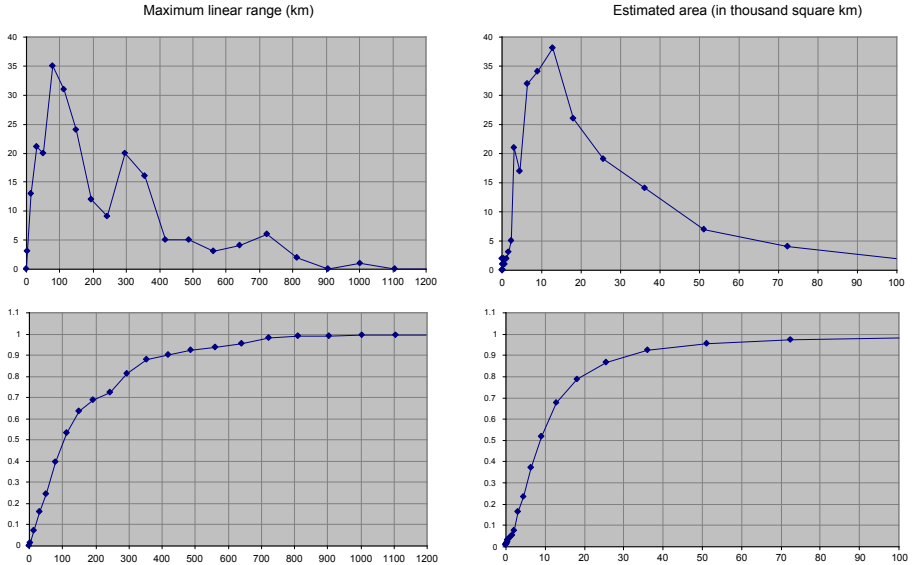
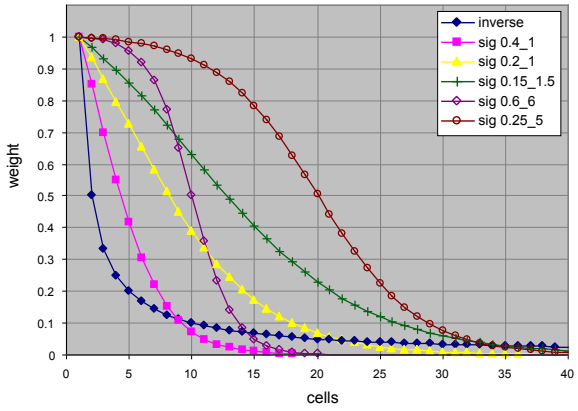


Figure 12. Relationship between endemicity and distribution range for the inverse and sigmoid functions. Various shapes of sigmoid function are shown depending on the shape parameters *a* and *b*, indicated in the legend.



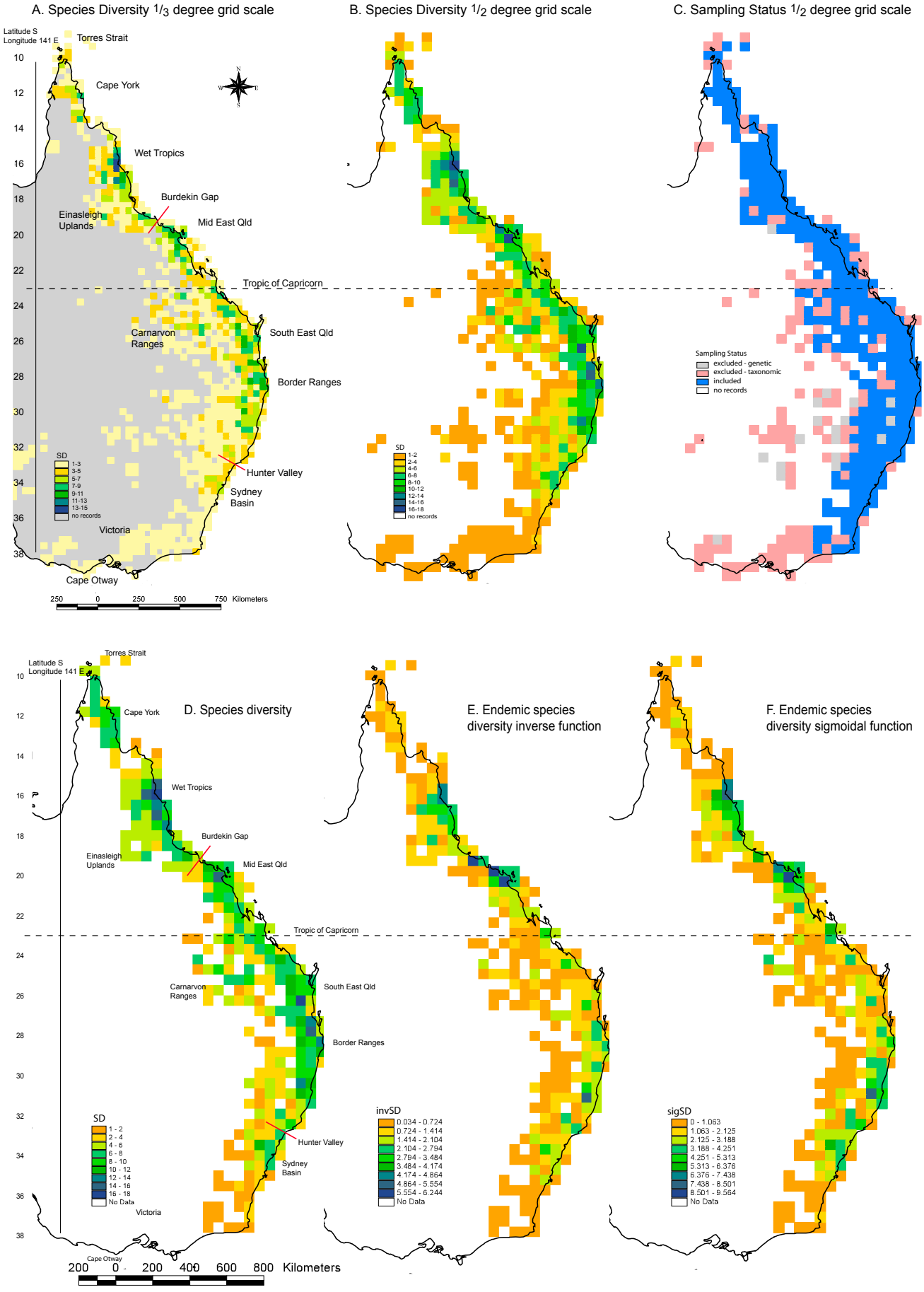
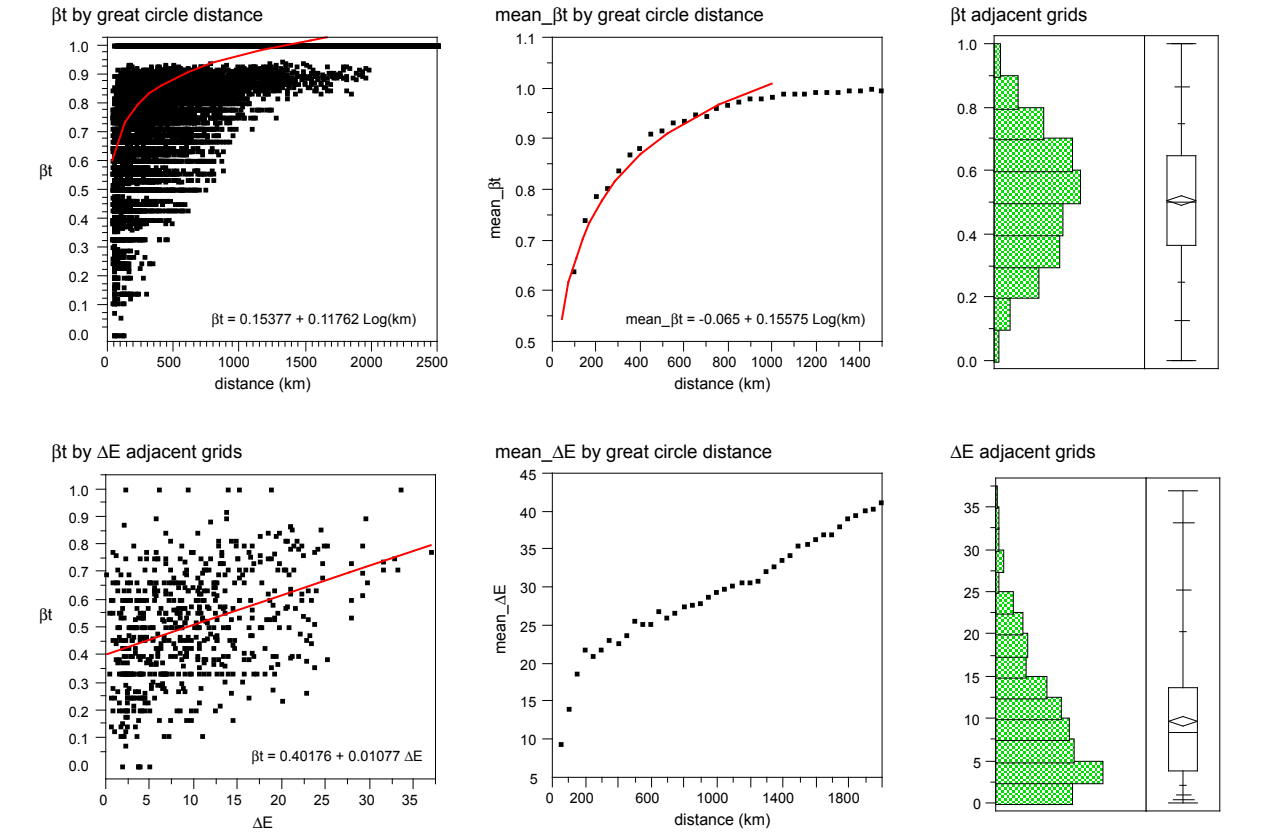


Figure 11. Sampling and species diversity of eastern Australian camaenid land snails. (Top) **A**: species diversity at 1/3 degree grid scale all records, pink cells are records from the Museum of Victoria (MoV); **B**: Species diversity at 1/2 degree grid scale all records; **C**: Sampling status of the 1/2 degree grid cells - pink cells are excluded due to inadequate number of records or taxonomy, grey cells are excluded due to inadequate genetic data. (Bottom) **D**: Species diversity for the included grid cells only; **E**: Endemic species diversity using the inverse function - this is the weighted endemism of Crisp et al. (2001); **F**: Endemic species diversity using the sigmoidal function with parameter values *a*=0.4 *b*=1.



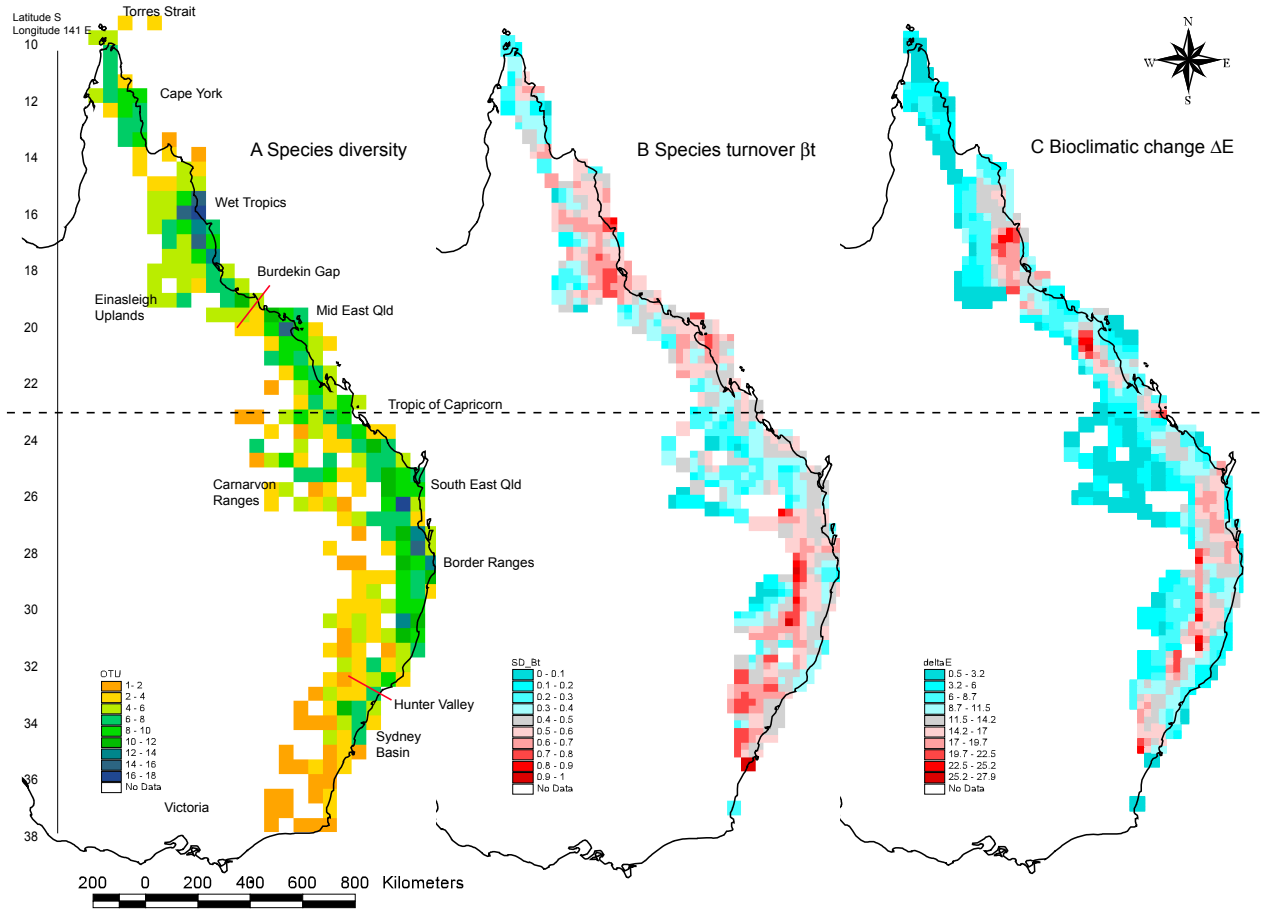


Figure 13. Species and bioclimatic turnover. Only grids with SD>2 have been used for turnover. Top panel shows various statistical characteristics. Mean turnover values are by 50km bins, with log function fitted to distances out to 1,000km only. Bottom panel shows maps. A: Species diversity. B: Species turnover, βt among adjacent grids; each SD grid allows 8 adjacent pairwise comparisons for turnover hence for the turnover maps each grid is sub-divided into 9 sections, with the centre equal to the average of the surrounding pairwise comparisons. C: Bioclimatic change using ΔE, the Euclidean distance among grid mean PGI.

The LINNEAN SOCIETY of London ZOOLOGICAL Journal of the Linnean Society

Zoological Journal of the Linnean Society, 2011, **161**, 531–572. With 7 figures

Beyond the prolegomenon: a molecular phylogeny of the Australian camaenid land snail radiation

ANDREW FORREST HUGALL[1]* and JOHN STANISIC[2]

[1]*School of Earth and Environmental Sciences, University of Adelaide, North Terrace, SA 5005, Australia*
[2]*Queensland Centre for Biodiversity, Queensland Museum, Brisbane, Qld 4000, Australia*

Received 8 September 2008; accepted for publication 4 November 2009

From an analysis of over 900 specimens of camaenid land snails, we have assembled a molecular phylogeny of 327 tips covering > 70% genera across the entire continent of Australia and including > 90% of eastern species. Our approach emphasizes sampling to identify lineage flocks from populations down to build a hierarchical gene-by-taxa tapestry or supermatrix dataset using three mitochondrial genes, then analysed with Markov chain Monte Carlo and fast maximum likelihood methods. Similarity amongst taxa set results suggests missing data cause only minor distortions. This is supplemented by a separate higher level 28S rDNA phylogeny for a global scale perspective. The shallow divergence of Australasian forms, and their nesting within South-East Asian groups within the Helicoidea supergroup extending from Europe to North America, is consistent with the Solem hypothesis of Laurasian immigration of *c.* Miocene origin, and so being more than 400 species in 80-plus genera spread across the continent of Australia from rainforest to desert, forms an immense radiation. There is a major distinction between eastern and western lineages, with some key exceptions. Finer scale patterns of relictual endemics indicate that many ancestral lineages were in place before the major decline and breakup of the Tertiary mesic forest realm that once dominated Gondwanan Australia, and so chart the phylogenetic turnover of ecosystem change from mesic to xeric. The various higher classification schemes proposed all founder on the sheer scale of this radiation. Of 30 polytypic genera tested, at least 18 are not monophyletic, highlighting (1) the repeated radiation of shell forms, and (2) that the current higher taxonomy is unacceptable. Here we provide a phylogenetic and biogeographically condign arrangement as the basis for future elaborations.

 Zoological Journal of the Linnean Society, 2011, **161**, 531–572.
doi: 10.1111/j.1096-3642.2010.00644.x

ADDITIONAL KEYWORDS: biodiversity – Camaenidae – missing data – supermatrix – tree-of-life.

INTRODUCTION

Macroevolutionary analyses of biodiversity are best served by large complete, tree-of-life phylogenies, overcoming reliance on Linnean ranks (e.g. Purvis, 1996; Blackburn & Gaston, 2004; Bininda-Emonds, Cardillo & Jones, 2007). Invertebrate groups showing fine scale patterns of diversification may be ideal; however, description of many of these groups faces practical and conceptual difficulties: the amount of work daunting with basic alpha-taxonomy often incomplete and archaic or otherwise limited higher taxonomies with subjective discrimination, particularly of allopatric diversity (e.g. Ponder & Lunney, 1999; Wheeler & Meier, 2000; Godfray & Knapp, 2004; Dubois, 2007).

In 1990 the pre-eminent American malacologist Alan Solem published a 'Prolegomenon for a Molecular Phylogeny.' of an Australian camaenid land snail radiation, comprising five species (Woodruff & Solem, 1990). In his last monograph on the camaenids of the central and western half of the continent (published posthumously in 1997) he lamented that 'These anticipations [of a biogeographical scenario] have

*Corresponding author.
E-mail: ahugall@unimelb.edu.au
Current address: Department of Zoology, University of Melbourne, Parkville 3010, Victoria, Australia

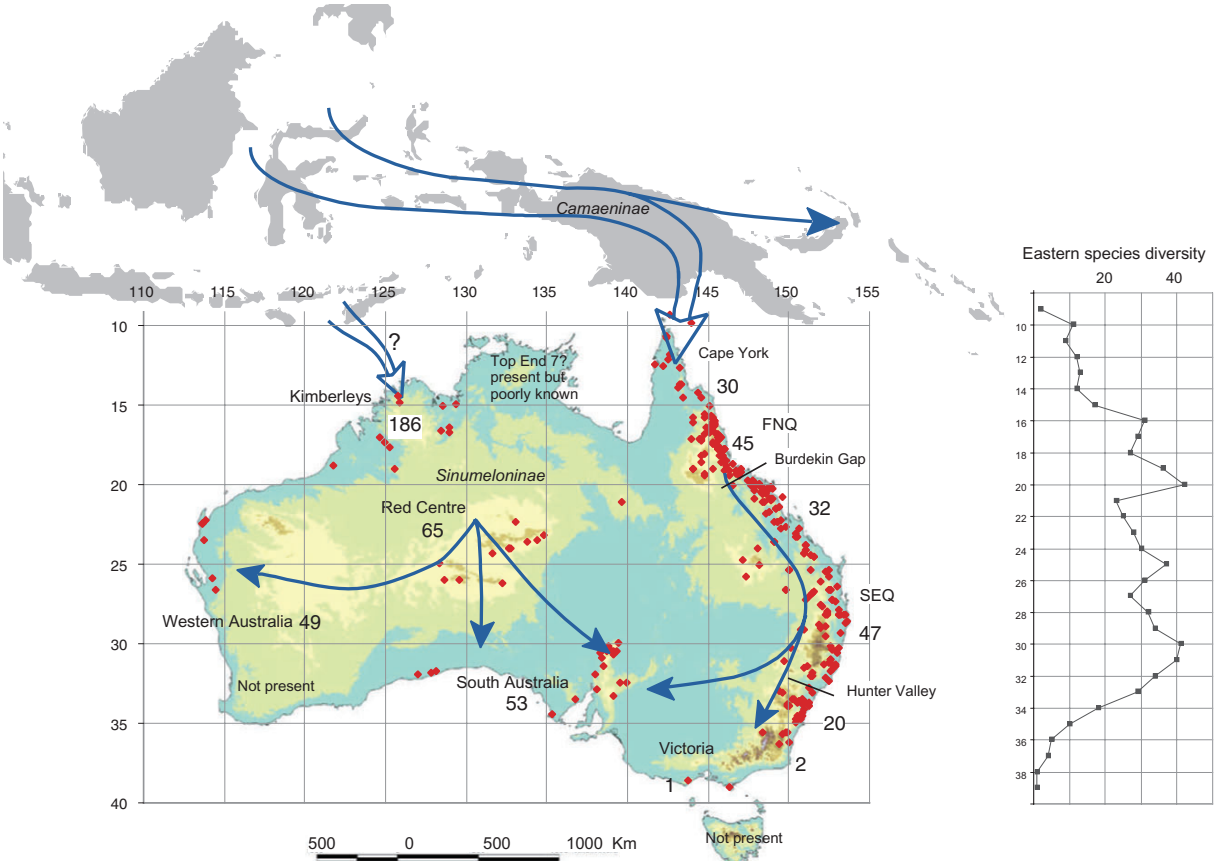


Figure 1. Map of Australia with broad scale patterns of camaenid species diversity and Solem schema of immigration from Laurasia via South-East Asian archipelago. Distribution of samples used in this study shown by red diamonds. Numbers indicate examples of regional species diversity. East coast diversities include informal codes. Graph shows museum database species diversity east of latitude 142°E by one degree bands.

foundered on the fact that only 28 of the 51 genera have been allocated to subfamily units' (Solem, 1997). And this does not even consider the 30-plus genera of eastern taxa.

The Camaenidae (*sensu lato*) are probably the most diverse family of land snails in Australia (see Fig. 1). Based on the broad north–south patterns of species diversity and some stratigraphical evidence, Solem (1979a, b, 1997) proposed that they stem from Miocene immigrants from a large pool of Laurasian taxa, with eastern mesic forests colonized via Melanesia and separate origins for western taxa but there was then little or no phylogenetic evidence for this (Scott, 1996b, 1997). Following the first species description by Férussac in 1819, 19th and early 20th century conchology led to a plethora of names and much duplication, culminating in the huge taxonomic expansion by Iredale (1933, 1937a, b, 1938). Although this has been successively tamed by Solem (1979b, 1997), Burch (1976), Smith (1984, 1992), Clark (2005, 2009), the alpha taxonomy as it stands may still be an underestimate of the real diversity, and the higher systematics is considered a sketch, at best.

These camaenid snails represent a microcosm of the invertebrate biodiversity debate – diverse enough to contain substantial alpha (ten to 20 species), beta (across most habitats), and gamma diversity (400-plus species), abundant enough to represent the bulk of the large terrestrial land snail biota in most of Australia, recondite enough to be refractory to more than a hundred years of taxonomy, but not so diverse as to overwhelm current phylogenetic methods (cf. other invertebrates groups such as beetles; Farrell, 1998). Here we provide a comprehensive molecular phylogeny of the camaenid land snails of eastern Australia, within a framework for the entire group. Sampling from populations up (in at least a large portion of the tree) fulfils two objectives: (1) to assess phylogenetic diversity not limited by the existing formal taxonomy, and (2) to illuminate the complete

shape of the tree necessary for macroevolutionary inference when allied to the distribution data associated with the museum collections (e.g. Hugall, Stanisic & Moritz, 2003; Davies *et al*., 2007; Heard & Cox, 2007; Weir & Schluter, 2007).

Historical biogeography of Australia

The recent (late Tertiary) evolution of the Australian biota is dominated by two trends: (1) colonization from the north by Malesian elements (Hooker, 1859, Burbidge, 1960) as the continent drifted northwards towards Asia; (2) progressive aridification – with occasional oscillations – from the late Miocene leading to fragmentation of a formerly widespread and rich mesic biota and the emergence of a vast xeric zone biota (Bowler, 1982; Archer, Hand & Godthelp, 1991; Adam, 1992; Truswell, 1993). Molecular phylogenetic information is now providing the means for fleshing out these broad themes, such as the antiquity and scale of xeric/desert domain (Bowler, 1982; Mcgowran *et al*., 2004; Fujioka *et al*., 2005), and the significance and timing of Malesian colonization vs. endemics of Gondwanan origin, with the camaenid snails as a possible immigrant element [e.g. Maekawa *et al*. (2003) blattoid cockroaches; Cook & Crisp (2005) *Nothofagus*; Oliver & Sanders (2009) diplodactylid geckoes; Sanders & Lee (2007) elapid snakes].

Camaenid land snails are widely distributed across Australia (see Fig. 1) with areas of high diversity concentrated along the relatively mesic forests east of the Great Dividing Range but also in the north-west (especially the Kimberleys) and Central Australian Ranges (the 'Red Centre'). Along the east coast, species diversity drops markedly in south-eastern Australia (south of the Hunter Valley, *c.* 33°S, regarded as a major biogeographical break; Burbidge, 1960; Adam, 1992) with very low diversity in Victoria, and absence from the south-west corner of the continent and from Tasmania.

Along this large eastern latitudinal 'transect' there is an east–west gradient in species diversity, reflecting rainfall, and general patterns seen in mesic dominant biota (Nix, 1982; Stanisic, 1994). At the species level many camaenid species have narrow ranges (50% of eastern taxa have < 100 km maximum linear range), making it impossible to delineate larger scale biogeographical patterns from species distributions, and higher taxonomic ranks are unreliable or even non-existent. This 'eastern track' may reflect general hypotheses about the connection between eastern Australian and mid-montane New Guinea rainforest (the 'Tumbunan' biota: Schodde & Calaby, 1972) and biogeography of eastern mesic forests (e.g. Cracraft, 1986; Joseph & Moritz, 1994; Crisp, Linder & Weston, 1995).

Hierarchical sampling and nested supermatrix method

To resolve a maximum of lineages with a minimum of sequence, we apply a hierarchically nested supermatrix approach containing a large amount of missing data, analysed by Bayesian and maximum likelihood phylogenetic inference. This is an extension of the supermatrix approach (Gatesy *et al*., 2002; Driskell *et al*., 2004; de Queiroz & Gatesy, 2007; Pirie *et al*., 2008), and recent methodological studies have indicated that it can make the most use of such data with minimal disadvantages because of the large number of taxa and complex 'tapestry' of missing data. The approach takes advantage of the general finding that both more sites and more taxa are valuable in improving phylogenetic inference (Hillis, 1996; Zwickl & Hillis, 2002; Hedke, Townsend & Hillis, 2006), even where there is missing data (Wiens, 2003, 2006; Driskell *et al*., 2004; Wiens *et al*., 2006; Wiens & Moen, 2008), although this depends upon various factors including an appropriate spread of common data and phylogenetic nestedness to minimize artefacts (Philippe *et al*., 2004; Manos *et al*., 2007; Dunn *et al*., 2008; Pirie *et al*., 2008; Lemmon *et al*., 2009). Therefore, we provide various descriptions of how the nested data structure affects phylogenetic inference. A separate 28SrDNA nuclear gene analysis provides a global scale perspective on the relationship of Australasian taxa to other camaenids and the Helicoidea supergroup (as defined by Wade, Mordan & Clarke, 2001; Wade, Mordan & Naggs, 2006; Wade *et al*., 2007).

MATERIAL AND METHODS

Specimens, sampling, and taxonomy

In response to the taxonomic difficulties, the Queensland Museum has, over the last 30 years, undertaken to develop a new complete and thorough geographically based collection consisting of more than 50 000 specimens of camaenids from over 2000 locations along the east coast of Australia. Half of these specimens are stored in alcohol (the 'wet collection') and therefore available for genetic analysis. This collection goes beyond previously recognized taxonomic diversity, emphasizing the need for a thorough modern revision.

Specimens analysed here are largely from collections of the Queensland Museum (QM) and the Field Museum of Natural History (FMNH), with some from the Bishop Museum (BM) and also the Australian

Museum (AM). Some 16S sequences (total of 18; all in clade 3: see Fig. 7) were provided by Stephanie Clark (Clark, 2005). Taxonomy is based on the Zoological Catalogue of Australia (Smith, 1992) as presented in the Australian Faunal Directory web site as of 2007 (AFD 2007), the Solem monographs (Solem, 1979b, 1981a, b, 1984, 1985, 1992, 1993, 1997), and the combined Queensland Museum and Australian Museum databases. Nomenclature is in two or three parts: genus_species_location/subspecies/AFD taxa to be referred to. Genus names are according to AFD. Species names are from a combined QM and AM database (A. F. H., unpubl. data). The third part refers these species to names in the AFD, or location if it is an intraspecific phylogeographical lineage. The AFD contains some names not used in the museum databases, and the museum databases contain numerous as yet to be formally recognized taxon codes of eastern species (east of 142°E). These latter are all labelled as generic Camaenidae, with many to be described in a forth-coming book [Australian Land Snails Volume 1. A field guide to eastern Australian species (Stanisic, Shea, Potter, Griffiths, in press) Publishers: Bioculture Press, Mauritius.] The current AFD uses *languinosa* [sic], *Obstengenia* [sic], and *richmondianus* [sic] but we use original spellings *lanuginosa* (Gude, 1907), *Obsteugenia* (Iredale, 1933, 1938), and *richmondiana* (Reeve, 1852).

The mtDNA dataset described here is from a larger dataset of 839 Australian camaenid individuals and comprises three outgroups and 324 ingroup tips: 164 described (AFD) species, 90 museum database informal codes, 14 Melanesia taxa, and 56 intraspecific phylogeographical lineages. This is laid out in Table 2 listing nomenclature, source, geographical location, and GenBank accession numbers. Of the catalogued Australian fauna, we have 60 of 83 (72%) named genera and 164 of 444 named species (in the AFD). For eastern taxa (east of 142°E) we have genetic data on 31 of 33 AFD genera (all but the monotypic *Discomelon*, and *Meliobba* with a single Australian species), 110 of 133 AFD species, and 98 of 189 coded (but not formally described) types of possible new species. We have no information on 11 eastern AFD taxa and there are 11 AFD species names not used in the museum databases but we have equivalent material for nine of these. Twelve AFD taxa are referred to as belonging within tip lineages based on information in the full 839-specimen database; therefore we have information on 122/133 (92%) of eastern AFD species. Details on missing taxa and discrepancies amongst databases are listed in Supporting Information Table S1. Sampling effort was focused most intensely on the eastern region of Australia, particularly the north-east (latitudes 15–30°S), and particularly for the hadroid group (see Results for redefinition). The rest of the eastern region is slightly less densely sampled, with sparsest sampling of Central, South, and Western Australia, and Melanesia. The results indicate a profound separation of east and west lineages; therefore this sparser sampling is adequate to provide a background framework for the eastern units.

Molecular phylogenetic analysis

Technical details on DNA extraction, PCR, and sequencing are in Supporting Information Appendix S1; here we concentrate on supermatrix methodological questions. Most sequence data were obtained using Chelex DNA extraction, PCR, and automated sequencing. For anatomical reasons land snails were preserved in ethanol only and are therefore, serendipitously, available for genetic analysis. Most specimens were > 10 years old when analysed, up to a maximum of 40 years. As a result some specimens were difficult to PCR and therefore lack the desired amount of data.

28S rDNA analysis

A broad selection of taxa representing the major Australasian lineages were sequenced for a 5′ fragment of the 28S rDNA gene and added to a selection of the higher-level land snail phylogenetic data of Wade *et al*. (2001, 2006, 2007) to provide a global phylogenetic perspective for the Australasian taxa (see Table 1). Sequences were aligned with ClustalX (Thompson *et al*., 1997), and as the alignment was relatively simple, all sites were used. The final data matrix analysed here comprises 26 Australasian and another 43 Helicoidea taxa (*sensu* Wade *et al*., 2001, 2006) with nine outgroups to give 78 taxa; 729 included sites containing 105 parsimony-uninformative and 161 parsimony-informative characters.

Within the Helicoidea, and especially within the Australasian taxa, the level of variation is very low (with many equally parsimonious trees and zero length branches), insufficient to support a meaningful stochastic model-based approach; therefore, maximum parsimony (MP) was emphasized as the most straightforward efficient use of the data for phylogenetic inference. Particularly, Bayesian methods do not allow polytomies and are ill suited to this kind of data (Erixon *et al*., 2003; Felsenstein, 2004; Lewis, Holder & Holsinger, 2005). Because of the enormous number of equally parsimonious trees the following strategy was used to sample the distribution of MP trees: 50 000 trees were collected from a tree bisection-reconnection search using a random sequence addition and collapse = maxbrlens, run to completion. This was repeated 20 times, amounting to

Table 1. Taxa and GenBank accessions used in the 28S rDNA analysis

Region	Genus species	Accession	Location	Family
Australian	*Sphaerospira volgiola*	GQ850894	Broadsound Range, Qld	Camaenidae
	Spurlingia forsteriana	GQ850913	Wangetti, Qld	Camaenidae
	Hadra barneyi	GQ850912	Andoom, Cape York	Camaenidae
	Sphaerospira fraseri	GQ850908	Yandina, Qld	Camaenidae
	Austrochloritis porteri	GQ850888	Lamington NP, Qld	Camaenidae
	Austrochloritis agamemnon	GQ850887	Charmillan Creek, Ravenshoe, Qld	Camaenidae
	Obsteugenia inflecta	GQ850903	Mt Fisher, FNQ	Camaenidae
	Mussonena campbelli	GQ850900	Chillagoe, FNQ	Camaenidae
	Xanthomelon jannellei	GQ850893	Cape Melville, FNQ	Camaenidae
	Ordtrachia australis	GQ850904	Behn River, Kimberleys	Camaenidae
	Amplirhagada mitchelliana	GQ850886	Mitchell Plateau, Kimberleys	Camaenidae
	Jacksonena delicata	GQ850897	Mt Williams, Lamb Range, FNQ	Camaenidae
	Noctepuna mayana	GQ850902	Cooper Creek, Cape Tribulation, Qld	Camaenidae
	Thersites richmondiana	GQ850910	Tennyson Woods, D'Aguilar Range, Qld	Camaenidae
	Papuexul bidwilli	GQ850905	Port Macquarie, NSW	Camaenidae
	Rhynchotrochus macgillivrayi	GQ850907	Big Tableland, FNQ	Camaenidae
	Noctepuna cerea	GQ850901	Helenvale, FNQ	Camaenidae
	Rhagada capensis	GQ850906	North West Cape, WA	Camaenidae
	Trozena morata	GQ850911	Forty Mile Scrub, Qld	Camaenidae
	Chloritisanax banneri	GQ850891	McIlwraith Ranges, Cape York	Camaenidae
	Basedowena gigantea	GQ850889	Mann Ranges, Central Australia	Camaenidae
	Sinumelon bednalli	GQ850909	Alice Springs, Central Australia	Camaenidae
Melanesian	*Chloritis quercina*	GQ850883	Wagina, Solomon Islands	Camaenidae
	Megalacron sp.	GQ850885	Wagina, Solomon Islands	Camaenidae
	Sulcobasis sp.	GQ850892	PNG	Camaenidae
	Papustyla xanthochila	GQ850884	Wagina, Solomon Islands	Camaenidae
Asian	*Helicostyla lignaria*	AY841343	Bohol island, Philippines	Bradybaenidae
	Chloraea intorta	AY841344	Bohol island, Philippines	Bradybaenidae
	Bradybaena similaris	GQ850890	Mt Kinabalu, Borneo	Bradybaenidae
	Nesiohelix bipyramidalis	AY841341	Ryukyu, Japan	Bradybaenidae
	Ainohelix editha	AY841338	Shimamaki, Hokkaido, Japan	Bradybaenidae
	Acusta despecta	AY841337	Japan	Bradybaenidae
	Euhadra sandai	AY014141	Osaka City, Japan	Bradybaenidae
	Aegista vulgivaga	AY014139	(Osaka City, Japan	Bradybaenidae
	Trishoplita hachijoensis	AY841345	Niijima Island, Izu Islands, Japan	Bradybaenidae
	Amphidromus sp.	AY841318	Unknown	Camaenidae
	Moellendorffia tokunoensis	GQ850895	Mt Tanpatsu-zan Tokunoshima Island, Japan	Camaenidae
	Moellendorffia diminuta	AY841330	Ryukyu, Japan	Camaenidae
	Mandarina ponderosa	AY841320	Hahajima, Bonin Islands, Japan	Camaenidae
	Obba rota	AY841328	Bohol Island, Phillippines	Camaenidae
	Nipponchloritis hirasei	GQ850898	Matsune, Wakayama Pref. Japan	Camaenidae
	Nipponchloritis bracteatus	AY841319	Sendai, Japan	Camaenidae
	Satsuma japonica	AY014122	Osaka City, Japan	Camaenidae
	Satsuma jacobii	GQ850899	Nakayama, Kasuga Village, Gifu Pref.	Camaenidae
	Satsuma caliginosa picta	GQ850896	Mt Urabu-dake, Yonaguni-jima Island, Japan	Camaenidae
	Coniglobus mercatorius	AY841324	Kikai Island, Ryukyu, Japan	Camaenidae
Old World	*Marmorana scabriuscula*	AY014133	Sicily	Helicidae
	Theba pisana	AY014135	Sicily	Helicidae
	Cantareus aspersus	AY014128	Kettering, Northamptonshire, UK	Helicidae
	Otala lactea	AY841336	Unknown	Helicidae
	Helix pomatia	AY841333	Pulpit Down, Buckinghamshire, UK	Helicidae

Table 1. *Continued*

Region	Genus species	Accession	Location	Family
	Cepaea nemoralis	AY014130	Marlborough Downs, Wiltshire, UK	Helicidae
	Helicigona lapicida	AY014137	Deepdale, Derbyshire, UK	Helicidae
	Arianta arbustorum	AY014136	Deepdale, Derbyshire, UK	Helicidae
	Trochulus striolatus	AY014124	Deepdale, Derbyshire, UK	Hygromiidae
	Monacha cantiana	AY841332	Pulpit Down, Buckinghamshire, UK	Hygromiidae
	Cernuella virgata	AY014127	Porthcurnick, Cornwall, UK	Hygromiidae
	Cochlicella acuta	AY014126	Porthcurnick, Cornwall, UK	Hygromiidae
New World	*Pleurodonte sinuata*	AY841322	Green Grot Cave, Jamaica	Camaenidae
	Thelidomus asper	AY841321	Windsor, Jamaica	Camaenidae
	Polydontes lima	AY841323	Dorado, Puerto Rico	Camaenidae
	Polydontes undulata	AY014121	Dominican Republic	Camaenidae
	Zachrysia auricoma	AY841326	Nr. Dorado, Puerto Rico	Camaenidae
	Monadenia fidelis	AY014142	Oregon, USA	Helminthoglyptidae
	Cepolis streatori	AY841346	Grand Cayman	Helminthoglyptidae
	Triodopsis alleni	AY841316	Williams Creek, Iowa, USA	Polygyridae
	Mesodon thyroides	AY841315	York Co. Pennsylvania, USA	Polygyridae
	Vespericola columbiana	AY014120	Eugene, Oregon, USA	Polygyridae
	Sagda sp.	AY841347	Windsor, Jamaica	Sagdidae
Outgroups	*Lissachatina fulica*	AY014069	Unknown	Achatinidae
	Caryodes dufresnii	AY014086	Mt. Wellington, Hobart, Tasmania	Caryodidae
	Cerion incanum	AY014060	Florida Keys, USA	Cerionidae
	Discus rotundatus	AY014097	Kirkdale, Derbyshire, UK	Discidae
	Haplotrema vancouverense	AY014090	Eugene, Oregon, USA	Haplotrematidae
	Fastosarion brazieri	AY014099	Mossman, Queensland, Australia	Helicarionidae
	Euglandina rosea	AY014074	Moorea, Tahiti	Spiraxidae
	Gonaxis quadrilateralis	AY014076	Reunion	Streptaxidae
	Succinea putris	AY014057	Southampton, UK	Succineidae

Locality codes: WA, Western Australia; Qld, Queensland; NSW, New South Wales; FNQ, far north Queensland; NP, National Park; PNG, Papua New Guinea.

a search across more than 100 billion rearrangements. The resulting one million MP trees were then filtered down to a total of 787 439 unique MP trees of length 812 steps. These were used to construct a strict consensus tree (Fig. 2). The dataset was also subjected to (1) maximum likelihood (ML) analysis using RAxML v.7.0.4 (Stamatakis, 2006), (2) Bayesian analysis using MrBayes v.3.1.2 (Ronquist & Huelsenbeck, 2003); in both cases applying a general time-reversible model with gamma rate heterogeneity parameter (GTR-G). The results with respect to the position of Australasian taxa were essentially the same as the MP analysis, with the possibility that the Philippines taxa *Helicostyla* and *Chloraea* – but not *Obba* – may be embedded within the clade containing the Australasian taxa (see Supporting Information Fig. S1).

Hierarchical mtDNA supermatrix assembly

We emphasized sequencing more individuals rather than more base pairs per exemplar (to completely define Australian clades). This was followed by a heuristic approach to 'backfilling' a selection of taxa with enough data to stabilize the backbone of the tree (e.g. Morando, Avila & Sites, 2003; Driskell *et al*., 2004; Wiens *et al*., 2005, 2006).

Three mtDNA genes were used: cytochrome oxidase subunit II (COII), and 12S and 16S ribosomal RNA (12S, 16S). The particular choice of genes-by-taxa was determined from initial surveys with one gene (COII or 16S), with extra genes (COII, 16S, and 12S) added to subsets of taxa as groups became more divergent and less robust for any one gene [using simple MP and neighbor-joining (NJ) methods]. Stability of this nested method hinges on having a well-dispersed and structured backbone to hang subgroups upon and ideally some data common to all (Wiens, 2003; Driskell *et al*., 2004; Wiens & Moen, 2008). However, because of technical difficulties (rarity and age of some specimens) some taxa had less sequence than otherwise would be desired.

In a hierarchically nested data supermatrix, ideally all taxa would share at least one common gene (for example a relatively fast evolving mtDNA gene) but

Table 2. Specimens, nomenclature and GenBank accessions used in the mtDNA analyses

Clade	Taxon	16S	COII	12S	Genes	Location
1	*Monteithosites helicostracum*	GQ851007	GQ851270	GQ850940	3	Bakers Blue Mt, FNQ
1	*Sphaerospira coxi*	GQ850999	GQ851261	GQ850930	3	Andromache River, MEQ
1	*Sphaerospira etheridgei*	x	GQ851401	x	1	Andromache River, MEQ
1	*Sphaerospira etheridgei birchi*	GQ851062	GQ851316	x	2	Seaforth, MEQ
1	*Sphaerospira findera* Conder	x	GQ851402	x	1	Conder Hills, MEQ
1	*Sphaerospira findera* Ossa	x	GQ851403	x	1	Mt Ossa, MEQ
1	*Sphaerospira fortasse*	x	GQ851367	x	1	Haslewood Island, MEQ
1	*Sphaerospira gavisa*	x	GQ851363	x	1	Wilsons Beach, MEQ
1	*Sphaerospira incei curtisiana*	GQ851223	GQ851382	x	2	Targinie, Qld
1	*Sphaerospira incei lessoni*	x	GQ851375	x	1	Johannsens Caves, Qld
1	*Sphaerospira macleayi*	x	GQ851366	x	1	Hook Island, MEQ
1	*Sphaerospira macneilli*	x	GQ851370	x	1	Hook Island, MEQ
1	*Sphaerospira mattea*	GQ851106	GQ851351	x	2	Dawson Valley, Taroom, Qld
1	*Sphaerospira mazee*	x	GQ851405	x	1	Seaview Range, FNQ
1	*Sphaerospira mazee* WT9	x	GQ851395	x	1	Jourama Falls NP, Seaview Range, FNQ
1	*Sphaerospira mitifica*	x	GQ851376	x	1	Palm Island, FNQ
1	*Sphaerospira mourilyani*	GQ851040	GQ851301	GQ850978	3	Cardwell Range, FNQ
1	*Sphaerospira mourilyani*	GQ851103	GQ851348	x	2	Kurramine Beach, FNQ
1	*Sphaerospira rawnesleyi*	x	GQ851378	x	1	Mt Elliot, FNQ
1	*Sphaerospira reducta*	GQ851073	GQ851326	x	2	Bulburin, Qld
1	*Sphaerospira sardalabiata*	x	GQ851386	x	1	Horseshoe Bay, MEQ
1	*Sphaerospira starena*	GQ851002	GQ851265	GQ850934	3	Cedar Falls, MEQ
1	*Sphaerospira thorogoodi*	x	GQ851368	x	1	Andromache River, MEQ
1	*Sphaerospira tomsoni*	x	GQ851407	x	1	Pallarenda, FNQ
1	*Sphaerospira volgiola*	GQ851076	GQ851328	x	2	Broadsound Range, Qld
1	*Sphaerospira zebina*	GQ851049	GQ851310	GQ850990	3	Rishton Scrub, FNQ
1	*Sphaerospira zebina* EU4	x	GQ851409	x	1	Mt Zero, Paluma, FNQ
1	Camaenidae BL2	x	GQ851393	x	1	Glenprairie Station, Qld
1	Camaenidae BL23	GQ851066	GQ851320	x	2	Cape Upstart, MEQ
1	Camaenidae BL24	GQ851061	GQ851315	x	2	Mt Aberdeen, MEQ
1	Camaenidae BL25	x	GQ851364	x	1	Mt Inkerman, MEQ
1	Camaenidae BL26	x	GQ851362	x	1	Collinsville, MEQ
1	Camaenidae BL28	GQ851065	GQ851318	x	2	Pine Mts, MEQ
1	Camaenidae BL29	x	GQ851369	x	1	Collinsville, MEQ
1	Camaenidae BL3	GQ851100	GQ851346	x	2	Connors Range, MEQ
1	Camaenidae BL33	x	GQ851379	x	1	Mt Elliot NP, Cockatoo Creek, FNQ
1	Camaenidae BL35	GQ851071	GQ851324	x	2	Mt Abbot, MEQ
1	Camaenidae BL39	x	GQ851396	x	1	Hervey Range, FNQ
1	Camaenidae BL4	GQ851018	GQ851280	GQ850955	3	Sarina, Lotus Creek, MEQ
1	Camaenidae BL40	x	GQ851365	x	1	Townsville, Mt Stuart, FNQ
1	Camaenidae BL46	GQ851011	GQ851273	GQ850943	3	Cullen Island, MEQ
1	Camaenidae BL47	GQ851068	GQ851322	x	2	Seaforth, MEQ
1	Camaenidae BL47 challisi	x	GQ851394	x	1	St Lawrence, MEQ
1	Camaenidae BL61	x	GQ851383	x	1	Longford Creek, MEQ
1	Camaenidae BL62	x	GQ851385	x	1	Mt Bramston, MEQ
1	Camaenidae BL63	x	GQ851384	x	1	Bowen, MEQ
1	Camaenidae EU16	GQ851111	GQ851355	x	2	Paluma Range, FNQ
1	Camaenidae SQ14	GQ851074	GQ851327	x	2	Dan Dan SF, Qld
2	*Hadra barneyi*	GQ851047	GQ851308	GQ850988	3	Andoom, Cape York
2	*Hadra bartschi*	x	GQ851372	x	1	Murray Island, Torres Strait
2	*Hadra bipartita* Cow	GQ851013	GQ851275	GQ850950	3	Cow Bay, FNQ
2	*Hadra bipartita* Wonga	x	GQ851404	x	1	Wonga Beach, FNQ
2	*Hadra funiculata*	GQ851081	GQ851329	x	2	Saibai Island, Torres Strait
2	*Hadra webbi*	x	GQ851388	x	1	Kurrimine Beach, FNQ
2	*Jacksonena rudis*	GQ851016	GQ851278	GQ850953	3	Atherton Tableland, FNQ
2	*Spurlingia monticola*	GQ851093	GQ851341	x	2	Mt Carbine, FNQ
2	*Spurlingia dunkiensis*	GQ851033	GQ851295	GQ850971	3	Cardwell, FNQ
2	*Spurlingia forsteriana* Lakeland	x	GQ851408	x	1	Carols Bridge, Lakeland Downs, FNQ
2	*Spurlingia forsteriana* Mellville	GQ851105	GQ851350	x	2	Cape Mellville NP, FNQ

Table 2. *Continued*

Clade	Taxon	16S	COII	12S	Genes	Location
2	*Spurlingia forsteriana* Normanby	GQ851048	GQ851309	GQ850989	3	Lakeland Downs, Normanby River, FNQ
2	*Spurlingia gemma*	x	GQ851373	x	1	Chillagoe, FNQ
2	*Spurlingia portus*	x	GQ851392	x	1	Weary Bay, FNQ
2	*Spurlingia praehadra*	GQ851034	GQ851296	GQ850974	3	Chillagoe, FNQ
2	*Spurlingia tinarooensis*	x	GQ851374	x	1	Mt Molloy, FNQ
2	*Spurlingia tinarooensis* Hann	GQ851096	GQ851343	x	2	Hann Tableland, FNQ
2	Camaenidae CY10 Bamaga	x	GQ851387	x	1	Bamaga, Cape York
2	Camaenidae CY10 McIlwraith	GQ851082	GQ851330	x	2	McIlwraith Ranges, Cape York
2	Camaenidae EU12	GQ851104	GQ851349	x	2	Broken R iver, Qld
2	Camaenidae EU14	x	GQ851390	x	1	Palmerville Road, FNQ
2	Camaenidae EU15	x	GQ851391	x	1	Mt Mulligan, FNQ
2	Camaenidae EU17	x	GQ851371	x	1	Narelle Wilson Cave, Palmerville, FNQ
2	Camaenidae EU8	x	GQ851377	x	1	Mt Surprise, FNQ
2	Camaenidae EU9	x	GQ851381	x	1	Greenvale, Qld
2	Camaenidae WT13	x	GQ851389	x	1	Herberton, FNQ
	Camaenidae BL7	GQ851027	GQ851289	GQ850964	3	Salvator Rosa NP, Qld
3	*Adclarkia dawsonensis*	GQ851005	GQ851268	GQ850937	3	Taroom, Mt Rose Station, Qld
3	*Contramelon howardi*	GQ851155	x	x	1	Nepouie Spring Gorge, SA
3	*Cooperconcha bunyerooana*	GQ851004	GQ851267	GQ850936	3	Bunyeroo Gorge, SA
3	*Cupedora lincolnensis*	GQ851157	x	x	1	Cowell, Eyre Peninsula, SA
3	*Cupedora lorioliana*	GQ851070	GQ851323	x	2	Willochra Creek, Flinders Ranges, SA
3	*Cupedora luteofusca*	GQ851075	x	GQ850944	2	Mt Remarkable, SA
3	*Cupedora patruelis*	GQ851069	x	GQ850939	2	Mt Dutton, SA
3	*Cupedora sublorioliana*	GQ851217	x	x	1	Gammon Ranges, Mt Aroona, SA
3	*Galadistes liverpoolensis*	GQ851077	x	GQ850945	2	Mt Baloola, Gunnedah, NSW
3	*Glyptorhagada janaslini*	GQ851162	x	x	1	Paralana Hot Springs Flinders Ranges, SA
3	*Glyptorhagada kooringensis*	GQ851079	x	GQ850948	2	Mt Pullen, SA
3	*Glyptorhagada silveri*	GQ851080	x	GQ850949	2	Pepuarta Hill, SA
3	*Glyptorhagada tattawuppana*	GQ851163	x	x	1	Yunta, SA
3	*Meridolum grayi*	GQ851248	x	x	1	Wyee, NSW
3	*Meridolum middenense*	GQ851085	GQ851333	x	2	Hornsby, NSW
3	*Meridolum bennetti*	GQ851218	x	x	1	Bell, Kondai Creek, Qld
3	*Meridolum corneovirens*	GQ851226	x	x	1	Sydney, Prospect Reservoir, NSW
3	*Meridolum gulosa* Bass	GQ851249	x	x	1	Bass Point NR, NSW
3	*Meridolum gulosa* Keira	GQ851250	x	x	1	Mt. Keira, NSW
3	*Meridolum jervisensis* Gerringong	GQ851246	x	x	1	Gerringong, Seven Mile Beach, NSW
3	*Meridolum marshalli*	GQ851024	GQ851286	GQ850961	3	Royal NP, NSW
3	*Meridolum maryae* Maroubra	GQ851251	x	x	1	Maroubra Beach, NSW
3	*Meridolum middenense* Narrabeen	GQ851086	GQ851334	x	2	Narrabeen Lagoon, NSW
3	*Pallidelix greenhilli*	GQ851222	x	x	1	Charlevue Ck, Qld
3	*Pallidelix greenhilli expeditionis*	GQ851227	x	x	1	Ka Ka Mundi NP, Bunbuncundoo Spring, Qld
3	*Pallidelix greenhilli* Zamia	GQ851215	x	x	1	Mt Zamia, Qld
3	*Petraconcha griffithsi*	GQ851242	x	x	1	Hawkins Hill, Hill End, NSW
3	*Pommerhelix carmelae*	GQ851239	x	x	1	Seal Rocks, NSW
3	*Pommerhelix stanisici*	GQ851171	x	x	1	Sofala, NSW
3	*Pommerhelix depressa*	GQ851023	GQ851285	GQ850960	3	Jenolan Caves
3	*Pommerhelix duralensis*	GQ851020	GQ851282	GQ850957	3	Hornsby, NSW
3	*Pommerhelix mastersi*	GQ851244	x	x	1	Buckenbowra SF, NSW
3	*Pommerhelix mastersi* Cleatmore	GQ851170	x	x	1	Cleatmore Caves, NSW
3	*Pommerhelix monacha*	GQ851237	x	x	1	Yarramundi Regional Park, NSW
3	*Pommerhelix monacha* BlueMts	GQ851238	x	x	1	Blue Mountains, Sassafras Gully, NSW
3	*Ponderconcha murphyi*	GQ851160	x	x	1	Coffs Harbour, NSW
3	*Ponderconcha gilberti*	GQ851021	GQ851283	GQ850958	3	Gambubal, Qld
3	*Ponderconcha ianthostoma*	GQ851084	GQ851332	x	2	Girraween NP, Qld
3	*Ventopelita leucocheilus*	GQ851214	GQ851361	x	2	Cambridge Plateau, NSW
3	*Ventopelita mansueta*	GQ851112	x	GQ850985	2	Toowoomba, Qld

Table 2. *Continued*

Clade	Taxon	16S	COII	12S	Genes	Location
3	Camaenidae BL12	GQ851072	GQ851325	x	2	Dulacca, Qld
3	Camaenidae BL13	GQ851210	GQ851357	x	2	Dalby, Qld
3	Camaenidae BL17	GQ851006	GQ851269	GQ850938	3	Dawson Valley, Boggomoss Station, Qld
3	Camaenidae MV14	GQ851231	x	x	1	Point Plomer, Port Macquarie, NSW
3	Camaenidae MV2	GQ851078	x	GQ850946	2	Kempsey, NSW
3	Camaenidae NE14	GQ851113	x	GQ850986	2	Mt Mackenzie, NSW
3	Camaenidae NE16	GQ851243	x	x	1	Macintyre R Falls, NSW
3	Camaenidae NE18	GQ851252	x	x	1	MacIntyre Falls, NSW
3	Camaenidae SN16	GQ851087	GQ851335	x	2	Lansdowne SF, NSW
3	Camaenidae SN17	GQ851236	x	x	1	Cessnock, Pokolbin, NSW
3	Camaenidae SN19 Kangaroo	GQ851169	x	x	1	Kangaroo Valley, NSW
3	Camaenidae SN19 Saddleback	GQ851245	x	x	1	Saddleback Mt. NSW
3	Camaenidae SN20	GQ851240	x	x	1	Annangrove Park, NSW
3	Camaenidae sp. BlueMts	GQ851247	x	x	1	Mt. Victoria, Blue Mountains, NSW
3	Camaenidae sp. Elizabeth	GQ851241	x	x	1	Elizabeth Lookout, NSW
3	Camaenidae SQ9	GQ851022	GQ851284	GQ850959	3	Rome Creek, Bunya Mts, Qld
	Vidumelon wattii	GQ851209	x	x	1	Hale River, NT
4	*Gnarosophia bellendenkerensis* Finnegan	AY151072	GQ851317	x	2	Mt Finnegan, FNQ
4	*Gnarosophia bellendenkerensis* Kirrama	AY151070	GQ851263	AY151061	3	Koomboolooba Dam, Qld
4	*Gnarosophia bellendenkerensis* Lamb	GQ851000	GQ851262	AY151062	3	Mt Sheridan, FNQ
4	*Gnarosophia bellendenkerensis* Lewis	x	GQ851356	AY151063	2	Mt Lewis, FNQ
4	*Gnarosophia bellendenkerensis* Mackay	AY151076	AY048390	x	2	Mt Mackay, FNQ
4	*Gnarosophia bellendenkerensis* Thornton	AY151073	GQ851319	x	2	Thornton Peak, FNQ
4	*Sphaerospira arthuriana*	x	AY151311	x	1	Scawfell Island, MEQ
4	*Sphaerospira blomfieldi* Gurgeena	AY151079	AY151298	x	2	Gurgeena Plateau, Qld
4	*Sphaerospira blomfieldi* Kroombit	GQ851037	AY048419	AY151055	3	Kroombit Tops,Qld
4	*Sphaerospira blomfieldi* Larcom	x	AY151306	x	1	Targinie, Qld
4	*Sphaerospira fraseri* Connodale	x	AY151316	x	1	Connodale Ranges, Qld
4	*Sphaerospira fraseri* Glorious	x	AY151326	x	1	Mt Glorious, Qld
4	*Sphaerospira fraseri* Peregian	AY151080	AY151299	x	2	Peregian Beach, Qld
4	*Sphaerospira fraseri* Spicers	AY151067	GQ851300	AY151058	3	Spicers Gap, Qld
4	*Sphaerospira fraseri* Stotts	x	AY151333	x	1	Stotts Island, NSW
4	*Sphaerospira informis* Blackwood	GQ851039	AY151294	AY151059	3	Mt Blackwood, MEQ
4	*Sphaerospira informis* Cameron	x	AY151340	x	1	Cameron Creek, MEQ
4	*Sphaerospira mortenseni*	AY151078	AY151297	x	2	Byfield, Qld
4	*Sphaerospira oconnellensis* Cameron	AY151069	AY151295	AY151060	3	Cameron Creek, MEQ
4	*Sphaerospira oconnellensis* Hillsborough	AY151081	AY151300	x	2	Cape Hillsborough, MEQ
4	*Sphaerospira rockhamptonensis*	AY151077	AY151296	x	2	Johannsens Caves, Qld
4	*Sphaerospira sidneyi* Bauple	x	AY151314	x	1	Bauple, Qld
4	*Sphaerospira sidneyi* Nangur	x	AY151328	x	1	Nangur SF, Qld
4	Camaenidae SQ1	GQ851038	GQ851299	AY151057	3	Bulburin SF, Qld
4	Camaenidae CY1	AY151082	AY151301	x	2	Iron Range, Cape York
4	Camaenidae WT1	GQ851010	AY151292	AY151056	3	Thornton Peak, FNQ
4	Camaenidae SQ4	GQ851026	GQ851288	GQ850963	3	North Aramara SF, Qld
4	*Mussonena spinei*	GQ851089	GQ851337	x	2	Kroombit Tops, Qld
4	*Offachloritis dryanderensis*	GQ851184	x	x	1	Mt Dryander, MEQ
4	*Ramogenia challengeri*	GQ851099	GQ851345	x	2	Tweed Heads, NSW
	Chloritobadistes victoriae	GQ850997	GQ851259	GQ850928	3	Otway Ranges, VIC
	Aslintesta camelus	GQ851054	x	GQ850919	2	Gammon Ranges, SA
5	*Austrochloritis separanda*	GQ851136	x	x	1	Gunialda, SEQ
5	*Austrochloritis separanda* Gatton	GQ851057	GQ851312	x	2	Gatton, SEQ
5	*Calvigenia blackmani* BR5	GQ851067	GQ851321	x	2	Keperra, Brisbane
5	*Gloreugenia blackalli*	GQ851161	x	x	1	Shute Harbour, Qld
5	*Gloreugenia cognata*	GQ851211	GQ851358	x	2	Rockhampton, Mt Archer, Qld
5	*Gloreugenia coxeni*	GQ851234	GQ851400	x	2	Andromache River, Qld
5	*Gloreugenia praecursoris*	GQ851138	x	x	1	Eungella, MEQ

Table 2. *Continued*

Clade	Taxon	16S	COII	12S	Genes	Location
5	*Gloreugenia praecursoris* MQ3	GQ851118	x	x	1	Eungella, MEQ
5	*Melostrachia acuticostata*	GQ851168	x	x	1	Copperfield River Gorge, FNQ
5	*Melostrachia glomerans*	GQ851019	GQ851281	GQ850956	3	Almaden, Qld
5	*Neveritis aridorum*	GQ851029	GQ851291	GQ850966	3	Bell, SEQ
5	*Neveritis misella*	GQ851224	x	x	1	Blackdown Tableland, Qld
5	*Neveritis poorei*	GQ851158	x	x	1	Paluma FNQ
5	*Obsteugenia inflecta*	GQ851091	GQ851339	x	2	Mt Fisher, FNQ
5	*Ramogenia lanuginosa*	GQ851193	x	x	1	Dan Dan Scrub, Qld
5	*Ramogenia mucida*	GQ851194	x	x	1	Connors Range, MEQ
5	Camaenidae BL56	GQ851182	x	x	1	Cape Cleveland, Qld
5	Camaenidae BL56 poorei	GQ851183	x	x	1	Maiden Mountain, Qld
5	Camaenidae BL6	GQ851055	x	GQ850922	2	Carnarvon NP, Qld
5	Camaenidae BR4	GQ851120	x	x	1	Conondale Ranges, Qld
5	Camaenidae MQ2	GQ851125	x	x	1	Flame Tree Ck, MEQ
5	Camaenidae SQ6	GQ851028	GQ851290	GQ850965	3	Bulburin SF, Qld
5.5	*Austrochloritis agamemnon* Bluewater	GQ851124	x	x	1	Blue Water Range, FNQ
5.5	*Austrochloritis agamemnon* Kirrama	GQ851060	GQ851314	x	2	Kirrama SF, FNQ
5.5	*Austrochloritis agamemnon* Tully	GQ851117	x	x	1	Tully, FNQ
5.5	*Austrochloritis astaeus*	GQ851116	x	x	1	Cardwell Range, FNQ
5.5	*Austrochloritis pusilla*	GQ850996	GQ851258	GQ850927	3	Kearneys Falls, FNQ
5.5	Camaenidae WT4	GQ851058	GQ851313	x	2	Thornton Peak, FNQ
5.5	Camaenidae WT8	GQ851122	x	x	1	Mt Father Clancy, FNQ
5.5	Camaenidae WT2	GQ851145	x	x	1	Cow Bay, FNQ
5.5	Camaenidae WT5	GQ851132	x	x	1	Lambs Head, FNQ
5.5	Camaenidae WT6	GQ851059	x	GQ850923	2	Black Mt, FNQ
5.5	Camaenidae WT7	GQ850993	GQ851255	GQ850924	3	North Bell Peak, FNQ
6	*Gloreugenia hedleyi*	GQ851012	GQ851274	GQ850947	3	Hann Tableland, FNQ
6	*Mussonena campbelli*	GQ851025	GQ851287	GQ850962	3	Chillagoe, FNQ
6	*Tolgachloritis jacksoni*	GQ851042	GQ851303	GQ850980	3	Herberton, FNQ
6	Camaenidae EU10	GQ851156	x	x	1	Rosella Plains Station, FNQ
6	Camaenidae EU20	GQ851205	x	x	1	Broken River, FNQ
7	*Austrochloritis buxtoni*	GQ850995	GQ851257	GQ850926	3	Punsand Bay, Cape York
7	*Austrochloritis layardi* CY	GQ851143	x	x	1	Heathlands, Cape York
7	*Austrochloritis layardi* McIvor	GQ851144	x	x	1	McIvor River, FNQ
7	*Chloritid sp.*	GQ851153	x	x	1	Papua New Guinea
7	*Chloritis conamcephala*	GQ851114	x	x	1	Wagina, Solomon Islands
7	*Chloritis eustoma*	GQ851063	x	GQ850931	2	Woodlark Island, Guasopa, PNG
7	*Chlorits minnigerodei*	GQ851148	x	x	1	Normanby Island, PNG
7	*Sulcobasis* sp. Crater	GQ851017	GQ851279	GQ850954	3	Crater Mountain, Papua New Guinea
7	*Sulcobasis* sp. PNG	GQ851008	GQ851271	GQ850941	3	Papua New Guinea
7	*Sulcobasis* sp. Wau	GQ851009	GQ851272	GQ850942	3	Wau, Papua New Guinea
7	*Torresitrachia torresiana*	GQ851043	GQ851304	GQ850981	3	Myall Creek, Cape York
8	*Jacksonena delicata* Carbine	GQ851230	GQ851399	x	2	Carbine Tableland, FNQ
8	*Jacksonena delicata* Kirrama	GQ851229	GQ851398	x	2	Mt Kooroomool Kirrama Range, FNQ
8	*Jacksonena delicata* Lamb	GQ851014	GQ851276	GQ850951	3	Lamb Range, Mt Williams, FNQ
8	*Jacksonena delicata* Tribulation	GQ851232	x	x	1	Cape Tribulation, FNQ
8	*Noctepuna mayana*	GQ851031	GQ851293	GQ850969	3	Cape Tribulation, FNQ
8	*Semotrachia bagoti*	GQ851198	x	x	1	Central Australia
8	*Semotrachia esau*	GQ851199	x	x	1	Palm Valley, Central Australia
8	*Semotrachia setigera*	GQ851101	x	GQ850977	2	Bitter Springs Creek, NT
8	*Thersites darlingtoni*	GQ851044	GQ851305	GQ850983	3	Binna Burra, Qld
8	*Thersites mitchellae*	GQ851107	GQ851352	x	2	Byron Bay, NSW
8	*Thersites novaehollandiae* Burraga	x	GQ851406	x	1	Burraga Swamp, NSW
8	*Thersites novaehollandiae* Dorrigo	GQ851235	x	x	1	Dorrigo Plateaux, NSW
8	*Thersites novaehollandiae* Gibraltar	GQ851233	x	x	1	Gibraltar Range, NSW
8	*Thersites novaehollandiae* Superbus	GQ851041	GQ851302	GQ850979	3	Mt Superbus, Qld
8	*Thersites richmondiana* Connondale	GQ851109	GQ851353	x	2	Connondale Ranges, Qld
8	*Thersites richmondiana* Tennyson	GQ851110	GQ851354	x	2	Tennyson Woods, Qld

Table 2. *Continued*

Clade	Taxon	16S	COII	12S	Genes	Location
9	*Austrochloritis ascensa*	GQ851126	x	x	1	Barrington Tops NP, NSW
9	*Austrochloritis bellengerensis*	GQ851228	x	x	1	Bellengen SF, NSW
9	*Austrochloritis brevipila*	GQ851141	x	x	1	Dorrigo NP, NSW
9	*Austrochloritis disjuncta*	GQ851127	x	x	1	Booti Booti NP, NSW
9	*Austrochloritis metuenda*	GQ851225	x	x	1	Robertson NR, NSW
9	*Austrochloritis nambucca*	GQ850991	GQ851253	GQ850920	3	North Brother Mt, NSW
9	*Austrochloritis nundinalis*	GQ851139	x	x	1	Baldy Knob, NSW
9	*Austrochloritis porteri*	GQ850994	GQ851256	GQ850925	3	Lamington NP, Qld
9	*Papuexul bidwilli*	GQ851032	GQ851294	GQ850970	3	Port Macquarie, NSW
9	*Posorites conscendens* Bunya	GQ851098	GQ851344	x	2	Bunya Mts NP, Qld
9	*Posorites conscendens* Lamington	GQ851097	x	GQ850973	2	Lamington NP, Qld
9	*Rhynchotrochus macgillivrayi* BT	GQ851036	GQ851298	GQ850976	3	Big Tableland, FNQ
9	*Rhynchotrochus macgillivrayi* Kurrimine	GQ851197	x	x	1	Kurrimine Beach, FNQ
9	Camaenidae BR1	GQ851216	x	x	1	Tamborine Mt, Qld
9	Camaenidae BR7	GQ851134	x	x	1	Mt Mitchell, Cunninghams Gap, SEQ
9	Camaenidae MV12	GQ851219	x	x	1	Mt Seaview Station, NSW
9	Camaenidae MV3	GQ851154	x	x	1	Kippara SF, NSW
9	Camaenidae MV4	GQ851220	x	x	1	Cairncross SF, NSW
9	Camaenidae MV8	GQ851212	GQ851359	x	2	Kempsey, NSW
9	Camaenidae MV9	GQ851221	x	x	1	Nambucca Heads, NSW
9	Camaenidae NE23	GQ851128	x	x	1	Mt Kaputar NP, NSW
9	Camaenidae NE5	GQ851142	x	x	1	Crawney Pass, NSW
9	Camaenidae NN2	GQ851119	x	x	1	Broken Heads, NSW
9	Camaenidae NN3	GQ851129	x	x	1	Girrard SF, NSW
9	Camaenidae SN1	GQ851135	x	x	1	Barrington Tops NP, NSW
9	Camaenidae SN10	GQ851121	x	x	1	Mt Wilson, Blue Mountains NP, NSW
9	Camaenidae SN11	GQ851213	GQ851360	x	2	Yerriyong SF, NSW
9	Camaenidae SN4	GQ851133	x	x	1	Cattai, NSW
9	Camaenidae SN5	GQ850992	GQ851254	GQ850921	3	Mt Dromedary, NSW
9	Camaenidae SN6	GQ851130	x	x	1	Misty Mt, NSW
9	Camaenidae SN6 Kybeyan	GQ851056	GQ851311	x	2	Kybeyan, NSW
9	Camaenidae SN8	GQ851123	x	x	1	Tuglow Caves, NSW
9	Camaenidae SN9	GQ851131	x	x	1	Mt Emperor Blue Mts NSW
9	Camaenidae ST2	GQ851137	x	x	1	Tumut, NSW
9	Camaenidae VC1	GQ851140	x	x	1	Wilsons Promontory, VIC
9	*Crikey steveirwini*	GQ851092	GQ851340	x	2	Mt Spurgeon, FNQ
10	*Noctepuna cerea* BT	GQ851030	GQ851292	GQ850968	3	Helenvale, FNQ
10	*Noctepuna cerea* Thornton	GQ851090	GQ851338	x	2	Roaring Meg Creek, FNQ
10	*Trachiopsis mucosa*	GQ851108	x	GQ850982	2	Dipperu NP, Qld
10	*Trachiopsis strangulata*	GQ851208	x	x	1	Punsand Bay, Cape York
10	*Trozena morata*	GQ851045	GQ851306	GQ850984	3	Forty Mile Scrub, Qld
10	Camaenidae CY3 Coen	GQ851083	GQ851331	x	2	Coen, Cape York
10	Camaenidae CY3 Musgrave	GQ851167	x	x	1	Musgrave, Peninsula Road, Cape York
10	Camaenidae CY5	GQ851206	x	x	1	Wenlock River, Cape York
10	Camaenidae EU3	GQ851207	x	x	1	Copperfield River Gorge, FNQ
11	*Xanthomelon jannellei*	x	GQ851380	x	1	Cape Melville, FNQ
11	*Xanthomelon pachystylum* Dawson	GQ851046	GQ851307	GQ850987	3	Dawson Valley, Qld
11	*Xanthomelon saginatum*	x	GQ851397	x	1	Chillagoe, FNQ
11	*Amplirhagada alta crystalla*	GQ851173	x	x	1	Mitchell Plateau, Kimberleys
11	*Amplirhagada mitchelliana*	GQ851174	x	GQ850918	2	Mitchell Plateau, Kimberleys
11	*Kimboraga micromphala*	GQ851175	x	x	1	Windjana Gorge, WA
11	*Mesodontrachia fitzroyana*	GQ851180	x	x	1	Fitzroy Station, WA
11	*Mouldingia orientalis*	GQ851177	x	x	1	Lissadell Homestead, Kimberleys
11	*Ningbingia octava*	GQ851181	x	GQ850967	2	Ningbing Ranges, Kimberleys
11	*Ordtrachia australis*	GQ851179	x	x	1	Behn River, Kimberleys
11	*Ordtrachia intermedia*	GQ851178	x	x	1	Spring Creek Station, Kimberleys
11	*Basedowena gigantea*	GQ850998	GQ851260	GQ850929	3	Mann Ranges, Central Australia

Table 2. *Continued*

Clade	Taxon	16S	COII	12S	Genes	Location
11	*Basedowena squamulosa*	GQ851147	x	x	1	Palm Valley, Central Australia
11	*Basedowena vulgata*	GQ851146	x	x	1	Tomkinson Range, WA
11	Camaenidae CC1	GQ851064	x	GQ850932	2	Mt Unbunmarro, Qld
11	*Falspleuroxia overlanderensis*	GQ851159	x	x	1	Hamelin Pool, WA
11	*Lacustrelix eyrei*	GQ851166	x	x	1	Frome River, SA
11	*Micromelon nepouieana*	GQ851088	GQ851336	x	2	Yankaninna, SA
11	*Minimelon colmani*	GQ851172	x	x	1	Mt Russell, WA
11	*Montanomelon reynoldsi*	GQ851176	x	x	1	Reynolds Range, NT
11	*Plectorhagada carcharias*	GQ851186	x	x	1	Kennedy Range, WA
11	*Pleuroxia adcockiana*	GQ851094	x	GQ850972	2	Krichauff Ranges, NT
11	*Pleuroxia italowiana*	GQ851188	x	x	1	Italowie Gorge Finders Ra., SA
11	*Pleuroxia oligopleura*	GQ851095	GQ851342	x	2	Eucla, SA
11	*Pleuroxia phillipsiana*	GQ851189	x	x	1	Moro Springs Flinders Ra., SA
11	*Pleuroxia polypleura*	GQ851187	x	x	1	Madura, WA
11	*Promonturconchum superbum*	GQ851190	x	x	1	North West Cape, WA
11	*Quistrachia lefroyi*	GQ851191	x	x	1	Cape Range, WA
11	*Quistrachia monogramma*	GQ851192	x	x	1	Tunnel Gorge, Napier Range, WA
11	*Sinumelon bednalli*	GQ851202	x	x	1	Alice Springs, Central Australia
11	*Sinumelon hullanum*	GQ851201	x	x	1	Lassiters Cave, Petermann Ranges, NT
11	*Sinumelon nullarboricum*	GQ851200	x	x	1	Kuthala Pass, WA
11	*Sinumelon serlense*	GQ851102	GQ851347	x	2	Flinders Range, SA
11	*Strepsitaurus ningaloo*	GQ851203	x	x	1	Ningaloo, WA
11	*Tatemelon herberti*	GQ851204	x	x	1	Musgrave Range, SA
	Chloritisanax banneri	GQ851003	GQ851266	GQ850935	3	McIlwraith Ranges, Cape York
Melanesia_b	*Chloritis quercina*	GQ851050	x	GQ850914	2	Wagina, Solomon Islands
Melanesia_b	*Megalacron boyerii*	GQ851149	x	x	1	Woodlark Island
Melanesia_b	*Megalacron* sp.	GQ851053	x	GQ850917	2	Wagina, Solomon Islands
Melanesia_b	*Rhynchotrochus* sp.	GQ851152	x	x	1	Tagula Island
Melanesia_b	*Rhynchotrochus woodlarkianus*	GQ851151	x	x	1	Woodlark Island
Melanesia_b	*Rhynchotrochus woodlarkianus* Sinamata	GQ851150	x	x	1	Woodlark Island, Sinamata
Melanesia_c	*Papuina cingulata*	GQ851185	x	x	1	Opio, Papua New Guinea
Melanesia_c	*Papuina mendana*	GQ851052	x	GQ850916	2	Wagina, Solomon Islands
Melanesia_c	*Papuina* sp.	GQ851051	x	GQ850915	2	Wagina, Solomon Islands
Melanesia_c	*Papustyla xanthochila*	GQ851115	x	x	1	Wagina, Solomon Islands
	Kendrickia ignivenatus	GQ851165	x	x	1	Red Bull Bore, Napier Range, WA
	Rhagada capensis	GQ851195	x	x	1	North West Cape, WA
	Rhagada dringi	GQ851035	GQ851297	GQ850975	3	Port Headland, WA
	Rhagada torulus	GQ851196	x	x	1	Carnarvon, WA
Outgroup	*Bradybaena similaris* Brisbane	GQ851001	GQ851264	GQ850933	3	Brisbane, Qld
Outgroup	*Bradybaena similaris* Borneo	GQ851164	x	x	1	Borneo, Mt Kinabalu
Outgroup	*Satsuma jacobii*	GQ851015	GQ851277	GQ850952	3	Nakayama, Kasuga Village, Gifu Pref.

Taxonomy is based on the Zoological Catalogue of Australia and the Australian Fauna Directory: (AFD, 2007). Names are in two or three parts: Genus_species_location/subspecies/AFD taxa to be referred to. Informal museum codes referred to as Camaenidae_'code'. Columns for 16S, COII, 12S refer to GenBank accessions; ×, no data. Lineage refers to groups labelled in Figure 7. Locality codes: WA, Western Australia, Qld, Queensland, NSW, New South Wales, MEQ, Mid East Queensland, FNQ, far north Queensland, NT, Northern Territory, SEQ, South East Queensland, SA, South Australia, VIC, Victoria, NP, National Park, NR, Nature Reserve; SF, State Forest, PNG, Papua New Guinea.

in the matrix here, a proportion of pairs of taxa have no data in common: for the 327-taxa data matrix there are 53 301 pairwise comparisons, of which 7980 (15%) have no data in common. However this is not as big a problem as it may at first seem, as these are between taxa separated by many nodes with numerous intermediate nodes sharing data. Based on the final topology (see Fig. 7), there is one case of taxa with no common data being separated by three nodes, and four cases separated by four nodes (involving two *Thersites novaehollandiae* and a *Jacksonena delicata*, and *Xanthomelon jannellei* with an *Amplirhagada*). Less than 0.2% of pairwise comparisons are separated by fewer than ten nodes. The weakest part of the matrix concerns some of the distant extralimital lineages; however, the limited resolution is adequate to

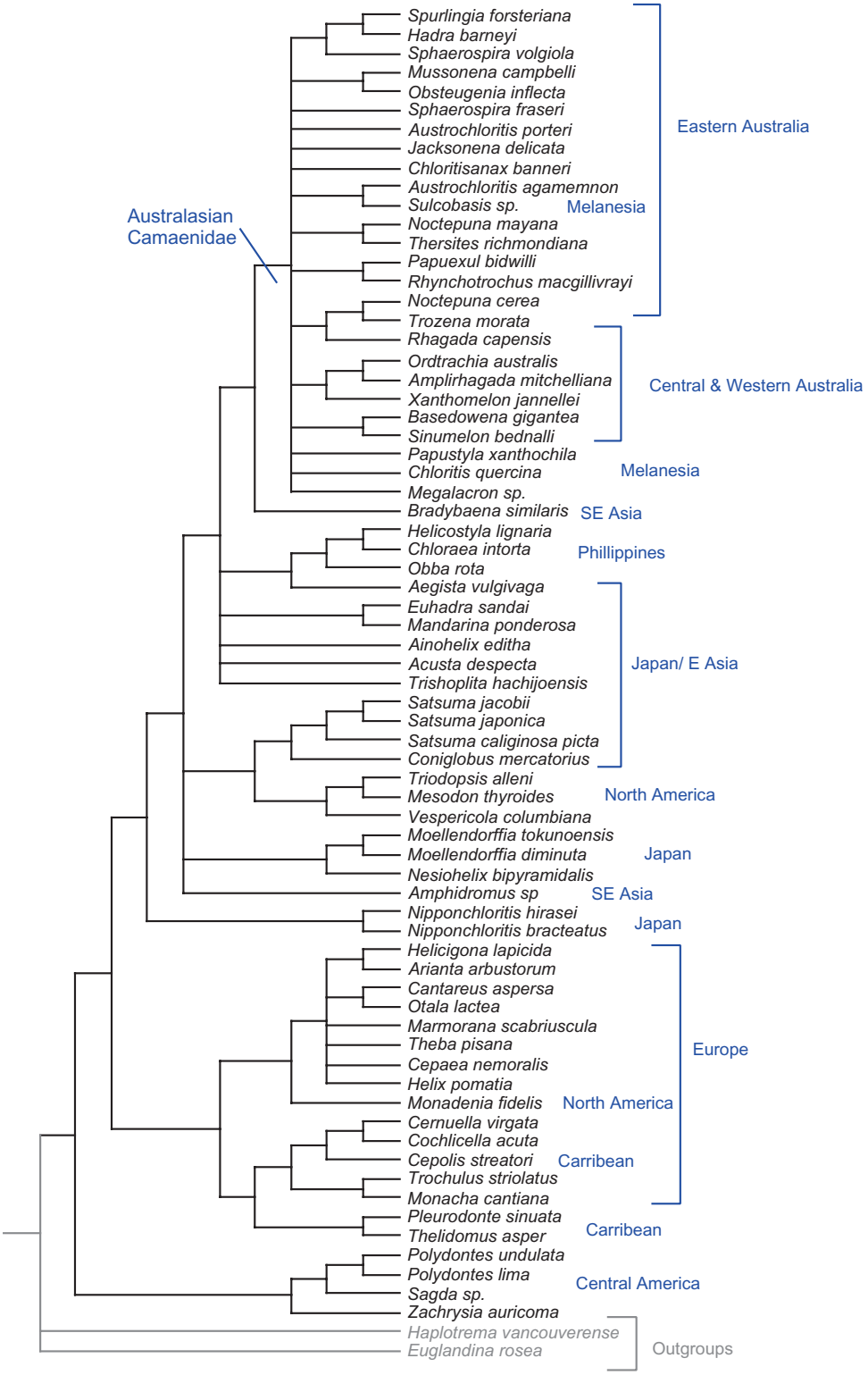


Figure 2. 28s rDNA phylogeny: strict consensus of 787 439 unique maximum parsimony trees from 20 random addition heuristic tree bisection-reconnection searches, each collecting 50 000 trees. Basal polytomy rooted with *Haplotrema vancouverense*. General geographical localities indicated right; further details in Table 1.

Table 3. Gene contributions to the mtDNA supermatrix data trees

	327 taxa tree			147 taxa tree		
Contributing data	Branches	%branches	%divergence	Branches	%branches	%divergence
One gene only						
16S	168	25.8	30.8			
COII	65	10.0	3.7			
12S	0	0.0	0.0			
Two genes only						
16S + COII	121	18.6	12.5	71	24.4	18.7
16S + 12S	41	6.3	15.5	27	9.3	23.3
COII + 12S	1	0.2	0.1	1	0.3	0.1
All three genes	255	39.2	34.5	192	66.0	57.9

%branches: percentage of the number of branches.
%divergence: percentage of sum total branch length.

distinguish these from the east Australian groups (see Discussion). The characteristics of the nested matrix are outlined in Table 3 giving a breakdown of the number of branches (and proportion of overall tree branch length) based upon one, two, and three genes. As all data are from one locus (mtDNA) the gene tree/species tree issue (Maddison, 1997) is not explored here, and we present the estimated mtDNA phylogeny as a good proxy of the organismal tree.

MtDNA alignment and sequence characteristics

The aligned mtDNA sequence data matrix for this analysis comprises 327 tips containing: 176 COII sequences of 480 sites, 274 16S rDNA of 450 sites, 86 12S rDNA of 522 sites, for a total of 1452 sites. Sixty-two taxa have all three genes, 85 taxa have two genes, 180 taxa have only one gene: a quarter of a million defined characters in a matrix of a half a million, the remaining being missing data. Total matrix of 474 804: 235 208 ACGT states, 218 IUPAC ambiguity states, 222 132 missing states, and 17 246 gap states. Sequences were aligned with ClustalX (Thompson *et al*., 1997) using standard parameters, visually inspected for major anomalies, and where available, compared to 16S and 12S secondary structure models.

All sites were used, including gap regions because: (1) amongst closely related sequences there is much less ambiguity; therefore these regions contain valuable information; (2) including a site rate heterogeneity parameter can effectively weigh the degree of uncertainty in homology by the level of overall divergence. Altogether, this gives better resolution near the tips and branch lengths across the depth of the tree. Nevertheless, we tested the effect of removing complex alignment sections in the 62 taxa all-gene data matrix (a total of 7.5% of sites excluded).

Three levels of taxon sampling

The full 327-tip supermatrix contains three levels of data completeness corresponding to datasets where (1) all taxa had all genes, (2) at least two genes, and (3) any gene. Analyses were performed on all three: the 62t analysis had 62 taxa with all three genes and no missing data; the 147t analysis had 147 taxa with at least two genes (62 all three genes and 85 with two of the three genes); and the 327t analysis had 327 taxa (62 with all three genes, 85 with two of the three genes, and 180 with only one gene). Table 2 lists the genes available for each taxon.

Model based estimation of tree support space

Phylogenetic inference using stochastic models evaluated by likelihood where missing states are handled by integrating over all possible states is ideally suited to this nested design of molecular sequence data, for both topology and for branch length estimation (Swofford, 2000; Ronquist & Huelsenbeck, 2003; Driskell *et al*., 2004; Felsenstein, 2004; de Queiroz & Gatesy, 2007; Wiens & Moen, 2008). Dense taxon sampling including many sequences of a wide range of differences can improve within group alignment, characterization of site by site sequence evolution, and phylogenetic accuracy (Sullivan, Swofford & Naylor, 1999; Zwickl & Hillis, 2002; Blouin, Butt & Roger, 2005; Hedke *et al*., 2006). Large numbers of taxa bring an exponential increase in computational burden making the analysis of such a data matrix difficult until development of Bayesian Markov Chain Monte Carlo (MCMC) and fast ML methods, whereas missing data make measures of support such as bootstrapping and Bremer support problematical (Lee & Hugall, 2003; Wiens, 2003). Bayesian estimation of tree space support may be the best practical method for such a large nested data supermatrix, providing a

way of rationalizing the amount of sequence needed to be gathered for tree shape (both topology and branch length) and measures of support. Therefore our analysis was based on MCMC using MrBayes v. 3.1.2 (Ronquist & Huelsenbeck, 2003), backed up with ML using RAxML 7.0.4 (Stamatakis, 2006) with fast bootstrapping.

Preliminary MCMC runs were undertaken to determine optimal MCMC run conditions (chain mixing, burnin, parameter effective sample size; posterior probability stability, e.g. Wilgenbusch, Warren & Swofford, 2004; Beiko *et al*., 2006; Drummond & Rambaut, 2006). Standard (default) MrBayes settings were adequate for the 62t analyses. Additional chains (above the standard four) made little difference to the 147t analysis but lifted mixing in the 327t runs to levels considered suitable. For the final 327t analyses, an approximate starting tree (derived by parsimony) was used to speed up convergence and therefore more efficiently use runtime.

Final run conditions used were:

62 taxa dataset: 5 million steps sampling every 50 steps with a burnin of 20%, four chains with heating temperature = 0.2 (i.e. standard), random starting tree.

147 taxa dataset: 10 million steps sampling every 100 steps, 20% burnin, six chains with heating temperature = 0.15, random starting tree.

327 taxa dataset: 20 million steps sampling every 100 steps, 20% burnin, eight chains with heating temperature = 0.075, approximate MP starting tree.

Analyses were conducted on an e-linux parallel cluster running one chain per processor: the 327t 20 million × 8 chain runs took 250–300 h each. Each analysis was carried out twice and then combined for the final Bayesian estimates of posterior probability (PP).

Sequence evolution model design

Model choice and partition strategies were evaluated with the second-order Akaike information criteria (AICc: Burnham & Anderson, 2003; Lee & Hugall, 2006). For assessing partition strategies AICc used the MCMC equilibrium average lnL (Hugall *et al*., 2008). To get the best estimate of phylogeny with branch lengths, the combined total data were used with branch length estimates linked. However, each gene may have a different overall rate. Despite missing data, for each branch the union of genes can contribute to branch length estimation, and by linking branch lengths all sites can contribute information, with a relative rate parameter (m) for each gene. Given the hierarchical dataset and hence blocks of missing data, partition by gene was considered desirable as the rate parameter may account for different rates amongst genes and hence some correction amongst branches estimated from different genes (Ren, Tanaka & Yang, 2008; Lemmon *et al*., 2009). A further two strategies were compared. As a result of limited secondary structure information, rDNA was split into length variable ('loops') vs. length-stable regions ('stems'). Therefore the three models investigated were: (1) single partition model; (2) three partition model (3p: COII, 12S, 16S); (3) four partition model (4p: COII 1^{st} + 2nd, 3rd, rDNA 'loops', rDNA 'stems'). AICc indicated that the most general model (general time-reversible with gamma rate heterogeneity and proportion of invariant sites parameters: GTR-G-inv) was appropriate for each partition type, and that the three and four partitions, but not more complex arrangements, were considered improvements over the single partition strategy.

ASSESSING TREE STABILITY AND ROBUSTNESS

The different models and taxa sets may give different results – a sensitivity we wished to address. By comparing the three partition models for each of the three taxa sets, within and amongst taxa set variation can be addressed. Across partition models we can (1) accept only one best model; (2) weight the model results by the relative information content of the models (i.e. AIC weight), or (3) accept a set of suitable models. For the latter, as all the models are misspecified to some extent, they would then be weighted equally (Burnham & Anderson, 2003). Therefore our final posterior estimates used for phylogenetic inference combined the MCMC output of the three different model types (single, 3p and 4p; most simply carried out by combining the.trprobs files from each run). This is similar to the idea of congruence amongst analysis types, and is intended to take a conservative approach to the estimate of PP support because of model misspecification (Erixon *et al*., 2003; Huelsenbeck & Rannala, 2004). The final three models combined ≥ 0.95 PP consensus trees were no more resolved than any one model tree, verifying this contention. Throughout we use this ≥ 0.95 PP as our basis for taxonomic interpretation.

Nonparametric bootstrap support was assessed by fast bootstrapping using RAxML 7.0.4 (Stamatakis, 2006) applying the 3p partition strategy. These analyses used default settings with 5000 bootstrap replicates via the –f a function, which conducts fast bootstrapping using the GTR-CAT model then uses a subset of these bootstrap trees as inputs for a series of more thorough ML searches using the GTR-G-inv model.

Assessing effect of nested data

To assess the effect of taxon sampling and the nested missing data, analyses were performed on the three

taxa sets. To allow comparison, for both topology and branch length, the larger taxa set results were pruned down to match the smaller taxa sets. Thus all three taxa set analyses could be compared for the common 62 taxa, and by pruning down to the common 147 taxa the 327 taxa set results could be compared with the 147 taxa set results. For the pruned results, taxa were first pruned from the MCMC samples (retaining branch lengths) and then PP and consensus trees calculated, including branch lengths. By pruning the original MCMC samples both topology and branch length PP could be calculated. Supporting Information Figures S2 and S4 show the branch length and PP estimation from pruned MCMC samples.

Topology was compared several ways: (1) by topological distances between trees (number of nodes collapsed in a strict consensus); (2) by SH tests (Shimodaira & Hasegawa, 1999); (3) by ≥ 0.95 PP consensus. The similarity amongst MCMC analyses including both topology and branch length can be compared by the overlap in log likelihood scores of a sample of trees drawn from the MCMC chains. Two thousand samples were drawn from each MCMC (at equilibrium), pruned to the largest common taxa set, and the lnL calculated, using the MCMC sample branch lengths. To provide a common currency for the SH tests and the MCMC overlap, likelihood scores were calculated in PAUP v.4.10b using the optimal single partition GTR-G-inv model. SH tests were also conducted in RAxML v7.0.4 using the 3p and 4p partition strategies (applying the GTR-G-inv model to each partition). Branch lengths were compared across all common bipartitions in the MCMC analyses and differences amongst models and taxa sets compared to the variance seen in the MCMC analysis of the most data complete analysis (i.e. the one with the least variance). This used the sample of MCMC trees (as above) but can also be carried out from the branch lengths and variance values in the MrBayes sumt.parts summary file.

Both the 28SrDNA and mtDNA alignments are available via TreeBase.

RESULTS

28S rDNA phylogeny

We start by describing the results of the 28S rDNA analysis first, to deal with the larger scale origins of the Australasian taxa. There are a very large number of equally parsimonious trees: multiple separate runs from random starts accumulated 787 439 equally parsimonious trees yielding a strict consensus shown in Figure 2. The bulk of the many MP trees is a result of the very low diversity amongst Australasian taxa, the remainder of the tree being reasonably well resolved. For the Australasian 'ingroup' subtrees, 80% were nonbinary with up to five collapsed nodes (zero length branches). The MP consensus tree (and the ML and Bayesian versions in Supporting Information Fig. S1) can be compared with those in Wade *et al*. (2001, 2006, 2007): considering the low support values, there are no significant relevant differences amongst trees. Bootstrap analysis with these data is not practical but an indication of support can be gained from the Bayesian analysis (Supporting Information Fig. S1B): PP of the Australasian Camaenidae is 0.71 or 0.88 excluding *Helicosyla* and *Chloraea*. Notwithstanding weak power, this 28S analysis suggests that Australasian camaenids are a subset of the Asian camaenids (*s.l.*), which are in turn a subset of the global Helicoidea supergroup (Wade *et al*., 2001, 2006, 2007). Thus the Australasian lineages appear to represent a local part of a global biogeographical series.

The Australian Camaenidae are most likely relatively young. Even the small sample of Asian taxa has greater phylogenetic diversity (divergence depth) than the entire Australasian set of taxa. Maximum divergence amongst the Australasian taxa is only 2.1% (pairwise observed difference) but 3.7% amongst the east Asian taxa (this translates to 2.3 vs. 4.4% using optimized GTR-G model corrected divergences). This is further emphasized by considering that a diverse array of taxa, spread across the entire Australian east coast range north to south (10–40°S) are identical for this amount of 28S (*Sphaerospira fraseri*, *Ponderconcha gilberti*, *Chloritobadistes victoriae*, *Gnarosophia bellendenkerensis*). The 28S data have little information within the Australasian lineages. The relatively young age and rapid lineage diversification rate of Australasian groups with respect to the substitution rate for 28S suggest that the Australasian clades should be tractable to phylogenetic analysis with the relatively fast mtDNA data available.

The hierarchically nested mtDNA data supermatrix phylogeny

The ultimate goal of this analysis was to produce a single phylogeny with branch lengths containing all 327 taxa (tips) from a supermatrix that contains much missing data. To assess the reliability of this approach the MCMC analyses for the three taxa set levels of data were compared to each other. The comparisons focused on the similarities and differences in MCMC posterior consensus topology, node posterior probability, and branch length. The principle sequence evolution model used was the three

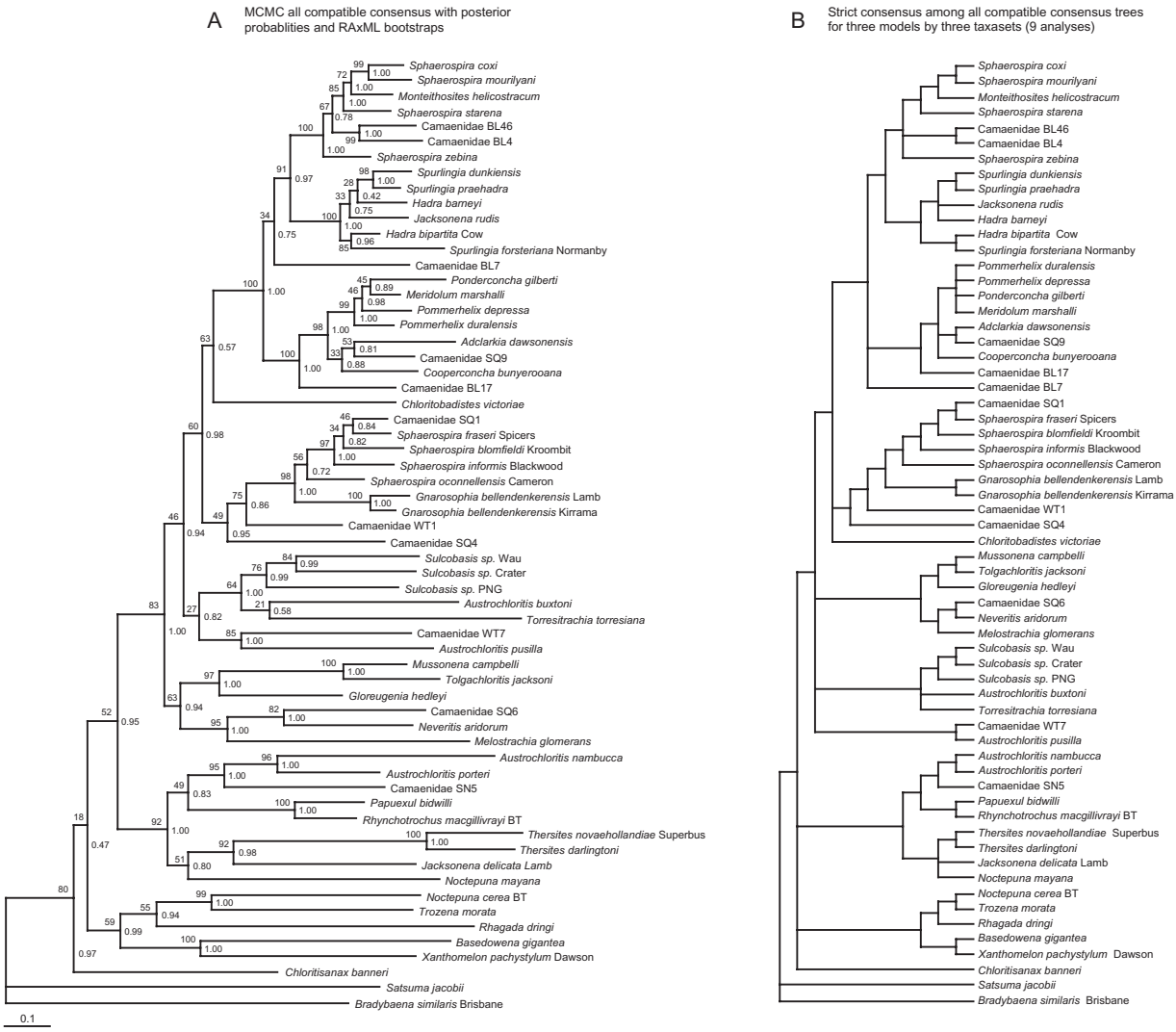


Figure 3. Consensus trees of the 62 taxa that have all three genes, using the 3p model. A, 62t Markov chain Monte Carlo (MCMC) all-compatible consensus tree with branch lengths, posterior probabilities (PP; decimal format to the right of nodes) and RAxML bootstrap values (5000 replicates; percentages to left of nodes); B, strict consensus amongst all-compatible consensus trees for three models by three taxa sets (nine analyses).

gene partition strategy (3p), with all sites included; however the effects of the complex alignment regions and of different partition model strategies was also considered.

Our comparison of analyses has six aspects:

1. The all-compatible consensus trees with branch lengths, pruned to the largest common set of taxa.
2. Strict consensus of all-compatible posterior consensus trees amongst different models and taxasets, pruned to the largest common set of taxa.
3. The ≥ 0.95 PP consensus across models, pruned to the largest common set of taxa.
4. SH tests.
5. Overlap in MCMC sample tree likelihood distribution.
6. Variation in branch lengths amongst model and taxa set analyses, compared to within MCMC variation.

For the sake of brevity, only the general findings of (3), (4), and (5) are described here, with full details retained in the Supporting Information.

Figures 3A, 4A, and 7, respectively, show the MCMC posterior all-compatible consensus for the 62t taxa set with all genes, the result for the 147t taxa set with at least two genes, and final 327t supermatrix tree, all using the three

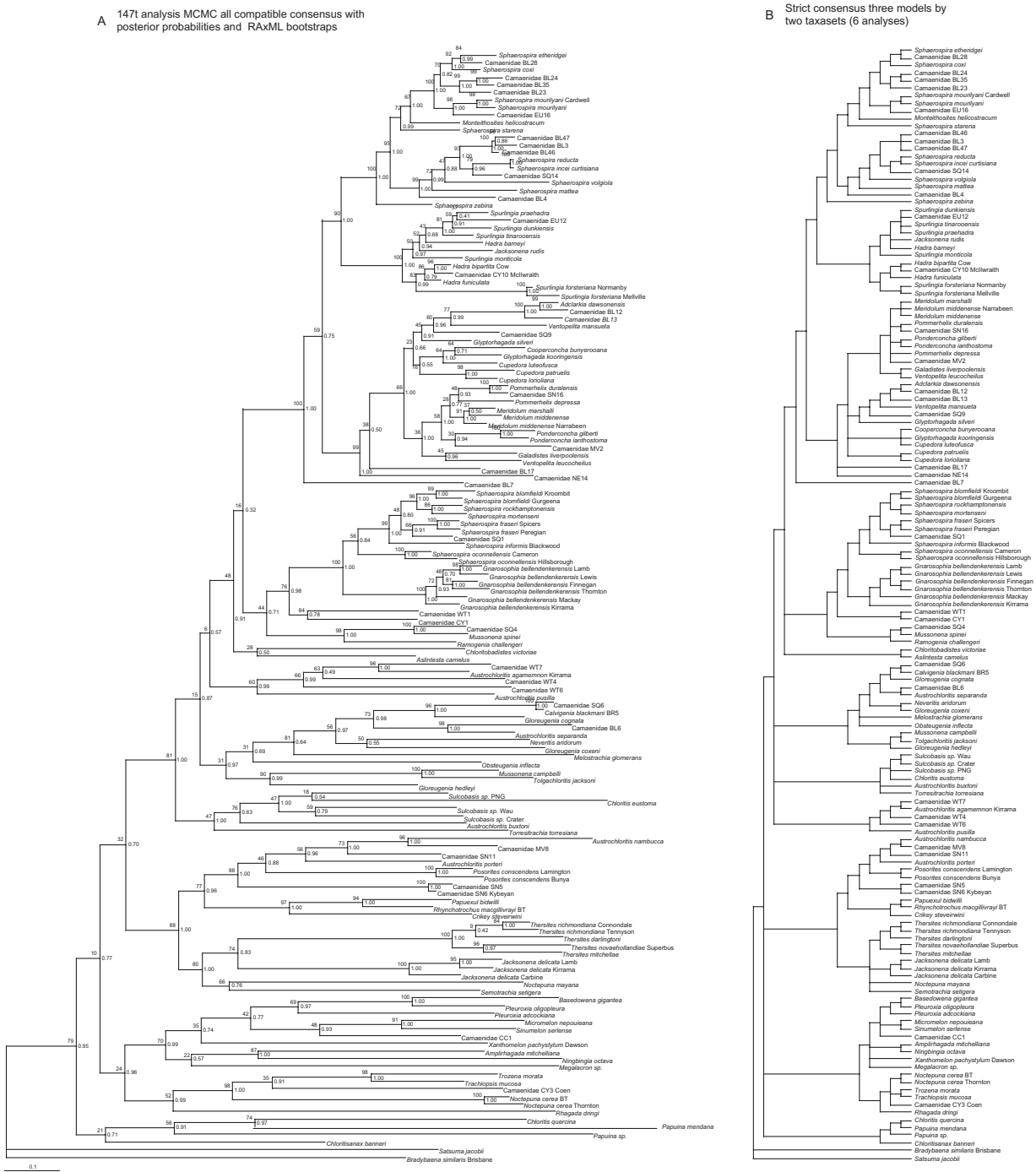


Figure 4. Consensus trees of the 147 taxa that have at least two genes, using the 3p model. A,147t Markov chain Monte Carlo (MCMC) all-compatible consensus with branch lengths, posterior probabilities (PP; decimal format to the right of nodes) and RAxML bootstrap values (5000 replicates; percentages to left of nodes); B, strict consensus of three models by two taxa sets (six analyses).

partition model. Matching consensus trees of analyses pruned down to the common 62 and 147 taxa are shown in Supporting Information Figures S2 and S3. These illustrate the overall similarity in shape (topology, PP support, and branch lengths) across the nested datasets. This is now looked at in more detail.

Topology stability

Within the 62t analysis, differences in all-compatible consensus trees because of model type (the three partition strategies) and trimming the complex alignment regions (7.5% of sites excluded) are minor subsets of the differences amongst taxa sets. Only three nodes are collapsed, and the trimmed data result differs by only one of these nodes. Amongst the 62t, 147t, and 327t MCMC analyses, all using the same three partition model, ten nodes are collapsed, with the 147t and 327t trees more similar to one another than they are to the 62t tree. The 62t tree differs by seven and nine nodes (out of a total 60) from the 147t and 327t trees, respectively, which differ from each other by only four nodes. Figure 3B shows the strict consensus amongst all taxa sets and partition models: the three taxa sets each analysed using three different models (nine MCMC all-compatible posterior consensus trees). A total of 13 nodes are collapsed. For the 147 taxa results, amongst models within the 147t analysis ten (out of a total 145) nodes are collapsed; for the same (3p) model between the 147t and 327t taxa set trees 16 nodes differ, and across all models and taxa sets (six trees), 33 nodes are collapsed (Fig. 4B). The trimmed data effectively recovers the same tree (and PP support) as using all sites. As the complex alignment did not cause discrepancies at this highest divergence level, all sites were used for all further analyses. In summary, the principle difference was amongst taxa sets: the difference because of model was largely a subset of the differences amongst taxa sets.

Comparing the all-compatible posterior consensus trees themselves highlights nodes that differ amongst analyses irrespective of posterior support. The next approach compared the ≥ 0.95 posterior probability consensus trees, indicating which nodes have consistently high support. This combined the MCMC results (the.trprobs files) of the three different models (the three different partition strategies). Therefore support needed to be high in all three for the combined PP to be ≥ 0.95 (at worst ≥ 0.85 in one analysis if the other two are 1.00). Results are shown in full in Supporting Information Figures S4 and S5. Amongst analyses, for matching taxa sets pruned from larger taxa sets, there was only one case of contradictory ≥ 0.95 PP but several nodes that were ≥ 0.95 in one analysis were < 0.95 in others. For the 62 taxa comparisons there was one conflict between the 62t and 147t analyses – the position of *Meridolum duralensis*. For the 147 taxa comparisons, there were no contradictory PP but there was fluctuating support for a key node of interest (see Discussion below): including *Aslintesta camelus* this node had 0.89 PP but ignoring it, the support rose to 0.95 PP.

In summary, instability in the tree was localized to the same regions across models and taxa sets, and these had consistently low PP support, whereas robust regions had consistently high support across models and taxa sets. Essentially the same nodes lost amongst taxa sets (Figs 2B, 3B) are the nodes collapsed in the ≥ 0.95 PP consensuses (Supporting Information Figs S4, S5). This was matched by the SH tests in which none of the trees were significantly worse than each other, whatever the partition model used to assess the likelihood differences (Supporting Information Table S3).

The next comparison was not of topology but of a general similarity amongst MCMC sample trees in both topology and branch length. This measured the degree of overlap in the distribution of a summary likelihood score for these sample trees, using a single partition GTR-G-inv model (shown in Supporting Information Fig. S7). The degree of overlap is a general indication of the similarity both in topology and branch length amongst the analyses, with overlap ranging from 0 to 1 and may be related to the 95% confidence intervals (95% CI) although there is no certain statistical interpretation of this (cf. Kulback Leibler divergence; Burnham & Anderson, 2003). There was little difference amongst partition models (minimum overlap = 0.81 for the 62t models; 0.65 for the 147t models). There was a 0.15 overlap between the 62t and 147t runs, only 0.03 overlap between the 62t and 327t runs, with 0.50 overlap between the 147t and 327t runs, and 0.19 overlap between the 147t and 327t runs. Essentially this mirrors the previous comparisons: minor differences amongst partition models; some differences amongst taxasets, with the 147t and 327t analyses more similar to each other than they are to the 62t results.

Branch length estimation and stability

Finally, we specifically compared branch lengths among models and taxa sets for all internodes seen in the all-compatible posterior consensus trees, in relation to the variation seen within the MCMC analysis of the most data complete analysis. Results are shown in Figures 5 and 6 (and summarized in Supporting Information Table S4). The linked partition relative rate parameter (m) for each gene for each of the 62t, 147t, and 327t analyses was very similar, giving 16S : COII : 12S m = 0.77:1.30:0.92. As expected, COII was the 'fastest' but counter to expectations 12S turned out to be slightly 'faster' than 16S.

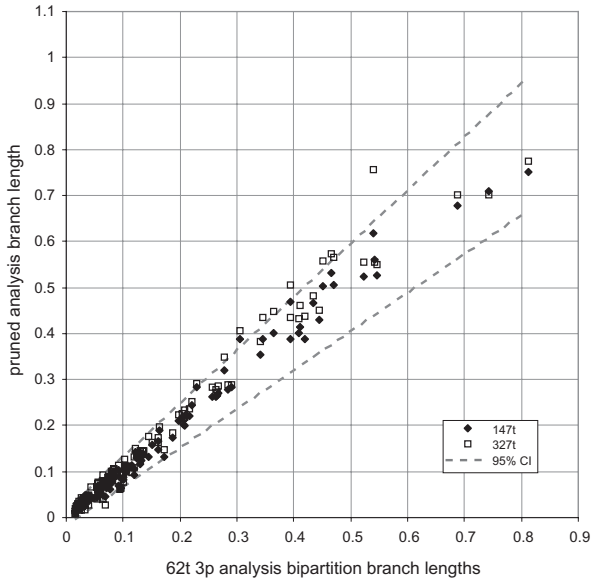


Figure 5. Branch length comparisons across models and taxa sets for 62 taxa. All common bipartition branch lengths from each of the three taxa set analyses (pruned to the 62 taxa) plotted against the 62t analysis 3p model bipartition branch lengths. Confidence intervals (CI) based on MCMC variation in the most data complete analysis, are approximated by dashed lines.

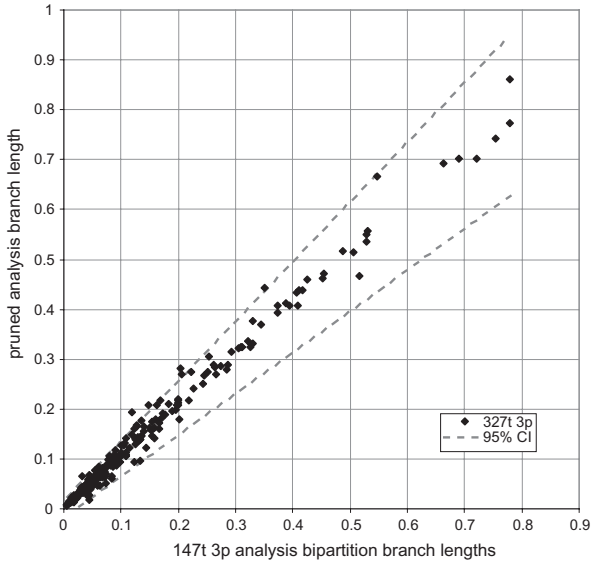


Figure 6. Branch length comparisons for 147 taxa. All common bipartition branch lengths plotted against the 147t 3p analysis bipartition branch lengths. Confidence intervals (CI) as for Figure 9.

For the 62t analyses, the relative branch length estimates amongst the three models (the three partition strategies) were very similar (correlation coefficient all ≥ 0.999, with no branches falling outside the 0.95 PP); less than the variation amongst taxa sets. They were also very similar amongst models for the 147t analyses (correlation coefficient all ≥ 0.997, with no branches falling outside the 0.95 PP), but less so in the 327t analyses (correlation coefficient all ≥ 0.996, with six branches falling outside the 0.95 PP). Figure 5 shows the relationship amongst the branch lengths in the 62 taxa trees across taxa sets using the same 3p model. The 95% CI are from the MCMC variance seen in the 3p model analysis. Figure 6 shows the equivalent for 147 taxa. Overall the fit was lower than within models but few branches fell outside the 95% CI, with the larger taxa sets tending to have longer branches. This can be summarized by comparing the tree length parameter (TL = sum of all branches) distributions amongst analyses (Supporting Information Fig. S8). For 62 taxa (using the same 3p model) the 62t and 147t analyses essentially returned the same TL, whereas the 327t analysis had a generally longer tree (median 20.25:20.44:21.61) but still just within the 95% CI (and distribution overlap = 0.23). For 147 taxa, the 327t median fell outside the 147t 95% CI (median 35.63:37.92; distribution overlap 0.15).

Larger taxa sets returning longer subtrees (for a matching set of taxa) might be expected because increased taxon sampling increases divergence estimates by revealing more sequence changes (Hugall & Lee, 2007). Substitution model parameter estimates contribute little to this effect as applying each set of (posterior median) parameter values to a common topology gives essentially the same TL. However, as the 327t TL differs more noticeably from the other two, this may also be because of limitations of the gene-by-taxa sampling in the 327t data matrix; for example, instability in the topological placement of *Xanthomelon jannellei* in the 327t analysis causing long branches in the associated lineages.

Summarizing the overall results of comparisons of the nested supermatrix dataset MCMC analyses, there is general agreement amongst taxa sets in topology, with conflict amongst taxa sets restricted to regions with low PP. There is also a general similarity in relative branch length. The additional taxa with subsets of data insert into the phylogeny with little perturbation of the mainframe work, in keeping with their nested position in the phylogeny. The regions of low support are constrained by nearby well-supported and stable nodes and therefore cause only localized uncertainty. There are some exceptions and these are mostly because of divergent taxa for which it was not possible to get the extra sequence data that the hier-

archical sampling strategy suggested were necessary in the first place (e.g. long branch taxa for which there are no close relatives and therefore no way of breaking up long branches). However, these caused only minor degradation of information because they are sufficiently constrained by nearby nodes. There are some statistically measurable differences amongst taxa sets, in particular between the 327t and 62t sets in MCMC overlap (Supporting Information Fig. S7) and absolute branch lengths (Fig. 5), but in terms of the overall picture and interpretation of the results for discussion, there is no significant discrepancy.

Similar comparisons can be made between the MCMC and RAxML v.7.0.4 trees. RAxML trees for the 62t and 147t taxa sets are shown in Supporting Information Figures S2D and S3C. These are very similar to the MCMC trees, and show reasonable correlation between PP and BS with no nodes in the 62t and 147t analyses incompatible between the PP $\geq$ 0.95 and BS $\geq$ 70% consensuses. For the 147t taxa set bootstrap support (BS) for the 'hadroid group' is also sensitive to *Aslintesta camelus*: 48% included, increasing to 59% when ignored in calculating BS. The pattern of branch lengths is much the same but with RAxML returning overall longer (i.e. higher divergence) branch lengths: 28.5%, 24.5% longer for the 62t and 147t analyses, respectively, whereas the 327t analysis returned much the same branch lengths (3% longer).

The big tree

We are now in a position to describe the final summary single supermatrix phylogeny of all taxa, with PP support, which will be used for evolutionary interpretation in the discussion (along with the 62 and 147 taxa trees). This tree is shown in Figure 7. It is one of the 3p model MCMC all-compatible consensus trees; however, significant support is determined as $\geq$ 0.95 PP after combining the MCMC outputs of the two runs each of the three different model types (single, 3p and 4p). Support is indicated by vertical bars on internodes $\geq$ 0.95 PP in the combined 327t analyses. In addition, nodes with RAxML fast bootstrap values (BS) $\geq$ 70% are indicated by circles. There is reasonable correspondence between PP and BS: 158 nodes have both PP $\geq$ 0.95 and BS $\geq$ 70%, with 22 nodes having only PP $\geq$ 0.95 and 16 with only BS $\geq$ 70%. Lineages inferred from all three genes are indicated by thick blue lines; those inferred from two genes by medium red lines, and those based on one gene by thin black lines. To aid discussion some major clades are labelled. Additional taxonomic and biogeography information is arranged to the right. Some pictures of animals and shells are added for illustration.

This final big tree is consistent with the various preliminary analyses of subgroups, and with the 62t and 147t analyses (notwithstanding the above-mentioned $\geq$ 0.95 PP inconsistencies). Amongst the three models within the 327t analysis, there is no significant ($\geq$ 0.95 PP) topological conflict, and relative branch lengths are similar (Supporting Information Table S4).

First some anomalies need to be noted:

Three taxa in particular showed excessive topological instability as a result of long branches and limited sequence data in the 327 taxa set: *Vidumelon wattii*, *Chloritis minnigerodei*, *Xanthomelon jannellei* (see Fig. 7 legend). These were included in the MCMC analysis but ignored for calculation of PP (i.e. they were pruned out before calculating the summary PP). However, these three 'rogue' taxa caused only local instability: *Vidumelon wattii* flips between where shown and within clade 7; vice versa for *Chloritis minnigerodei*; *Xanthomelon jannellei* is drawn toward *Mesodontrachia*. This causes some distortion of posterior mean branch lengths in this part of the tree because of poor data overlap to the rest of the lineages it associates with (COII vs. 16S and 12S). These taxa are more stable in the RAxML analysis and BS support was calculated using all taxa. However two other long branch taxa were less stable (*Obstengenia inflecta* and *Torresitrachia torresiana*), compromising BS support in neighbouring regions.

The large relative branch length difference between *Hadra bipartita* and *Spurlingia forsteriana* is principally because of COII in *S. forsteriana*. The long path length for the deep Melanesian papuinids (e.g. *Papuina medana*) might be exaggerated as a result of these taxa being mostly represented only by the slower genes 16S and 12S. This may also be the case for the *Mesodontrachia* to *Ordtrachia* group.

The grade of *Simumelon* taxa (*Falspleuroxia* to *Quistrachia*) near the base of clade 11, and also the terminalization (pectinate shape) of the *Austrochloritis* group (*Austrochloritis ascensa* to Camaenidae MV3) may also be a result of them being limited to only 16S data in this complex model MCMC: simpler models (e.g. single partition HKY-G) and methods (e.g. NJ) have these as more symmetrical clades, but in the same relative position to the remainder of the tree. Similarly, nesting of *Kendrickia* within *Rhagada* (with PP < 0.95) rather than as sister, may be an artefact. However, caveat one specimen, clearly it is related to *Rhagada*.

A key area at the base of the 'hadroid group', separating clades 1–4 from 5 to 7, shows some instability: it is significant ($\geq$ 0.95 PP) in the 62t analyses, marginal in the 147t analyses (largely because of *Aslintesa camelus*), and significant in the 327t analyses (after ignoring the above three 'rogue taxa'). However, in all cases the overall tree shape is the same, only differing in the placement of the

nearby lineages (*Aslintesa camelus*, *Chloritobadistes victoriae*, and the *Offachloritis – Mussonena spinei* lineages).

There is no resolution amongst the deep groups (e.g. clades 8 + 9, 10, 11, both groups of Melanesia papuinid, and *Chloritisanax banneri*). However this has little effect on interpretation as they are all distinct from one another and outside all the other lineages, and resolution in this part of the tree is of little meaning at this stage as this is likely to be a subsample of near extralimital diversity (Melanesia and South-East Asia).

Despite relatively complete data there is limited resolution amongst major lineages in clades 2 and 3, and the lineages near the base of the 'hadroid group'. By contrast, with similar data, clades 1 and 4 are reasonably well resolved, whereas a pastiche of data is sufficient to resolve the major lineages in clades 8 and 9. Therefore, irrespective of the numbers of nodes with < 0.95 PP, the shapes of [the sampled lineages in] these clades are probably all similarly well defined.

Lastly, although the 28S rDNA tree (Fig. 2) has low resolution, it is worth noting that it is not inconsistent with the mtDNA trees (e.g. grouping of clades 1 + 2 + 3; the polyphyly of *Noctepuna mayana* and *Noctepuna cerea*; some support for grouping of Simumeloninae taxa *sensu* Solem, 1997).

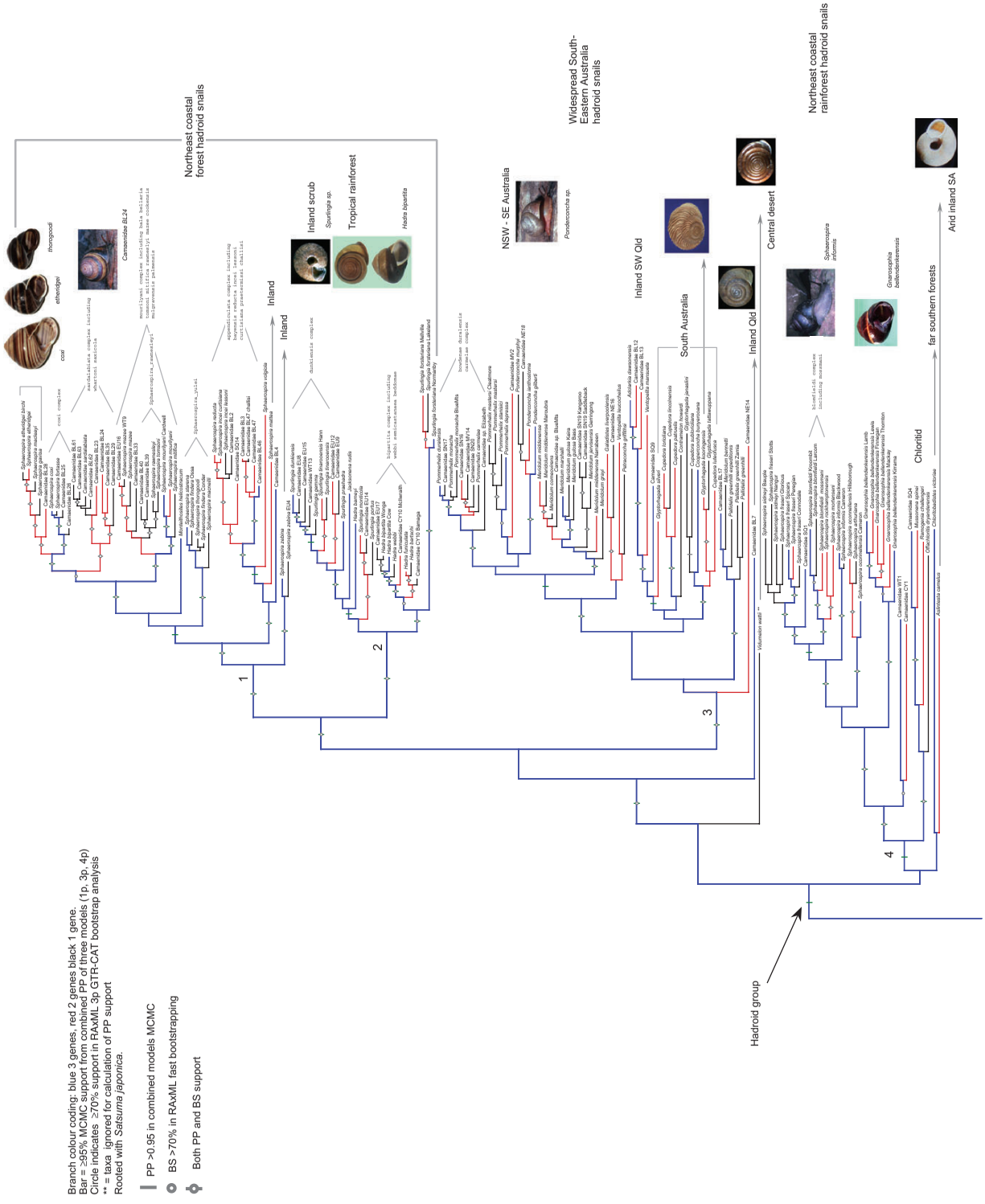
DISCUSSION

Performance of the nested supermatrix approach

Despite limitations, PP is high for many nodes across the tree, firmly demarking the phylogenetic framework. The initial survey of individuals and heuristic accumulation of additional gene data outlines the major lineages, for which the final full-length data are sufficient to define major groups (as seen in the 62t and 147t analyses). The data are sufficiently dense and hierarchical (the tree shape is phylogenetically nested) to provide effective support where needed, with the additional data-limited taxa appended to a relatively stable backbone that is based on more complete data. The addition of these extra taxa has only a small effect on the shape and stability (e.g. PP support) of the backbone, as illustrated by the similarity of the subtrees (Supporting Information Figs S2, S3). Increasing the number (perhaps better thought of as increasing the density) of taxa does not adversely affect the backbone, as long as taxa are closely linked to another with more data.

Here we could only compare subtrees to investigate the effect of missing data, rather than subsample gene data from a complete matrix. However,

Figure 7. Australian camaenid hierarchical nested supermatrix phylogeny with special emphasis on eastern species. Combined mtDNA data of 16S, COII, 12S: 450, 480, 522 sites, total = 1452 sites, 47% missing data. Markov chain Monte Carlo (MCMC) all-compatible posterior consensus from 160 000 post burnin samples using gene partition (3p) GTR-G-inv model. Number of genes contributing information to each branch indicated by colour and thickness: blue, three genes; red, two genes; black, one gene. Bar indicates 95% posterior probability (PP) support from combined two runs each of three models (1p, 3p, 4p). Asterisk (*) indicates taxa ignored for calculation of this PP support. Circle indicates ≥ 70% support with RAxML fast bootstrapping (5000 replicates). Basal polytomy rooted with *Satsuma japonica*. Some major clades are labelled, various groupings indicated, and some description and pictures added to aid discussion and for general interest. See text for details.


Northeast coastal forest hadroid snails
Inland scrub
Tropical rainforest
Inland
Inland
NSW - SE Australia
Widespread South-Eastern Australia hadroid snails
Inland SW Qld
South Australia
Inland Qld
Central desert
Northeast coastal rainforest hadroid snails
far southern forests
Arid inland SA
Chloritid
Hadroid group
1
2
3
4
Branch colour coding: blue 3 genes, red 2 genes black 1 gene.
Bar = ≥95% MCMC support from combined PP of three models (1p, 3p, 4p)
Circle indicates ≥70% support in RAxML 3p GTR-CAT bootstrap analysis
** = taxa ignored for calculation of PP support
Rooted with Satsuma japonica.
PP >0.95 in combined models MCMC
BS >70% in RAxML fast bootstrapping
Both PP and BS support

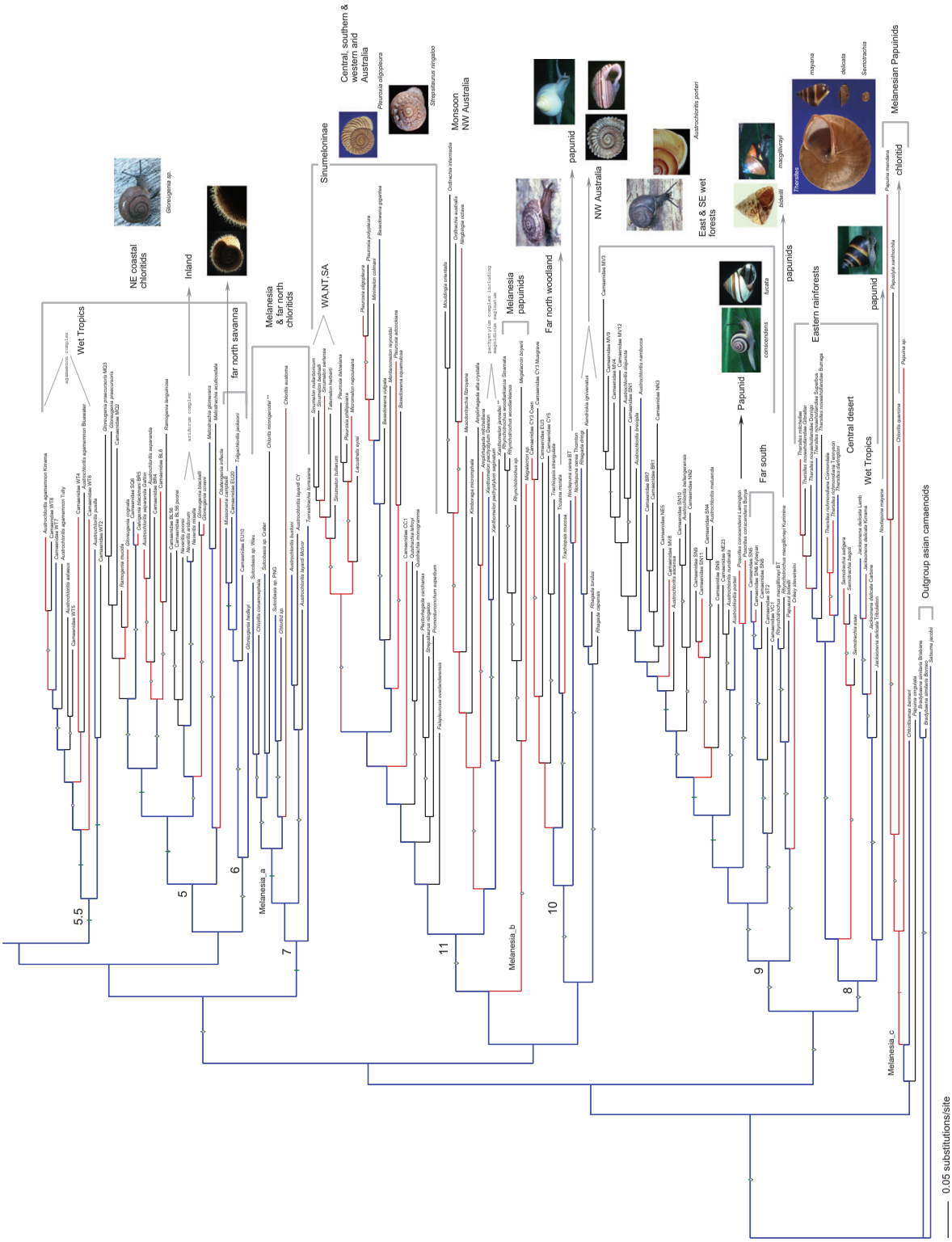


Figure 7. *Continued*

simulation studies on the effect of missing data indicate that with the appropriate density of data, accurate results can be obtained despite high proportions of missing sites albeit with some caveats (see Material and methods for references). It is still a question of judgment as to whether the fewer taxa but more complete matrix should be favoured over the more taxa but less data complete matrix (the more taxa or more data debate, with evidence pointing to the value of both). The results indicate that the larger 147t and 327t analyses are topologically more similar to one another than they are to the 62t analyses. Therefore we suggest that the 147t analysis represents the best balance between taxon sampling and data completeness for phylogenetic inference amongst the deeper lineages: the additional information outweighs any distortion and instability resulting from missing data.

Another approach used to build large trees from smaller sets of (overlapping) data is the 'divide-and-conquer' supertree method of assembling of separate (e.g. smaller) taxa-set trees into a synthetic topology (without branch lengths) (e.g. Bininda-Emonds, 2004). In contrast, the hierarchically nested supermatrix approach (for all its flaws) uses all of the data simultaneously to infer a complete tree (with branch lengths) in a model-based analysis, which may be thought of by comparison as the 'united-we-stand' philosophy. There is a redundancy as a result of phylogenetic nesting such that large parts of the data are not necessary. This and the similarity of our subtree comparisons suggest that the final nested supermatrix is adequate for the phylogenetic inferences made below.

Taxonomic implications

The following discussion principally refers to the 327-tip tree (Fig. 7) and for the purposes of organizing the results some of the major clades on that tree have been labelled (1–11 and Melanesia a, b, c). Nomenclature is based around the Zoological Catalogue (Smith, 1992; AFD, 2007). The combined model ≥ 0.95 PP across all three taxa sets is used as the basis for determining phylogenetic status.

We define several broad regions: Western Australia, South Australia, Top End, Kimberleys, Red Centre, and Eastern Australia (the 'east coast'). The first five are as defined in the Solem monographs (Solem, 1979b, 1981a, b, 1984, 1985, 1992, 1993, 1997). We further subdivide the east coast into Cape York (CY: Torres Strait to the Laura Gap), far north Queensland (FNQ: including the Wet Tropics; from the Laura Gap to the Burdekin Gap), mid-east Queensland (MEQ: from the Burdekin Gap to the St. Lawrence Gap), south-east Queensland (SEQ: the St. Lawrence gap to the Border Ranges), north-east New South Wales (NENSW: The Border Ranges to the Hunter Valley), and south-east Australia (SEAust: NSW south of the Hunter Valley into Victoria). Some of these are indicated on Figure 1. These divisions are primarily for descriptive purposes but they are based around natural centres of diversity and biogeography (Burbidge, 1960; Bishop, 1981; Cracraft, 1986; Joseph, Moritz & Hugall, 1993; Stanisic, 1994; Crisp *et al*., 1995, 2001; Hugall *et al*., 2003; Moritz *et al*., 2004).

Rational approach to nomenclature

The taxonomy as it currently stands (Zoological Catalogue AFD, 2007) can be split in two: the taxa covered by the Solem monographs, and the eastern taxa. The Solem (1979b, 1981a, b, 1984, 1985, 1992, 1993, 1997) monographs are a comprehensive modern revision using both shell and anatomical characters; the taxonomy of the eastern groups is by-and-large 19th and early 20th century conchology, with some more recent rationalization but little additional analysis (Burch, 1976; Bishop, 1978; Smith, 1992; Stanisic, 1994, 1996, 1998, 2001; Scott, 1996a; Clark, 2005, 2009). The inadequacy of the current eastern taxonomy and the difficulties blocking any real progress have been well accepted and obvious, right from the outset [Mörch, 1867; Cox, 1868; Pilsbry, 1890 (series); Hedley, 1912; Iredale, 1933, 1937a, b, 1938; Burch, 1976; Solem, 1979b, 1997; Stanisic, 1996]. Therefore it is amongst these groups that the molecular data make the essential contribution for revision, providing a rational basis for any further elaborations (e.g. splitting genera). This is principally for the eastern hadroid taxa. We also note major taxonomic discrepancies in the other eastern lineages (e.g. *Noctepuna*, *Austrochloritis*). As the remainder (e.g. central, western, and extralimital taxa) are only a subsample, we only comment on broad biogeographical implications.

We have intraspecific information on numerous 'hadroid' taxa and informal codes (in clades 1, 2, and 4, and to a lesser extent clade 3) that can contribute in part to questions of specific status, but irrespective of their status, the formal names and informal codes can be assigned to lineages (demarked by clades) and hence higher ranks (e.g. genera). Quite a number of the current eastern taxa form zoogeographical series (cf. Cain, 1954: 48–72). With many of these zoogeographical complexes, the specific names refer to shell variation, sometimes restricted to a locality within the broader range. Many of these combinations of shell variant and geographical locality are sufficiently vague to leave only the type specimens as unequivocal members of the taxon. Leaving aside questions of the biological significance of these taxa, they are a useful guide to sampling but create a practical problem. Fortunately, sufficient sampling across the entire range should allow a phylogeographical demarcation

of these zoogeographical complexes. The original formal names may then quite reasonably be assigned to lineages, at some level within the phylogeny.

As it currently stands the Zoological Catalogue has 83 Australian genera, 33 of which are monotypic. Of the remaining 50 genera, we have multiple species representatives in 28. Leaving aside decisions of specific status, of 30 polytypic genera where we have more than one species analysed, at least 18 (60%) are not monophyletic (using the criterion of ≥ 0.95 PP in all taxa sets, e.g. Figs 3, 4, 7). Thus the majority of tested genera are not monophyletic, highlighting (1) the repeated evolution of shell forms and (2) that the current higher taxonomy is untenable. Importantly, for most of these this is because taxa are spread amongst major lineages and not amongst just relatively closely related sibling lineages, where phylogenetic uncertainty and gene tree sorting could be raised as an explanation (for example compare the non-monophyly of *Jacksonena* with that of *Hadra*).

We do not have representatives of all the genera Solem assigned to three subfamilies (Solem, 1997: 1868–1870) but the sampling is adequate to investigate all three. Of these, only the Sinumeloninae is consistent with the mtDNA. All the genera assigned to Sinumeloninae are grouped (Fig. 7 clade 11), which also includes at least another ten (AFD) genera, all of which Solem left unassigned. The Pleurodontinae cannot be monophyletic as the Australian members belong to at least three separate lineages, not to mention that it also refers to Caribbean taxa (Cuezzo, 2003; Wade *et al*., 2006). The Australian taxa Solem placed in Camaeninae are a subset of clades 2 and 3, containing the South Australian genera and *Hadra* (leaving aside *Damochlora* for which there is yet no molecular information). Thus, even leaving aside the (probably unrelated) Asian members, the Australian Camaeninae deserves reassessing.

Hadroids

We are particularly interested in focusing on the generally large often striped hadroid type snails that are a dominant snail fauna of the eastern coastal forests. By excluding *Thersites* (which belongs to the unrelated clade 8), we can redefine the hadroid group as comprising the well-resolved group of clades 1, 2, and 3, plus clade 4, and probably *Vidumelon watti*, *Aslintesa camelus*, *Chloritobadistes victoriae*, and the *Offachloritis – Mussonena spinei* lineage of 'chloritids'. These are either within this hadroid group or near the base but in any event are sufficiently close to need be accounted for. Sampling for clade 3 is known to be incomplete, missing several (see Supporting Information Table S1) western NSW taxa and having only a subsample of the South Australian (Camaeninae *sensu* Solem, 1997) genera (11 of 31 species in five of six genera). The remainder of this hadroid group (clades 1, 2, and 4) are well sampled (all AFD species).

The phylogenetically distinct assemblage of clades 1, 2, and 3 contains a large proportion of eastern species, from Torres Strait to near the southern limit, and contains all the South Australian Camaeninae (*sensu* Solem, 1997) genera except *Aslintesta camelus* [there are no molecular data for *Pseudcupedora* but it probably belongs, as Solem (1992) considered it allied to *Cupedora*]. Clades 1, 2, and 3 have adjacent ranges with very little overlap, spanning Cape York to the NSW/Victorian border, and west to South Australia. Each contains numerous species (with local sympatric diversity of up to three species per clade). Allied to this is a monotypic lineage Camaenidae BL7 isolated to ranges in inland central southern Queensland. Within the north-eastern part, clade 4 has an overlapping distribution (from Cape York to NENSW).

Severing Sphaerospira

The most speciose genus in this Hadroid group is *Sphaerospira* (Mörch, 1867), which has been used to describe the large globose often striped taxa of north-eastern Australia. However as it currently stands *Sphaerospira* is polyphyletic, with members found in clades 1 and 4 (Stanisic, 1996). The type species is *Sphaerospira fraseri* (Griffith & Pidgeon, 1833), and as it belongs to the clade 4 group of *Sphaerospira*, we prefer to use the name *Sphaerospira* to refer to this assemblage of species. Together with the sister/sibling lineage of *Gnarosophia* (and probably CY1 and WT1) this group has a biogeographical and environmental milieu of the mesothermal and subtropical rainforest archipelago (Nix, 1982; Joseph *et al*., 1993; Hugall *et al*., 2003) and so forms a natural group, which we call the *Sphaerospira* lineage.

Notwithstanding some limitations of resolution, the group spanning *Offachloritis – Mussonena spinei* has a parallel distribution across the range MEQ to NENSW. All the taxa in this lineage have been referred to as Chloritidae (Iredale, 1933, 1938). This group also contains *Ramogenia challengeri* but *Ramogenia* is not monophyletic, with *Ramogenia mucida* and *Ramogenia lanuginosa* belonging within clade 5 (where these two do not appear to be monophyletic either). Neither is *Mussonena* monophyletic, with *Mussonena campbelli* belonging to a separate lineage with clade 5. *Offachloritis* is monotypic; *R. challengeri* is the type species of *Ramogenia*, and *Mu. spinei* is the type species of *Mussonena*. Therefore, a simple approach here would be to restrict the use of *Ramogenia* and *Mussonena* to these types and related lineages (e.g. SQ4). This leaves the whole lineage as a group of monotypic genera but each of considerable conchological difference to one another (as the original names implied).

The remainder of *Sphaerospira* species comprises all of the taxa in clade 1, except the monotypic *Monteithosites* and *Temporena*, along with numerous informal code units. This clade has a mostly coastal distribution from south-east Queensland to the southern Wet Tropics. *Temporena* is embedded within a complex of taxa belonging to the *Sph. sardalabiata* complex (Stanisic, 1998). *Monteithosites* is a disjunct northern outlier endemic to parts of the Wet Tropics. Three taxa also occur in inland more xeric areas, one in the north (*Sph. zebina*) and two further south (*Sph. mattea* and *Sph. volgiola*).

The full COII dataset contains: 20 individuals in the *sardalabiata* complex; 36 in the *mitifica–rawnesleyi–mazee* clade; 47 in the *coxi–etheridgei* clade; 33 in the *yulei* complex; seven in the *zebina* lineage. Thorough sampling across the range establishes a series of phylogeographical units corresponding to the *etheridgei*, *coxi*, *sardalabiata, 'mourilyani'*, and *yulei* complexes, together with the disjunct northern outlier *Monteithosites* (indicated on Fig. 7). Clade 1 is highly structured phylogenetically and biogeographically. Within subclades there is phylogeographical structuring according to informal codes and to a lesser degree AFD names. Within each of these lineages there are a set of names each corresponding broadly to a zoogeographical series. There is also some higher biogeographical structuring amongst subclades, with the highest level of diversity in mid-east Queensland where members of the *etheridgei*, *coxi*, and *yulei* groups can be found in sympatry (top illustration Fig. 7). Immediately to the north of this (within the Burdekin Gap, a major biogeographical feature; Cracraft, 1986; Joseph *et al*., 1993) lies the *sardalabiata* complex, which is basically confined to small pockets of tropical seasonal vine scrub. Further north (of the Burdekin Gap) is a diverse group containing numerous names, which broadly form a phylogeographical series, across lowland scrub and into upland mesothermal forest. An exception to this is at Mount Elliot, where there is an overlap of two lineages, one from the (lowland) *rawnesleyi* complex and another endemic to upland Mount Elliot, the morphologically distinct Camaenidae BL33. Upland Mount Elliot is a southern outlier of the Wet Tropics mesothermal biome, with a subset of taxa and several endemic vertebrates (Williams, Pearson & Walsh, 1996; Hoskin & Couper, 2004). It is tempting to suggest that BL33 represents an endemic re-evolution of a montane rainforest adapted species. Whatever the case it deserves recognition.

The northern '*mourilyani*' complex incorporates the taxa *bellaria*, *cookensis*, *mazee*, *mitifica*, *mourilyani*, *bala*, *rawnesleyi*, and *tomsoni*. Several of these are defined by islands: *bala* Magnetic Island; *cookensis* Brook Island; *bellaria* Hinchinbrook Island; *mitifica* Palm Islands (treating the reference to Mulgrave River as referring only to far north Queensland). Broadly the type geography forms a zoogeographical series; therefore allowing reasonable association with the phylogeographical series indicated by the mtDNA. More work is needed to delimit species but overall this lineage appears to be undergoing diversification.

Whereas *Monteithosites* and *zebina* are phylogenetically distinct, *Temporena* is not. There is a number of generic names that might be used to label the discreet clusters within clade 1: *Bentosites*, *Varohadra*, *Temporena*, *Sphaerospira*, *Gnarosophia*, and *Figuladra* (Iredale, 1933, 1937b; Burch, 1976; Smith, 1992). However, as originally defined, none of them is consistent with the phylogeny. As *Sphaerospira* and *Gnarosophia* can be fixed to clade 4 taxa (*Sph. fraseri* is the type species of *Sphaerospira*; *G. bellendenkerensis* is the type species of *Gnarosophia*: Stanisic, 2001), the remaining four are candidates for clade 1 generic units. Taxa in the *etheridgei-coxi* lineages have been given the genus name *Bentosites* (Iredale, 1937b) with *Bentosites macleayi* the designated type species. Taxa in the clade containing *mattea*, *appendiculata*, and *volgiola* have been called *Figuladra* (Iredale, 1933) with *curtisiana* as the designated type species. Currently *curtisiana* is regarded as a subspecies of *incei*; Iredale regarded *Figuladra* as a subgenus of *Varohadra* and synonymized it with *Varohadra* (Iredale, 1937a).

A simple practical resolution is to use *Bentosites* for the *etheridgei-coxi* lineages, *Figuladra* for all in the BL4-*mattea-appendiculata* lineage, *Temporena* for all in the *sardalabiata* complex, and retain *Monteithosites*. This leaves the *mitifica-mazee* lineage, the *zebina* lineage, and the *yulei* complex to be re-assigned. Iredale (1937b) put all the *yulei* complex taxa into *Varohadra* with clade 4 taxon *oconnellensis* as the type species, which is now committed to *Sphaerospira* (Smith, 1992; see also Stanisic, 1996; Hugall *et al*., 2003). Consequently a new name will have to be proposed. The only names associated with *zebina* are *Sphaerospira* (and originally *Helix*; Smith, 1992); therefore *zebina* also requires a new generic name. Taxa in the *mitifica-mazee* lineage have variously been assigned to *Gnarosophia*, *Varohadra*, and *Sphaerospira*: none of which is suitable. Therefore this lineage also requires a new name. With this arrangement, any group can be further elaborated (split) if and when necessary.

There is a large scale replacement of clade 1 animals by clade 2 and clade 4 north of the southern Wet Tropics. Clade 1 distribution extends a little more northerly on the offshore islands than on the mainland. In several places clade 1 is found in upland rainforest but this is only where clade 4 is absent (Mt Elliot, Paluma, Seaview, Cardwell) with limited

overlap in the Kirrama region, and along the coast north of Tully. *Monteithosites* is a northern outlier on the isolated ranges inland of the northern Wet Tropics but again where clade 4 is absent. Here it overlaps with clade 2 lineages, and there is limited overlap between clades 1 and 2 in the southern Wet Tropics, in the Cardwell Range and on some offshore islands.

Clade 2

Clade 2 is phylogenetically distinct; however, structuring within the clade is less pronounced. This clade is distributed from Torres Strait down the east side of Cape York, through Far North Queensland down to the southern Wet Tropics, and west out to the Einasleigh Uplands. In the coastal areas the *Hadra* taxa dominate whereas inland it is *Spurlingia*.

The full COII dataset contains: 78 individuals in the *H. bipartita* complex in the *bipartita* complex; 22 in the *S. dunkiensis* clade; 21 in the *S. portus* clade; six in *H. barneyi*; two in *Jacksonena rudis*; seven in the *S. forsteriana* lineage. Excepting *S. forsteriana*, the *Spurlingia* taxa (including numerous codes) form a phylogeographically structured series across the steep rainfall gradient from east to west. Of the formal names within *Spurlingia*, *dunkiensis* is inadequate to describe the lineages associated with that name. Molecularly there are broadly two major lineages (each with multiple units); one can be linked to *S. dunkiensis*, therefore the other might be associated with *S. portus*. Despite limited resolution neither *Hadra* nor *Spurlingia* appear monophyletic, even ignoring *Spurlingia forsteriana*, which Iredale (1933) made the type species of *Zyghelix*. Embedded within the remaining *Spurlingia* are *Jacksonena rudis* and *Hadra barneyi*. *Jacksonena* is not monophyletic, with the other species *Jacksonena delicata* belonging to clade 8; therefore, *Jacksonena* should be kept for the type species *rudis* (Iredale, 1937b). Hedley (1912) originally placed *delicata* in *Planispira*, which we believe should be restricted to unrelated South-East Asian taxa. Considering some phylogenetic uncertainty (as there is some instability amongst analyses and all lineages are relatively close siblings), *barneyi* could be left in *Hadra*. Alternatively we could resort to Iredale (1933, 1937b) and place *barneyi* in the monophyletic *Micardista*, allowing a paraphyletic *Spurlingia*. It would then be a simple elaboration to erect a new genus for taxa in the *portus* lineage (e.g. *portus*, CY9), if it proves worthwhile.

The generic *Hadra* is probably best reserved for the phylogeographical (and roughly parallel taxonomic zoogeographical) series encompassing the Wet Tropics *H. webbi* and *H. bipartita* (the type species), the island *semicastanaea*, Torres Strait *H. funiculata* and *H. bartschi*, together with the Cape York CY10 and inland EU17. The taxon CY10 spans two distinct mtDNA lineages, one in the central Cape York (McIllwraith and Iron) Ranges; the other further north, allied to the Torres Strait taxa. This latter component of CY10 may be better referred to the Torres Strait taxa. The Lizard Island endemic *Hadra semicastanea* is not phylogenetically distinct but embedded within northern FNQ haplotypes of *bipartita*. *Hadra webbi* is associated with southern upland sites. Although there is geographical structuring, *H. webbi* is not phylogenetically distinct by mtDNA but a subset of southern haplotypes. In some upland areas *H. webbi* takes on such a flattened and angulate form that it has been mistaken for *Jacksonena rudis*, which is co-distributed on the Atherton. Thus *H. webbi, H. bipartita, H. semicastanea*, and EU17 form a phylogenetic series most simply interpreted as a phylogeographical extension of *bipartita*, with *webbi* as an indistinct southern form, EU17 a peripheral inland isolate, and *semicastanea* an island population. CY10 and the Torres Strait taxa are then phylogenetically distinct sibling lineages. Whereas the Torres Strait lineage is phylogenetically distinct, there appears to be little clear distinction between *H. bartschi* and *bartschi*, other than island typology. This whole area was land as little as 7000 years ago (Yokoyama, Purcell & Lambeck, 2001).

Overall, clade 2 lineages show substantial morphological variation, correlated to environment. There is a strong correlation of shell size, aspect ratio, and thickness with rainfall and temperature, so that rainforest animals are large and globose whereas island taxa are smaller, thinner and flatter. The latter appears also to be influenced by temperature so that higher elevation animals tend to have flatter shells. The flatter the shell, the more angulate, and this may be related to paedomorphosis, as juveniles of even the most globose *H. bipartita* have a pronounced angulation. As a whole, clade 2 is spread across a wide rainfall and temperature range. It also spans considerable conchological variation and both of these are seen at multiple phylogenetic scales. The environmental and morphological variation in one lineage may be seen in another, and that seen within one species complex mirrors that seen across genera. The *Spurlingia* lineages can be seen as an evolutionary extension of the variation seen in the *H. bipartita* complex: the extension in geographical and environmental space matched by extension in conchological space. Thus *Spurlingia praehadra* and Camaenidae EU9 the furthest west and xeric are the smallest and flattest. Therefore the molecular phylogeny can be interpreted as indicating repeated geographically based parallel evolution of dry adapted forms, hence the interdigitation of genera.

We do not have molecular information on the remaining AFD *Hadra* species *Hadra wilsoni*,

endemic to the Kimberleys region. This species was described on the basis of poorly presented shells and placed here pending the availability of more data (Solem, 1979b). Given the distribution of the others, a relationship to clade 2 is questionable.

Clade 3

Clade 3 has a very large distribution from inland central Queensland down through eastern NSW to the Victorian border, and across central western NSW to South Australia. In the more northern and north-western part of its range, clade 3 is replaced by clade 1, with overlap principally confined to areas within the distribution of '*Sphaerospira mattea*' in central western and south-east Queensland. Through south-east Queensland and coastal NENSW clade 4 overlaps with clade 3.

Taxa of this group dominate the south-eastern part of Australia (NSW), containing the genera *Meridolum*, *Galadistes*, and *Ventopelita*. There are numerous informal codes in this clade with substantial phylogenetic structuring; therefore the species diversity as it currently stands in the Zoological Catalogue is likely to be a substantial underestimate. The recent revision of *Meridolum* taxa by Clark (2005, 2009) using morphological and genetic data has gone some way in elaborating this, especially for the Sydney basin region, and we depart from the Zoological Catalogue to incorporate this new nomenclature. Amongst *Meridolum* taxa three complexes are worth mentioning: the *Meridolum monacha* (= *bowdenae*)-*duralensis*-*mastersi* complex, the *Meridolum corneovirens*-*middenense*-*gulosa* complexes, and *gilberti* and allied lineages. In the Clark (2009) revision the former complex has been re-assigned to the new genus *Pommerhelix*, with *bowdenae* replaced by *Pommerhelix monacha*. Northern NSW and southern Queensland diversity associated with *Meridolum gilberti* and allied lineages has been assigned to taxa within the new genus *Ponderconcha* (and possibly Stanisicia). Each of these groups contains numerous molecular lineages, taxa of uncertain description, informal codes, and some recently described species (Clark, 2005, 2009). Although there is still some uncertainty, especially in the northern NSW taxa, the phylogenetic diversity is a broad indication of regional biological diversity. Several complexes (e.g. *Meridolum* and *Pommerhelix*) widely overlap the greater Sydney basin region, probably contributing to the taxonomic difficulties. The new montypic genus *Sauroconcha caperteeana* that has been recently discovered in this region (Zhang & Shea, 2008) appears to belong within clade 3 (M. Shea, pers. comm.).

Clade 3 as a whole is phylogenetically distinct and with a series of species clusters, but with little resolution amongst these, and some instability amongst analyses (e.g. see Figs 4, 7). The bulk of *Meridolum sensu lato* appears monophyletic, notwithstanding *Me. exocarpi* for which there is no molecular data, and *Me. bennetti*, which may be ill-defined; both of these taxa being known only from type localities, and the phylogeny is also broadly consistent with the recent generic revision by Clark (2005, 2009). We are missing three of seven *Galadistes* and several *Galadistes* taxa are best considered as ill-defined (Clark, 2005); however, it may also not be monophyletic.

Of the sampled taxa, several lineages are found widespread across inland regions: the phylogenetically distinct lineage containing *Adclarkia*, Camaenidae BL12, and Camaenidae BL13 is therefore widespread across central southern Queensland. Another phylogenetically distinct cluster also distributed across southern inland Queensland from the Darling Downs to the Expedition Ranges is associated with various names: *Meridolum bennetti, Galadistes expeditionis*, the monotypic *Pallidelix greenhilli*, and several informal codes. As both *Meridolum* and *Galadistes* are better preserved for other groups (on the principle of type species), this lineage might form the rational basis for the genus *Pallidelix*. Thus in southern inland Queensland, despite relatively low species diversity, there are (at least) two distinct divergent lineages within clade 3 with widespread overlap contributing to diversity.

This leaves the problem of *Ventopelita*, and several other distinct lineages. There are three species of *Ventopelita*, of which we only have information on two but these appear to belong to separate lineages within clade 3. The position of the remaining divergent lineage containing Camaenidae NE14 is not well resolved but is clearly not closely related to anything else in the phylogeny. Considering this lineage, and those associated with *Ventopelita*, Camaenidae MV2, *Galadistes liverpoolensis* and *Petraconcha griffithsi*, there is scope for substantially more overlapping diversity across eastern NSW.

Further sampling along the Great Dividing Range across NSW and southern Queensland is required to demark phylogeographical patterns sufficient for further taxonomic interpretation and resolution of *Galadistes*. Suffice to say that the current level of sampling captures a large proportion of regional phylogenetic diversity.

The South Australia component

There are ten genera in the South Australian region (Smith, 1992; Solem, 1992, 1997; defined as not including the far northern ranges, which are included in the Red Centre region). However these ten fall into two separate major groups or subfamilies: *Lacustrelix*, *Micromelon*, *Pleuroxia*, and *Sinumelon* all fall into clade 11 and were assigned to the Sinumeloninae

subfamily (Solem, 1997). The remaining six genera were assigned to the subfamily Camaeninae: of these *Cupedora*, *Contramelon*, *Cooperconcha*, *Glyptorhagada* (and probably *Pseudcupedora*) all fall in clade 3. Although we have only 11 of the 31 (AFD and Solem, 1992) taxa in these genera, as they all fall into one area in clade 3, it is probable that all the remaining taxa will also fall within clade 3. The monotypic *Aslintesta* is a separate lineage stemming from near the base of the hadroid group. There is sufficient sampling to suggest that *Glyptorhagada* may not be monophyletic but *Cupedora* may well be so. The 11 South Australian taxa sampled do not appear to form a monophyletic group with respect to more eastern lineages but there is some uncertainty (see Fig. 4). Solem (1992) considered that the South Australian 'Camaeninae' may have originated multiple times from eastern taxa; although we do not have all the species we have most of the genera and the evidence supports a relatively discrete eastern origin with the possibility even of reverse dispersal (e.g. the Queensland *Adclarkia* group). Further sampling both of South Australian taxa and western NSW taxa could provide a clear framework for the origin and evolution of the South Australian fauna.

Below the (above-defined) hadroid group is a series of well-defined groups (clades 5–7), the arrangement amongst which is not clearly resolved, but are all north-eastern Australian and Melanesian chloritid (*s.l.*) type animals.

Chloritids and papuinids

Two major types (or forms; cf. facies) of camaenid, the terrestrial chloritids (family Chloritidae; globose low aspect ratio < 1, brownish, frequently with periostracal hairs), and the arboreal papuinids (family Papuinidae; colourful conical high aspect ratio > 1, e.g. the celebrated Manus green tree snail), reiterate throughout the tree, and at relatively low phylogenetic levels (e.g. less than half the divergence height of the mtDNA trees).

Chloritid (*s.l.*) forms appear across the tree in as many as six separate lineages. Within this, the Australian genus *Austrochloritis* (as is currently defined: Smith, 1992; AFD, 2007) is split into at least four separate lineages: the lineage containing *agamemnon* and *pusilla* (clade 5.5), the lineage containing *separanda* (clade 5), the lineage containing *layardi* and *buxtoni* (clade 7), and the large complex in clade 9 spanning *porteri* – *disjuncta*. There is a major biogeographical division amongst all these Australian lineages, with the clade 9 group distributed from the southern limit in Victoria to the Border Ranges (separating SEQ and NENSW). All the remaining types are found north of here: SEQ, MEQ, FNQ, Cape York, with clade 7 also containing Melanesian chloritid taxa. A minor exception to this is *Neveritis* (in clade 5), which extends into southern Queensland and northern NSW west of the Great Dividing Range). The clade 9 austrochloritid complex may or may not contain the papuinid (*s.s.*) *Posorites conscendens* group, which is distributed across SEQ–NENSW. Leaving this aside, the clade 9 austrochloritids split into a far southern lineage, and a diverse series of overlapping lineages distributed successively from south-east Australia to the Border Ranges, with greatest diversity in NENSW.

Untangling Austrochloritis

As the type species for *Austrochloritis* is *Austrochloritis porteri* (Pilsbry, 1890; see Smith, 1992) this genus can be applied to all the taxa (and lineages) in clade 9 for which *Austrochloritis* has been applied – and are clearly of the chloritid (*s.l.*) form. This is, with the possible exception of the *Posorites conscendens* lineage, a monophyletic group in the mtDNA analyses. We were unable to get molecular data for the other two *Posorites* species, and given polyphyly of other papuinid taxa, it is unsafe to assume they are related to the *P. conscendens* lineage. Notwithstanding further complications from as yet unsampled lineages, even if *Posorites* does split the clade 9 austrochloritids, allowing for paraphyly these can all still be included in *Austrochloritis*.

Restricting *Austrochloritis* to the clade 9 assemblage requires reassigning the other four groups.

By returning *Austrochloritis victoriae* to *Chloritobadistes* (see Cox, 1868; Iredale, 1938; cf. Solem, 1979b), we affirm that this refers to the chloritid-like snail with periostracal 'scales' from the Otway Ranges as distinct to the chloritid types in eastern Victoria with periostracal 'hairs', which belong to a lineage within the clade 9 complex. Solem (1979b) used snails from Wilsons Promontory to represent *A. victoriae* and found them similar to *Austrochloritis disjuncta*, which is the case here with the Wilsons Promontory taxa Camaenidae VC1 belonging to the type species group (clade 9 that includes *A. disjuncta*). Iredale (1938) based his assessment on Cox (1868) shells and remarked that it was a curious development of the *Meridolum* form, and certainly not a member of the Chloritidae. The *Chloritobadistes* type is known to occur in the Grampians, the Otways, and the ranges east of Melbourne, where it overlaps with the *Austrochloritis* type (A. F. H., unpubl. data). Just how widespread and overlapping these two lineages are in Victoria, and the scale of any zoogeographical structuring within them, will require further collection and analysis.

All the members of clade 7 are nominally regarded as chloritids ('family' Chloritidae; Pilsbry, 1890), and all have northern distributions. *Torresitrachia* (see

Solem, 1979b: 45–92) has a wide distribution across northern Australia from the Kimberleys in the west to Cape York in the east, and also New Guinea. However, the phylogenetic status is untested as we only have *torresiana*, the sole eastern species. The lineage containing *Austrochloritis layardi* and *Austrochloritis buxtoni* spans Torres Strait, Cape York, across the Laura Gap to the northern part of FNQ. Across the Laura Gap *A. layardi* is split into two deep lineages. Iredale (1938) placed *A. layardi* and *A. buxtoni* in the monotypic genera *Nannochoritis* and *Patrubella*, respectively. The remaining component of clade 7 is a diverse group of Melanesian chloritid taxa. We have only a subsample of Melanesian chloritids from Papua New Guinea and the Solomon Islands. Needless to say, *Chloritis* (*s.s.*) is not monophyletic because *Chloritis quercina* belongs to a lineage unrelated to clade 7. *Chloritis* is an old name that was used for a wide array of chloritid form shells from South-East Asia to Australia (Beck, 1837; type species *Chloritis ungulina* Linnaeus, 1758), reflecting the phylogenetic complexity of Chloritidae (*s.l.*).

The lineage comprising the *Austrochloritis agamemnon* (Gude, 1906) complex and sister taxon *Austrochloritis pusilla* (clade 5.5) is of similar divergence level to the entire hadroid group but is restricted to the Wet Tropics mesic forest habitats, and comprises only three (named) species. *Austrochloritis agamemnon* is a complex of several overlapping lineages and phylogeographical units, although it is not yet possible to completely untangle the two. For example *A. pusilla* is split into a northern and a southern lineage, typical of many Wet Tropics species (Schneider, Cunningham & Moritz, 1998; Moritz *et al*., 2002). These overlap *A. agamemnon* which is further split into overlapping lineages (e.g. WT4 – *Bluewater*; WT5 – *astaeus*; *agamemnon* Tully – Kirrama). These overlapping lineages are associated with informal codes embedded within *A. agamemnon*. Whatever the case, the phylogenetic diversity underlying the informal codes better captures the biological diversity than the current taxonomy. *Austrochloritis agamemnon* (Gude, 1906) was originally placed in *Chloritis* but our data do not support this. Therefore the clade containing *agamemnon, astaeus*, and *pusilla* requires (at least one if not more) new generic names. As they are all currently under one name, and are endemic to the Wet Tropics region, a single unifying genus may be acceptable.

Austrochloritis separanda belongs to a large complex of chloritid (*s.l.*) type taxa distributed along the coast from north of the Border Ranges to the southern Wet Tropics (clade 5), and west from northern NSW to the Carnarvon Ranges in central western Queensland. Clade 5 contains *Ramogenia*, *Gloreugenia*, *Calvigenia*, *Neveritis*, and *Austrochloritis*, along with numerous informal codes. *Calvigenia* is monotypic (but *blackmani* probably represents part of a more widespread phylogeographical series, e.g. including Camaenidae SQ6). Of the remaining genera, the type species are: *Gloreugenia coxeni*, *Mussonena spinei*, *Ramogenia challengeri*, *Neveritis poorei*, and *Austrochloritis porteri*. Notwithstanding the missing *N. interna, Neveritis* appears to be monophyletic. This genus has a very wide distribution, from inland southern Queensland to the southern Wet Tropics. With such a large range it is not surprising that there will at least be some phylogeographical diversity as indicated by the associated informal codes. Allowing paraphyly because of the monotypic *Calvigenia*, and removing *separanda* from *Austrochloritis*, still leaves *Ramogenia mucida* and *R. lanuginosa* interdigitated between *Gloreugenia*.

The sibling lineages of *Melostrachia* and *Obsteugenia*, which appear to be associated with clade 5 (but with limited support) are distributed across the Wet Tropics (*Obsteugenia*) and to the west (*Melostrachia*). A further associated lineage (clade 6) spans the Wet Tropics, and across the Laura gap to the McIllwraith Ranges in Cape York.

Using the principle of type species *separanda*, *mucida*, and *lanuginosa* would be re-assigned. As *coxeni* is the type species for *Gloreugenia*, this suggests that by moving all three (and associated lineages) into *Gloreugenia*, and *Gloreugenia hedleyi* to another genus, *Gloreugenia* would then form a coherent group within clade 5 (allowing paraphyly because of *Calvigenia*). At later times this *Gloreugenia* could be split into more genera if desired to demark distinct complexes. *Mussonena campbelli* also needs to be re-assigned: a simple remedy would be to collapse it into *Tolgachloritis*, to which it clearly is related. Originally *Chloritis hedleyi* (Fulton, 1907), then placed in *Gloreugenia* (Iredale, 1938), neither of which are suitable, *hedleyi* could also be placed in *Tolgachloritis*, reflecting probable affinities with that lineage, although it may ultimately be deserving of generic status. *Gloreugenia hedleyi* is likely to be a phylogeographical series as it has a wide distribution from southern Cape York to the southern Wet Tropics. However, the two *Melostrachia* species span the range of the distribution, as does the *Tolgachloritis* lineage. Therefore the long stem internodes are probably not artefacts of sampling but real phylogenetic structuring. Within clade 5 there appears to be a series of overlapping zoogeographical complexes, although the sampling is not yet enough to fully demark all of these.

Papuinids

The Papuinidae is a complex 'family' of snails with great diversity through South-East Asia and Melane-

sia (Pilsbry, 1890; Clench & Turner, 1966). It is not within the scope of this work to investigate papuinid phylogeny but we can discuss the Australian taxa. There are currently recognized eight Australian papuinid species (in four genera). Of these we have five in four genera plus the recently named Wet Tropics endemic Crikey steveirwini (Stanisic, 2009). In addition we have a sampling of extralimital diversity (eight species in four genera). Phylogenetically these Australian papuinid tree snails have at least three separate sources: in clades 8, 9, and 10.

As it stands, *Noctepuna* is not monophyletic as *Noctepuna mayana* and *Noctepuna cerea* belong to unrelated lineages (clades 8 and 10, respectively). However as we do not have information on the type species *Noctepuna poiretiana* we refrain from suggesting any nomenclatural revision. From our sampling, *Rhynchotrochus* is also not monophyletic, with the sole Australian species *Rhynchotrochus macgillivrayi* unrelated to the Melanesian *Rhynchotrochus woodlarkianus* complex. As we do not have the type species of *Rhynchotrochus* – the New Guinea species *Rhynchotrochus tayloriana* – we cannot say whether *macgillivrayi* needs to be reassigned; however it is worth noting that the comprehensive study by Clench & Turner (1966) considered the two to be quite similar in both shell and anatomy. The clade 9 *Rhynchotrochus macgillivrayi* lineage also contains two other Australian species, *Papuexul bidwilli* and *Crikey steveirwini*. We have only one of the three *Posorites* species, *Posorites conscendens*, which is another lineage within clade 9. Thus the basal sequence of clade 9 comprises papuinid–chloritid–papuinid–chloritid. Notwithstanding phylogenetic uncertainty (even though all nodes are well supported) and unsampled (or extinct) lineages, the implication is that either the papuinid form is a derived lineage from within a pre-existing austrochoritid group, or vice versa. Militating against the latter is that the papuinid form of *Posorites* may be distinct from that of the *Rhynchotrochus macgillivrayi* lineage taxa, and that the chloritid form is more widely distributed. Either way, the two forms (papuinid and chloritid) are interdigitated. This is seen elsewhere in the tree, where *Noctepuna cerea* is related to the *Trachiopsis*/*Trozena* type of chloritid (clade 10) and (caveat limited data) *Chloritis quercina* is (more) associated with papuinid species (than any other chloritid lineage in the tree). Resolution of the sampled Melanesian papuinid taxa is limited but they fall into several divergent lineages not related to any of the sampled Australian lineages. Altogether, the sample of papuinids here shows multiple origins and considerable phylogenetic diversity.

In Australia papuinids *s.l.* are absent from MEQ with a 'leapfrog' distribution on either side (Cape York, Wet Tropics, SEQ–NENSW). This distribution is seen in a number of mesic forest taxa and is possibly because of severe attenuation of MEQ rainforest habitat during the Pleistocene (Joseph *et al*., 1993; Low, 1993; Norman *et al*., 2002). Within clade 1, in MEQ mesic habitats the *etheridgei* group have diverged from their sibling lineages with heightened aspect ratio (> 1) and occasional semi-arboreal habit. Thus they may be described as having proceeded in the direction of a papuinid (*s.l.*) form.

Clade 8

A corollary of the reiteration of papuinids and chloritids across the tree is that virtually the entire conchological, environmental and geographical range of the whole Australasian Camaenidae can be found within any one subgroup. For example clade 8, comprising *Noctepuna mayana, Semotrachia, Jacksonena delicata*, and *Thersites*, contains a tropical rainforest tree-snail, miniature desert rock snails, acutely keeled subdiscoidal upland tropical rainforest endemics, and globose large temperate forest snails (see Fig. 7 and associated shell pictures). *Noctepuna mayana* is endemic to northern Wet Tropics rainforest, *Semotrachia* is confined to gorges in the Central Australian ranges, *Jacksonena delicata* is endemic to upland rainforest in the Wet Tropics, and *Thersites* is distributed in mesic warm temperate forest in SEQ–NENSW.

We have only a subsample (three of 25 species) of *Semotrachia*; they span a large part of the distribution across the Red Centre but appear to be relatively closely related. *Semotrachia* species are highly endemic with limited ranges within gorges of the central Australian ranges; therefore the genus may show detailed regional zoogeographical patterns. The remaining members of clade 8 are comprehensively sampled; therefore the overall shape of (the Australian part of) this clade is not an artefact but a product of evolutionary history (either they never existed, are to be found elsewhere, or went extinct – probably the latter). *Thersites* comprises a phylogeographical complex with taxa associated with particular sets of lineages. The lowland subtropical rainforest endemic *Thersites mitchellae* allied to the *T. novahollandiae* complex, and *T. darlingtoni* an upland species and close relative of the more northerly distributed *T. richmondiana* complex. *Jacksonena delicata*, as with so many other Wet Tropics endemics, shows strong phylogeographical structuring (Schneider *et al*., 1998; Hugall *et al*., 2002). As the unrelated *Jacksonena rudis* is the type species, and *delicata* is distinct from anything else ('.of strange appearance, yet withal a beautiful form.'; Iredale, 1937b: 22 referring to *Jacksonena*), it will need to be placed in a new genus.

The remaining assemblage of eastern chloritid taxa, *Trachiopsis* and *Trozena*, fall into clade 10. *Trachiopsis strangulata* and allied lineages (e.g. CY5) have a northern distribution in Cape York vine scrub. The related lineage has a disjunct distribution: southern Cape York (CY3), and to the west of the Wet Tropics (EU3). *Trachiopsis mucosa* has an enormous distribution (second largest of all eastern species), from SEQ to central inland Queensland and north to the southern Wet Tropics, again mainly in vine scrub, and is likely to be a phylogeographical complex. Sister taxa *Trozena morata* is parapatric to this in the north, distributed to the west of the Wet Tropics. Clearly related to this *Trachiopsis* group of chloritids is the papuinid tree snail *Noctepuna cerea*, endemic to the northern Wet Tropics rainforests (see Figs 2, 3, 4, 7). Given the wide distribution, phylogenetic depth, and ecological diversity in clade 10, there may be other unsampled papuinid and chloritid groups with northern (and extralimital) distributions that will be associated with clade 10, such as the Cape York endemic tree snail *Noctepuna poiretiana*. We do not have information on the chloritid genera *Damochlora* and *Parlogenia* (total of five species) found in the Kimberleys; however relationship to clade 10 is possible (see Solem, 1979b). The monotypic *Trozena* falls within *Trachiopsis* but further sampling is desirable before considering generic changes.

The Cape York rainforest endemic *Chloritisanax banneri* (also considered a chloritid) is a divergent lineage not clearly associated with any other lineage yet sampled. Rather than being a Cape York palaeoendemic, it may have closer affinities with as yet unsampled Melanesian chloritid taxa.

Grade at the base of the hadroids: interpretation

This grade of chloritid (*s.l.*) type lineages below the hadroid group (see Fig. 7) may indicate something of the origins of the eastern diversity as a whole, and the hadroid animals in particular. Below the hadroids there is a series or grade of lineages of chloritid mesic forest groups with further lineages deeply embedded near the base of the hadroids. Together these lineages span the entire range of the eastern distribution from Melanesia to Victoria, and west to Central Australia (and possibly as far as the Kimberleys). Several of the lineages stemming from near the base or just within the hadroids have intermediate form, in that they have been assigned as either Chloritidae or Camaeninae. There is a concentration of mesothermal forest types near the base of the hadroids (the *Austrochloritis agamemnon* complex; *Offachloritis et al.*; the *Sphaerospira* lineage), with several relict (divergent low diversity lineages), endemic to extremities of the range (*Chloritobadistes*, *Aslintesta*, *Vidumelon*, and also note Camaenidae BL7). This pattern is consistent with ancestral widespread distribution of chloritid type snails that have persisted through subsequent periods of environmental change fragmenting the mesic forest habitat and giving rise to more xeric megathermal woodland habitats, with evolutionary consequences of attenuation of the pre-existing ancestral distribution and diversification of lineages adapted to the more xeric megathermal woodland habitats. As the chloritid form is very widely distributed, it may well be a plesiomorphic archetype through which other taxa repeatedly evolve. This aspect of the phylogeny may be useful in a comparative analysis on the evolution of the characteristic chloritid periostracal structures (cf. Goodfriend, 1986).

Clade 11: Sinumeloninae

We have a subsample of generic diversity in the western part of the continent, Western Australia, Kimberleys, Red Centre (see Fig. 1) amounting to 22 of 42 AFD genera.

The only other eastern genus, *Xanthomelon*, consistently allies with a very diverse clade 11, which contains many taxa from the western part of the continent. Broadly, *Xanthomelon* is a tropical monsoon forest and grassland group and has a *trans*-continental distribution, Kimberleys, Top End, and east coast but we do not have Top End and Kimberleys species. The Zoological Catalogue (Smith, 1992; AFD, 2007) lists five eastern *Xanthomelon* species: *Xanthomelon jannellei*, *Xanthomelon pachystylum*, *Xanthomelon distractum*, *Xanthomelon magnidicum*, and *Xanthomelon saginatum*. *Xanthomelon pachystylum* is the most widespread eastern species with a distribution encompassing much of the southern Brigalow Land Bioregion, south of the Fitzroy River; *X. distractum* from the Rockhampton region to near Bowen; the large *X. magnidicum* from the Whitsunday region; *X. saginatum* from the Einasleigh Uplands and *Xanthomelon jannellei* from Cape York, FNQ. Based on albeit limited sampling from Cape York to central Queensland (six specimens), *X. jannellei* appears phylogenetically distinct whereas *X. pachystylum* and *X. saginatum* may be parts of a more closely related zoogeographical complex.

The remaining taxa in clade 11 all belong to the western part of the continent, and this clade contains all of the taxa Solem (1997) assigned to the subfamily Sinumeloninae (that we have sampled: ten genera; Solem also included *Granulomelon*, *Eximiorhagada*, and *Caperantrum*, for which we have no information). Clade 11 also includes another ten genera that Solem left unassigned. The type genus *Sinumelon* is one of the most species diverse and widespread in Australia, distributed from north-western Australia across the Nullabour plain and through the Red Centre and

South Australia. *Sinumelon* contains what are considered to be true desert snails, with free-sealing ability and inhabiting spinnifex sand country (Solem, 1993, 1997: 1524–1526). We only have four species spanning a large part of the total range and these four appear to be all relatively closely related. Another 'desert' genus with a similar wide distribution, *Pleuroxia*, appears not to be monophyletic and much more phylogenetically diverse than *Sinumelon*, with the five sampled species falling into three different lineages within clade 11. These correspond to the three regions Western Australia (Nullabor), South Australia (Flinders Ranges), Red Centre, with South Australian *Pleuroxia* species allied to *Sinumelon*; thus partially supporting the Solem (1997: 1893) conjecture on multiple origins of *Pleuroxia*. Another Red Centre genus *Basedowena* also appears not to be monophyletic.

The enigmatic Camaenidae CC1 (which appears capable of free-sealing) from far western Queensland belongs to clade 11 but appears not to be related to *Sinumelon per se*. Clade 11 also includes the genus *Amplirhagada*, the most speciose camaenid genus in Australia with 55 species (of which we have only two), all confined to the Kimberleys region. Solem (1997: 1884) reported that preliminary allozyme data for 14 Kimberleys genera gave very large divergences (too large to be useful for relationships), but did support a link amongst *Mouldingia*, *Ordtrachia*, *Westraltrachia*, *Exiligada*, and *Prototrachia*. As the first two genera are in the mtDNA phylogeny, this suggests that the latter three genera also will be associated with clade 11. Thus broadly, all these species and genera form a coherent supergroup (clade 11) that could be regarded as an extended Sinumeloninae, which dominates the western and central part of the continent. Sampling and resolution are too limited to test other ideas (Solem, 1992, 1997) on origins of western and southern members from a central Australian source. However, the complexity suggests a long history with potential for multiple interchanges.

Another western genus with high species diversity, *Rhagada*, is not clearly associated with any other lineage in the datasets (although it shows some affinity with clade 10). *Rhagada* (*s.s.*) traditionally has an extralimital distribution that includes Indonesia but Solem (1985, 1997: 1668–1672) questioned this and considered *Rhagada* to be endemic. This remains to be tested.

Considering that we have only 22 of 42 of Australian genera outside the East Coast, the results indicate that there is substantial phylogenetic diversity in western regions especially the Kimberleys and northwest Western Australia. For example there are 19 endemic genera in the Kimberleys (Solem, 1997: 1864–1895) compared to two to four for south-eastern Australia (Smith, 1992; Clark, 2005), and this pattern appears to hold for phylogenetic diversity too.

Large scale biogeographical patterns

This paper is mainly concerned with sampling, phylogenetic reconstruction, and taxonomic implications, and we only broadly outline a number of biogeographical aspects, which may be elaborated upon elsewhere.

Origins

Overall the 28S rDNA data is broadly consistent with the immigration hypothesis but cannot clearly adjudicate more detail. The Australasian taxa appear to be a regional offshoot of a larger camaenid (*s.l.*) and even larger Helicoidea biogeography. Limited resolution given by the 28S, limited South-East Asian and Melanesian sampling, and that the sampled Australasian lineages are monophyletic with respect to the greater camaenid (*s.l.*) diversity, means that the result is not definite evidence of the immigration hypothesis over the Gondwanan hypothesis. However, the relatively shallow 28S rDNA divergence, terminal position and diversity of Melanesian lineages are consistent with the (approximately Miocene era) immigration scenario. More speculatively, the Australian lineages may actually be nested amongst a larger group of deep lineages found across Melanesia and South-East Asia with further sampling to fill this gap (for example: *Rhagada*, papuinid and chloritid groups). In terms of large scale biogeographical scenarios, two of the Solem (1997) subfamilies are unhelpful (Pleurodontinae and Camaeninae) but there is support for Sinumeloninae. Hence the 'two-track' hypothesis of Camaeninae for eastern and Sinumeloninae for central and western regions (Solem, 1979b, 1992, 1997) can be reconciled with the phylogenetic data, being consistent with (1) the phylogenetic distinction between east and west, and (2) the high phylogenetic diversity seen in the northwest. Nevertheless, deep mtDNA divergences and biogeographical patterns suggest a long enough history, pre-dating much of the New Guinean landmasses (Hall, 2001), that it is possible that some near-northern diversity actually originated from Australia. Further sampling and more precise dating would be required to detect this.

Ancestral vs. endemic phylogenetic diversity

Accepting the broad immigration scenario poses several questions: (1) how old, (2) how much of the phylogenetic diversity is endemic, and (3) how is it related to the north–south diversity gradients?

The phylogeny incorporates several groups that span the mesothermal archipelago of eastern Austra-

lia (Nix & Kalma, 1972; Nix, 1982), providing temporal benchmarks that are consistent with mtDNA divergence estimates. Wet Tropics phylogeography in *Gnarosophia bellendenkerensis*, *Jacksonena delicata*, and *Noctepuna cerea* is essentially congruent with vertebrate taxa, and probably dates from the late Pliocene–early Pleistocene (Joseph, Moritz & Hugall, 1995; Schneider *et al*., 1998; Hugall *et al*., 2002). The mesic forest dependent *Sphaerospira* lineage (Hugall *et al*., 2003) and the 'leapfrog' disjunct distribution in *Rhynchotrochus–Papuexul* rainforest tree snails, span the foremost biogeographical barrier in eastern Australia (the Burdekin Gap). This break has been estimated to date from the mid Miocene (10–15 Mya) for several taxonomic groups (Joseph *et al*., 1993; Crisp *et al*., 1995; Hugall *et al*., 2003, 2008).

The mtDNA tree contains representatives of (at least) seven very deep lineages: (1) the hadroid group plus associated grade of chloritids (clades 1–7); (2) combined clades 8 and 9; (3) clade 11, the Sinumeloninae; (4 and 5) at least two Melanesian lineages containing papuinids (which may include *Chloritisannax banneri*); (6) the northern chloritid dominated clade 10; and (7) *Rhagada*. Five of these are found in Australia and either contain extralimital taxa or are separated by extralimital lineages. Therefore it is prudent to consider that an unknown but substantial proportion of the deeper phylogenetic diversity may be ancestral to the dispersal and diversification on the Australian continent. Nevertheless, within Australia there are numerous biogeographically coherent groups of deep divergence, suggesting that there has been considerable history of endemic diversification and extinction. Therefore a substantial proportion may pre-date the mid-late Miocene ecosystem turnover from mesic to xeric dominated landscape.

Although the distinction between these ancestral and endemic components can never be exactly known (and may be at different levels for different lineages), the pattern of relict lineages may provide temporal benchmarks. This is perhaps best seen in relicts associated with the hadroid group, the *Sphaerospira* lineage, the *Austrochloritis agamemnon* complex, and the basal lineage structures in clades 8 and 9. These suggest that at least several groups had diversified across a large continental scale during mesic eras and before major fragmentation and attenuation of mesic habitat owing to aridification, which may have begun as early as mid Miocene (~15 Mya) (Bowler, 1982; Adam, 1992; Alley, 1998; Mcgowran *et al*., 2004; Martin, 2006). McMichael (1968) and Megirian *et al*. (2004) have reported possible Miocene age camaenid fossils from the Lake Eyre Basin, consistent with a relatively early immigration and widespread dispersal – and therefore long endemic history spanning subsequent environmental changes. Further possible fossil sites (e.g. Riversleigh Archer *et al*., 1991; Frome Basin T. Worthy, pers. comm.) may prove informative.

There is evidence that sclerophyll and semi-arid environments in the central and western parts of the continent have a long history, dating from the Oligocene (Adam, 1992; Truswell, 1993; Alley, 1998; Martin, 2006). However, the distribution (and phylogenetic position) of the phylogenetically distinct *Chloritobadistes victoriae* (endemic to the south-eastern edge of Australia), and *Vidumelon watti*, *Aslintesta camelus*, the *Semotrachia* lineage, and Camaenidae BL7 (desert/arid zone endemics of deep rock gorge microhabitat) may be best explained as relicts of earlier eras when Australia was dominated by mesic forests. There is reasonable support for linking clade 8 with 9 but cladistic interpretation of the basal relationships may overlook missing (or extinct) lineages that are implied by the shape and biogeographical distribution of the phylogenetic diversity. This clade has the hallmarks of extinction as all members are long-branch, long stem-to-crown, lineages, with a widely disjunct distribution. Possibly all primary lineages are ancestral, at least to the mid-late Miocene shift from mesic to xeric environment at the continental scale. Potential missing taxa may be searched for amongst the as-yet unfathomed diversity within South-East Asia and Melanesia.

East–west, north–south patterns

There is a large scale phylogenetic distinction between the eastern and western halves of Australia with very little lineage overlap, with none of the western lineages present in the phylogenetically more diverse eastern region (with the exception of grassland *Xanthomelon pachystylum*, the most widely distributed species, and possibly some of the western chloritid *Damochlora, Parlogenia*, *Torresitrachia*). In the Red Centre and the Flinders Ranges (in South Australia), the eastern and the western lineages overlap. The Red Centre contains diverse genera from the Sinumeloninae (clade 11, e.g. *Sinumelon*, *Basedowena*) but also has relatives of eastern lineages: *Semotrachia* (clade 8), *Vidumelon* (allied to the hadroid group), and possibly more in the unsampled genera (e.g. *Dirutrachia*, *Divellomelon*). In the Finders Ranges (and adjacent Gammon, Gawler, and Barrier Ranges) again the Sinumeloninae overlaps with hadroid clade 3 and *Aslintesta* lineages. Either the western lineages were never present in the east or went extinct. Relicts suggest that some of the eastern lineages were more widespread but have since become attenuated. However, deep endemism of western lineages suggests that they have never been present in the east, perhaps because of a lack of suitable

environments (at least until recently) in the east. Hence the deep western endemism is consistent with a long history of xeric environments in the western part of the continent (Alley, 1998; McGowran *et al*., 2004; Martin, 2006).

Solem (1979a, 1997) considered that the north–south patterns of diversity argued against Gondwanan origins and better fitted immigration from the north of a basically tropical monsoonal adapted group. In addition to a general north–south diversity cline, camaenids are not present in southern Western Australia, or Tasmania, and have low diversity in Victoria: they have limited diversity in microthermal regimes, where the land snail fauna is dominated by Rhytididae and Caryodidae (and Bulimulidae) – suggested as likely Gondwanan groups (Solem, 1979a; Bishop, 1981; Wade *et al*., 2006). Camaenid species and lineage diversity in south-eastern Australia is low compared to most other regions. However, several of the lineages are not derived subsets of more northern groups but divergent endemic lineages (e.g. *Chloritobadistes victoriae* and the austrochloritid VC1-SN5 lineages), suggesting a complex history shaped by long-term environmental changes.

At the other end of the distribution, ongoing connection with New Guinea may contribute to the pattern of diversity seen in Cape York. Cape York contains several major lineages each with relatively low diversity; (e.g. *Chloritisanax banneri*, clade 7, clade 10). Cape York does appear to have at least some palaeoendemics (for example Camaenidae CY1); however it is plausible that some of the divergent taxa will have northern extralimital (New Guinea/Melanesia) relationships, indicative of interchange, as is the case with numerous other groups (for example vertebrates: Schodde & Calaby, 1972; Kikkawa *et al*., 1981 Aplin, Baverstock & Donnellan, 1993; Norman *et al*., 2002). For example, clade 7 shows a relatively close relationship of some Melanesian and Australian groups. Other northern taxa such as *Rhynchotrochus macgillivrayi* and the unsampled northern endemics *Noctepuna poiretiana* and *Meliobba shafferyi* may also share links with northern extralimital (e.g. New Guinea) lineages. If the camaenids have had a relatively long history in Australia (Miocene) compared to a relatively young history of large parts of (Montane) New Guinea (Pliocene; Hall, 2001), there may well have been interchange north–south and vice versa.

Given the depth of phylogenetic diversity and geographical scale of dispersal of multiple groups, environmental (climatic as all as biotic) factors must govern the southern limit of diversity in camaenids, whereas geographical and environmental proximity to Melanesian (probably vast) sources augments northern diversity.

Regional endemic diversification

Along eastern Australia, the vast majority of sister species diversity is allopatric, some parapatric. Although there are few if any sympatric sister species, in each region there is consistent emergence of endemic sympatric diversity, with each major clade contributing two to four sibling sympatric taxa, with higher levels of local diversity (up to 12–15) accumulated from amongst lineages. For example, for local (within 5 km radius) species diversity in the eastern region clade 1 contributes up to three species, with similar numbers for other clades: clade 2, two to three; clade 3, two to three, clade 4, two to three; clade 5.5, two; clades 5 and 6, two to three; clades 8 and 9, two to four. One locality in the Macleay Valley of north-eastern NSW may have a local radiation of up to six clade 9 *Austrochloritis* species (Stanisic *et al*., 2007). Broadly this suggests a dominant mode of speciation beginning with vicariant or peripatric isolation, with local diversity governed by accumulation and maintenance.

Further developments

Shells are seductive – easy to collect and keep, measure, and draw, thus generating substantial inventories of diversity. To this we add detailed molecular phylogenetic data to help develop a rational and biogeographically coherent framework, which can be readily elaborated upon. In the eastern camaenids, although many generic names are probably polyphyletic (and many at least paraphyletic), this can usually be broken down into a series of monophyletic clusters, splitting up the complexity into smaller sibling groups to which traditional morphological methods can more usefully be applied. Given the vast camaenid (*s.l.*) diversity yet to be analysed and taxonomic limitations implied by the molecular results, it may be better to place new data into the current matrix, rather than basing tests on pre-existing classifications. Such a matrix can be added to both for lineages and for genes. Ideally there should be as much overlap as possible in genes but this is not strictly necessary as long as there is sufficient density of data to link lineages. Assembling supermatrices and deciding just what constitutes sufficient data is a complex topic dealt with elsewhere (Sanderson & Driskell, 2003; Yan, Burleigh & Eulenstein, 2005; Ciccarelli, Doerks & von Mering, 2006) but measures of tree stability across data levels are a key guide. Our phylogeny provides an almost 'tree-of-life' complete description of lineage and species diversity across the entire eastern distribution from Torres to Bass Straits. Combined with fine scale distribution data (associated with the museum collections) this will enable quantitative biogeographical and macro-

ecological analyses incorporating both species and phylogenetic diversity (e.g. Faith, 2002; Rabosky, 2006; Davies *et al*., 2007; Weir & Schluter, 2007).

ACKNOWLEDGEMENTS

Stephanie Clark (UWS), Regie Kawamoto (Bishop Museum), Chris Wade (BMNH), Rei Ueshima (Tokyo), Winston Ponder (AM), Darryl Potter (QM), Rudiger Bieler (FMNH), Damien Broderick, Nancy Irwin, Michael Cunningham, Patrick FitzHenry (SAPAC), Adnan Moussalli, Craig Mortiz, Mike Lee, Jeremy Austin, Sonya Clegg, and R. A. Hugall.

REFERENCES

Adam P. 1992. *Australian rainforests*. Oxford: Clarendon Press.

AFD. 2007. *Australian fauna directory website*. Available at http://www.environment.gov.au/biodiversity/abrs/online-resources/fauna/afd/taxa/CAMAENIDAE

Alley NF. 1998. Cainozoic stratigraphy, palaeoenvironments and geological evolution of the Lake Eyre Basin. *Palaeogeography, Palaeoclimatology, Palaeoecology* **144:** 239–263.

Aplin KP, Baverstock PR, Donnellan SC. 1993. Albumin immunological evidence for the time and mode of origin of the New Guinean terrestrial mammal fauna. *Science in New Guinea* **19:** 131–145.

Archer M, Hand SJ, Godthelp H. 1991. *Riversleigh. The story of animals in ancient rainforest of inland Australia*. Balgowlah, NSW: Reed Books, 170–187.

Beck HH. 1837. *Index molluscorum praesentis aevi musei principis augustissimi christiani frederici*. Hafniae: [Beck], Pt 2 101–124.

Beiko RG, Kieth JM, Harlow TJ, Ragan MA. 2006. Searching for convergence in phylogenetic markov chain monte carlo. *Systematic Biology* **55:** 553–565.

Bininda-Emonds ORP. 2004. The evolution of supertrees. *Trends in Ecology and Evolution* **19:** 315–322.

Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A. 2007. The delayed rise of present-day mammals. *Nature* **446:** 507–512.

Bishop MJ. 1978. Anatomical notes on the Australian land snail *Sphaerospira fraseri* (Griffith & Pidgeon). *Journal of Molluscan Studies* **44:** 200–208.

Bishop MJ. 1981. The biogeography and evolution of Australian land snails. In: Keast A, ed. *Ecological biogeography of Australia*. The Hague: Dr W. Junk, 924–954.

Blackburn TM, Gaston KJ. 2004. *Macroecology: concepts and consequences 43rd symposium of the British Ecological Society*. Cambridge: Cambridge University Press.

Blouin C, Butt D, Roger AJ. 2005. Impact of taxon sampling on the estimation of rates of evolution at sites. *Molecular Biology and Evolution* **22:** 784–791.

Bowler JM. 1982. Aridity in the late Tertiary and Quaternary of Australia. In: Barker WR, Greenslade PJM, eds. *Evolution of the flora and fauna of arid Australia*. Norwood, SA: Peacock Publications, 35–45.

Burbidge NT. 1960. The phytogeography of the Australian region. *Australian Journal of Botany* **8:** 75–210.

Burch JB. 1976. Outline of classification of Australian terrestrial molluscs (native and introduced). *Journal of the Malacological Society of Australia* **3:** 127–156.

Burnham KP, Anderson DR. 2003. *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd edn. New York: Springer-Verlag.

Cain AJ. 1954. *Animal species and their evolution*. London: Hutchinson House.

Ciccarelli FD, Doerks T, von Mering C. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science* **311:** 1283–1287.

Clark SA. 2005. Systematics, spatial analysis and conservation genetics of *Meridolum corneovirens* (Pfeiffer, 1851) and related forms (Gastropoda: Camaenidae) from the Sydney region of Australia. PhD thesis, University of Western Sydney.

Clark SA. 2009. A review of the land snail genus *Meridolum* (Gastropoda: Camaenidae) from central New South Wales, Australia. *Molluscan Research* **29:** 61–120.

Clench WJ, Turner RD. 1966. Monograph of the genus *Rhynchotrochus* (Papuininae: Camaenindae). *Journal of the Malacological Society of Australia* **1:** 59–95.

Cook LG, Crisp MD. 2005. Not so ancient: the extant crown group of Nothofagus represents a post-Gondwanan radiation. *Proceedings of the Royal Society of London B* **272:** 2335–2544.

Cox JC. 1868. *A monograph of Australian lands shells*. Sydney: Maddock.

Cracraft J. 1986. Origin and evolution of continental biotas: speciation and historical congruence within the Australian avifauna. *Evolution* **40:** 977–996.

Crisp MD, Laffan S, Linder HP, Monro A. 2001. Endemism in the Australian Flora. *Journal of Biogeography* **28:** 183–198.

Crisp MD, Linder HP, Weston PH. 1995. Cladistic biogeography of plants in Australia and New Guinea: congruent pattern reveals two endemic tropical tracks. *Systematic Biology* **44:** 457–473.

Cuezzo MG. 2003. Phylogenetic analysis of the Camaenidae (Mollusca: Stylommatophora) with special emphasis on the American taxa. *Zoological Journal of the Linnean Society* **138:** 449–476.

Davies RG, Orme CD, Webster AJ, Jones KE, Blackburne TM, Gaston KJ. 2007. Environmental predictors of global parrot (Aves: Psittaciformes) species richness and phylogenetic diversity. *Global Ecology and Biogeography* **16:** 220–233.

Driskell AC, Ane CA, Burleigh JG, McMahon MM, O'Meara BC, Sanderson MJ. 2004. Prospects for building the tree of like from large sequence databases. *Science* **306:** 1172–1174.

Drummond AJ, Rambaut A. 2006. *BEAST v1.4*. Available at http://beast.bio.ed.ac.uk/Tracer

Dubois A. 2007. Phylogeny, taxonomy and nomenclature: the problem of taxonomic categories and of nomenclatural ranks. *Zootaxa* **1519:** 27–68.

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sørensen MV, Haddock SHD, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452:** 745–750.

Erixon P, Svennblad B, Britton T, Oxelman B. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Systematic Biology.* **52:** 665–673.

Faith DP. 2002. Quantifying biodiversity: a phylogenetic perspective. *Conservation Biology* **16:** 248–252.

Farrell BD. 1998. 'Inordinate fondness" explained: Why are there so many beetles? *Science* **281:** 555–559.

Felsenstein J. 2004. *Inferring phylogenies*. Sunderland, MA: Sinauer.

Férussac AEJPJF d'A de. 1819. *Histoire Naturelle Générale et Particulière des Mollusques Terrestres et Fluviatiles*. Paris Livr. 1–5 pp. 1–96 [34] [pl. 28, figs. 3–4].

Fujioka T, Chappell J, Honda M, Yatsevich I, Fifield K, Fabel D. 2005. Global cooling initiated stony deserts in central Australia 2–4 Ma, dated by cosmogenic 21Ne-10Be. *Geology* **33:** 993–996.

Fulton HC. 1907. Descriptions of new species of Australian *Planispira* and *Chloritis*. *Proceedings of the Malacological Society of London* **7:** 362–363.

Gatesy J, Matthee C, DeSalle R, Hayashi C. 2002. Resolution of a supertree/supermatrix paradox. *Systematic Biology* **51:** 652–664.

Godfray HCJ, Knapp S. 2004. Taxonomy for the twenty-first century – introduction. *Philosophical Transactions of the Royal Society of London* **359:** 559–569.

Goodfriend GA. 1986. Variation in land-snail shell form and size and its causes: a review. *Systematic Biology* **35:** 204–233.

Griffith E, Pidgeon E. 1833. The Mollusca and Radiata. In: Cuvier GLCFD, ed. *The Animal Kingdom*. London: Whittaker & Co., Vol. 12 601 pp. 40 pls.

Gude GK. 1906. Further remarks on the genus Chloritis, with descriptions of eleven new species. *Proceedings of the Malacological Society of London* **7:** 105–118.

Gude GK. 1907. A further contribution to our knowledge of the genus *Chloritis*, with descriptions of eleven new species. *Proceedings of the Malacological Society of London* **7:** 228–233.

Hall R. 2001. Cenozoic reconstructions of SE Asia and the SW Pacific: changing patterns of land and sea. In: Metcalfe I, Smith JMB, Morwood M, Davidson I, eds. *Faunal and floral migrations and evolution in SE Asia-Australasia*. Lisse: A.A. Balkema Publishers, 35–56.

Heard SB, Cox GH. 2007. The shapes of phylogenetic trees of clades, faunas, and local assemblages: exploring spatial pattern in differential diversification. *The American Naturalist* **169:** E107–E118.

Hedke SM, Townsend T, Hillis DM. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Systematic Biology* **56:** 522–529.

Hedley C. 1912. On some land shells collected in Queensland, by Mr Sidney W. Jackson. *Proceedings of the Linnean Society of New South Wales* **37:** 253–270.

Hillis DM. 1996. Inferring complex phylogenies. *Nature* **383:** 130–131.

Hooker JD. 1859. Introductory essay. In: *The Botany of the Antarctic Voyage of H.M. discovery ships Erebus and Terror in the years 1839–1843: vol. 3, Flora Tasmaniae*. London: Reeve, i–cxviii.

Hoskin CJ, Couper PJ. 2004. A new species of *Glaphyromorphus* (Reptilia: Scincidae) from Mt Elliot, north-eastern Queensland. *Australian Journal of Zoology* **52:** 183–190.

Huelsenbeck JP, Rannala B. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology* **53:** 904–913.

Hugall AF, Foster R, Hutchinson M, Lee MSY. 2008. Phylogeny of Australasian agamid lizards based on nuclear and mitochondrial genes: implications for morphological evolution and biogeography. *Biological Journal of the Linnean Society* **93:** 343–358.

Hugall AF, Lee MSY. 2007. The likelihood node density effect and how it confounds evolutionary studies of molecular rates. *Evolution* **61:** 2293–2307.

Hugall AF, Moritz C, Moussalli A, Stanisic J. 2002. Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences, USA* **99:** 6112–6117.

Hugall AF, Stanisic J, Moritz C. 2003. The case of the Sphaerospira lineage and history of Queensland rainforests. In: Lydeard C, Lindberg D, eds. *Molecular systematics and phylogeography of mollusks*. Washington, DC: Smithsonian Institution Press, 270–301.

Iredale T. 1933. Systematic notes on Australian land shells. *Records of the Australian Museum* **19:** 37–59.

Iredale T. 1937a. A basic list of the land Mollusca of Australia. *Australian Zoologist* **8:** 287–333.

Iredale T. 1937b. A basic list of the land Mollusca of Australia – Part II. *Australian Zoologist* **9:** 1–39.

Iredale T. 1938. A basic list of the land Mollusca of Australia – Part III. *Australian Zoologist* **9:** 83–124.

Joseph L, Moritz C. 1994. Mitochondrial DNA phylogeography of birds in eastern Australian rainforests: First fragments. *Australian Journal of Zoology* **42:** 385–403.

Joseph L, Moritz C, Hugall A. 1993. A mitochondrial perspective on the historical biogeography of mideastern Queensland rainforest birds. *Memoirs of the Queensland Museum* **34:** 201–214.

Joseph L, Moritz C, Hugall A. 1995. Molecular support for vicariance as a source of diversity in rainforest. *Proceedings of the Royal Society of London B* **260:** 177–182.

Kikkawa J, Monteith GB, Ingram G. 1981. Cape York

Peninsula: major region of faunal interchange. In: Keast JA, ed. *Ecological biogeography of Australia*. The Hague: W. Junk, 1695–1742.

Lee MSY, Hugall AF. 2003. Partition likelihood support and the evaluation of data set conflict. *Systematic Biology* **52:** 15–22.

Lee MSY, Hugall AF. 2006. Model type, implicit data weighting, and model averaging in phylogenetics. *Molecular Phylogenetics and Evolution* **38:** 848–857.

Lemmon AR, Brown JM, Stanger-Hall K, Lemmon EM. 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Sysematic Biology* **58:** 130–145.

Lewis PO, Holder MT, Holsinger KE. 2005. Polytomies and Bayesian phylogenetic inference. *Systematic Biology* **54:** 241–253.

Linnaeus C. 1758. *Systema Naturae, per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Editio decima, reformata*. Holmiae: Laurentii Salvii Tom. 1 824 pp.

Low T. 1993. Last of the rainforests: rainforest refuges of the Mackay region. *Wildlife Australia* **4:** 18–21.

Maddison WP. 1997. Gene trees in species trees. *Systematic Biology* **46:** 523–536.

Maekawa K, Lo N, Rose HA, Matsumoto T. 2003. The evolution of soil-burrowing cockroaches (Blattaria: Blaberidae) from wood-burrowing ancestors following an invasion of the latter from Asia into Australia. *Proceedings of the Royal Society of London B* **270:** 1301–1307.

Manos PS, Soltis PS, Soltis DE, Manchester SR, Oh SH, Bell CD, Dilcher DL, Stone DE. 2007. Phylogeny of extant and fossil Juglandaceae inferred from the integration of molecular and morphological data sets. *Systematic Biology* **56:** 412–430.

Martin HA. 2006. Cenozoic climatic change and the development of the arid vegetation in Australia. *Journal of Arid Environments* **66:** 533–563.

McGowran B, Holdgate GR, Li Q, Gallagher SJ. 2004. Cenozoic stratigraphic succession in southeastern Australia. *Australian Journal of Earth Sciences* **51:** 459–496.

McMichael DF. 1968. Non-marine Mollusca from Tertiary rocks in northern Australia. *Bureau of Mineral Resources Bulletin* **80:** 133–159.

Megirian D, Murray P, Schwartz L, Von Der Borch C. 2004. Late Oligocene Kangaroo Well Local Fauna from the Ulta Limestone (new name), and climate of the Miocene oscillation across central Australia. *Australian Journal of Earth Sciences* **51:** 701–741.

Morando M, Avila LJ, Sites JW Jr. 2003. Sampling strategies for delimiting species: genes, individuals, and populations in the *Liolaemus elongata-kriegi* complex (Squamata: Liolaemidae) in Andean-Patagonian South America. *Systematic Biology* **52:** 159–185.

Mörch OAL. 1867. Abrégé de l'histoire de la classification moderne des mollusques basée principalement sur l'armature linguale. *Journal de Conchology* **15:** 232–258.

Moritz C, Hoskin C, Graham CH, Hugall A, Moussalli A. 2004. Historical biogeography, diversity and conservation of Australia's tropical rainforest herpetofauna. In: Gittleman JL, Brooks TM, Purvis A, eds. *Phylogeny and conservation*. Cambridge: Zoological Society of London, Cambridge University Press, 243–266.

Moritz C, Richardson KS, Ferrier S, Monteith GB, Stanisic J, Williams SE, Whiffin T. 2002. Biogeographical concordance and efficiency of taxon indicators for establishing conservation priority in a tropical rainforest biota. *Proceedings of the Royal Society of London B* **268:** 1875–1881.

Nix HA. 1982. Environmental determinates of biogeography and evolution in Terra Australis. In: Barker WR, Greenslade PJM, eds. *Evolution of the flora and fauna of arid Australia*. Frewville, SA: Peacock Publications, 47–66.

Nix HA, Kalma JD. 1972. Climate as the dominant control in the biogeography of northern Australia and New Guinea. In: Walker D, ed. *Bridge and barrier: the natural and cultural history of Torres Strait*. Canberra: Australian National University, 61–92.

Norman JA, Christidis L, Joseph L, Slikas B, Alpers D. 2002. Unravelling a biogeographical knot: origin of the 'leapfrog' distribution pattern of Australo-Papuan sooty owls (Strigiformes) and logrunners (Passeriformes). *Proceedings of the Royal Society of London. Series B* **269:** 2127–2133.

Oliver PM, Sanders KL. 2009. Molecular evidence for Gondwanan origins of multiple lineages within a diverse Australasian gecko radiation. *Journal of Biogeography* **36:** 2044–2055.

Philippe H, Snell EA, Bapteste E, Lopez P, Holland PWH, Casane D. 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Molecular Biology and Evolution* **21:** 1740–1752.

Pilsbry HA. 1890. Vols 3–8. In: Tryon GW, Pilsbry HA, eds. *Manual of conchology; structural and systematic*. Philadelphia, PA: Academy of Natural Sciences.

Pirie MD, Humphreys MA, Galley C, Barker NP, Verboom GA, Orlovich D, Draffin SJ, Lloyd K, Baeza CM, Negritto M, Ruiz E, Sanchez JHC, Reimer E, Linder HP. 2008. A novel supermatrix approach improves resolution of phylogenetic relationships in a comprehensive sample of danthonioid grasses. *Molecular Phylogenetics and Evolution* **48:** 1106–1119.

Ponder W, Lunney D, eds. 1999. *The other 99%. The conservation and biodiversity of invertebrates*. Mosman: Transactions of the Royal Zoological Society of New South Wales.

Purvis A. 1996. Using interspecies phylogenies to test macroevolutionary hypotheses. In: Harvey PH, Brown AJL, Maynard Smith J, Nee S, eds. *New uses for new phylogenies*. Oxford: Oxford University Press, 153–168.

de Queiroz A, Gatesy J. 2007. The supermatrix approach to systematics. *Trends in Ecology and Evolution* **22:** 34–41.

Rabosky D. 2006. Likelihood methods for temporal detecting shifts in diversification rates. *Evolution* **60:** 1152–1164.

Reeve LA. 1852. Monograph of the genus *Helix*. In: Reeve LA, ed. *Conchologia iconica*, Vol. 7. London: L. Reeve & Co., pls 63–132, 134–146.

Ren F, Tanaka H, Yang Z. 2008. A likelihood look at the supermatrix–supertree controversy. *Gene* **441:** 119–125.

Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572–1574.

Sanders K, Lee MSY. 2007. Molecular evidence for a rapid late Miocene radiation of Australasian venomous snakes. *Molecular Phylogenetics and Evolution* **46:** 1165–1173.

Sanderson MJ, Driskell AC. 2003. The challenge of constructing large phylogenetic trees. *Trends in Plant Science* **8:** 374–379.

Schneider CJ, Cunningham M, Moritz C. 1998. Comparative phylogeography and the history of endemic vertebrates in the Wet Tropics rainforests of Australia. *Molecular Ecology* **7:** 487–498.

Schodde R, Calaby JH. 1972. The biogeography of the Australo-papuan bird and mammal faunas in relation to Torres Strait. In: Walker D, ed. *Bridge and barrier: the natural and cultural history of Torres Strait*. Canberra: Australian National University, 257–306.

Scott B. 1996a. Systematics, phylogeny and biogeography of the larger camaenid land snails of eastern Queensland rainforests: (Pulmonata: Stylommatophora: Camaenidae). PhD thesis, James Cook University Townsville.

Scott B. 1996b. Phylogenetic relationships of the Camaenidae (Pulmonata: Stylommatophora: Helicoidea). *Journal of Molluscan Studies* **62:** 65–73.

Scott B. 1997. Biogeography of the Helicoidea (Mollusca: Gastropoda: Pulmonata): land snails with Pangean distribution. *Journal of Biogeography* **24:** 399–407.

Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16:** 1114–1116.

Smith BJ. 1984. Regional endemism of the south-eastern Australian land mollusc fauna. In: Solem A, Van Bruggen AC, eds. *World-wide snails. Biogeographical studies on non-marine Mollusca*. Leiden: E.J. Brill/Dr W. Backhus, 178–188.

Smith BJ. 1992. Non-marine mollusca. In: Houston WWK, ed. *Zoological catalogue of Australia*, Vol. 8 xii. Canberra: AGPS, pp. 110–174.

Solem A. 1979a. Land-snail biogeography: A true snail's pace of change. In: Nelson G, Rosen DE, eds. *Vicariance biogeography a critique*. New York: Columbia University Press, 197–221.

Solem A. 1979b. Camaenid land snails from Western and Central Australia (Mollusca: Pulmonata: Camaenidae) I Taxa with trans-Australian distributions. *Records of the Western Australian Museum Supplement* **10:** 5–142.

Solem A. 1981a. Camaenid land snails from Western and central Australia. II. Taxa from the Kimberley, Amplirhagada Iredale, 1933. *Records of the Western Australian Museum, Supplement* **11:** 147–320.

Solem A. 1981b. Camaenid land snails from Western and Central Australia (Mollusca: Pulmonata: Camaenidae). III. Taxa from the Ningbing Ranges and nearby areas. *Records of the Western Australian Museum, Supplement* **11:** 321–425.

Solem A. 1984. Camaenid land snails from Western and Central Australia (Mollusca: Pulmonata: Camaenidae). IV. Taxa from the Kimberley, Westraltrachia Iredale, 1933 and related genera. *Records of the Western Australian Museum, Supplement* **17:** 427–705.

Solem A. 1985. Camaenid land snails from Western and central Australia (Mollusca: Pulmonata: Camaenidae). V. Remaining Kimberley genera and addenda to the Kimberley. *Records of the Western Australian Museum, Supplement* **20:** 707–981.

Solem A. 1992. Camaenid land snails from southern and eastern South Australia, excluding Kangaroo Island. *Records of the South Australia Museum*, Monograph Series **2:** 1–425.

Solem A. 1993. Camaenid land snails from Western and central Australia (Mollusca: Pulmonata: Camaenidae). VI. Taxa from the Red Centre. *Records of the Western Australian Museum, Supplement* **43:** 983–1459.

Solem A. 1997. Camaenid land snails from Western and Central Australia (Mollusca: Pulmonata: Camaenidae) VII taxa from dampierland through the nullabor *Records of the Western Australian Museum, Supplement* **50:** 5–142.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.

Stanisic J. 1994. The distribution and patterns of species diversity of land snails in eastern Australia. *Memoirs of the Queensland Museum* **36:** 207–213.

Stanisic J. 1996. The large camaenids of Mideastern Queensland. *Australasian Shell News* **92:** 1–2.

Stanisic J. 1998. *Temporena* Iredale unraveled. *Memoirs of the Queensland Museum* **42:** 604.

Stanisic J. 2001. Taxonomy of the Australian rainforest snail, *Helix bellendenkerensis* Brazier 1875 (Mollusca: Eupulmonata: Camaenidae). *Memoirs of the Queensland Museum* **46:** 337–348.

Stanisic J. 2009. *Crikey steveirwini* gen. et sp. nov. from montane habitats in the Wet Tropics of northeastern Queensland, Australia (Gastropoda: Eupulmonata: Camaenidae). *Zootaxa* **2206:** 62–68.

Stanisic J, Cameron RAD, Pokryszko BM, Nekola JC. 2007. Forest snail faunas from S.E. Queensland and N.E. New South Wales (Australia): patterns of local and regional richness and differentiation. *Malacologia* **49:** 445–462.

Sullivan J, Swofford DL, Naylor GJP. 1999. The effect of taxon sampling on estimated rate heterogeneity parameters of maximum likelihood models. *Molecular Biology and Evolution* **16:** 1347–1356.

Swofford DL. 2000. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4 (beta). Computer program. Sunderland MA: Sinauer Associates.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24:** 4876–4882.

Truswell E. 1993. Vegetation changes in the Australian Tertiary in response to climatic and phytogeographic forcing factors. *Australian Systematic Botany* **6:** 533–557.

Wade CM, Hudelot C, Davison A, Naggs F, Mordan PB. 2007. Molecular phylogeny of the helicoid land snails (Pulmonata: Stylommatophora: Helicoidea), with special emphasis on the Camaenidae. *Journal of Molluscan Studies* **73:** 411–415.

Wade CM, Mordan PB, Clarke B. 2001. A phylogeny of the land snails (Gastropoda: Pulmonata). *Proceedings of the Royal Society of London* **268:** 413–422.

Wade CM, Mordan PB, Naggs F. 2006. Evolutionary relationships among the pulmonate land snails and slugs (Pulmonata, Stylommatophora). *Biological Journal of the Linnean Society* **87:** 593–610.

Weir JT, Schluter D. 2007. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science* **315:** 1574–1576.

Wheeler QD, Meier R, eds. 2000. *Species concepts and phylogenetic theory: a debate*. New York: Columbia University Press.

Wiens J. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. *Systematic Biology* **52:** 528–538.

Wiens JJ. 2006. Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics* **39:** 34–42.

Wiens JJ, Fetzner JW Jr, Parkinson CL, Reeder T. 2005. Hylid frog phylogeny and sampling strategies for speciose clades. *Systematic Biology* **54:** 719–748.

Wiens JJ, Graham CH, Moen DS, Smith SA, Reeder TD. 2006. Evolutionary and ecological causes of latitudinal diversity gradient in Hylid frogs: Treefrog trees unearth the roots of high tropical diversity. *The American Naturalist* **168:** 579–596.

Wiens JJ, Moen DS. 2008. Missing data and the accuracy of Bayesian phylogenetics. *Journal of Systematics and Evolution* **46:** 307–314.

Wilgenbusch JC, Warren DL, Swofford DL. 2004. *AWTY: a system for graphical exploration of MCMC convergence in bayesian phylogenetic inference*. Available at http://ceb.csit.fsu.edu/awty

Williams SE, Pearson RG, Walsh PJ. 1996. Wet tropics biodiversity. *Pacific Conservation Biology* **2:** 327–362.

Woodruff DS, Solem A. 1990. Allozyme variation in the Australian camaenid land snail *Cristilabrum primum*: a prolegomenon for a molecular phylogeny of an extraordinary radiation in an isolated habitat. *Veliger* **33:** 129–139.

Yan CH, Burleigh JG, Eulenstein O. 2005. Identifying optimal incomplete phylogenetic data sets from sequence databases. *Molecular Phylogenetics and Evolution* **35:** 528–535.

Yokoyama Y, Purcell A, Lambeck K, Johnston P. 2001. Shore-line reconstruction around Australia during the Last Glacial Maximum and Late Glacial Stage. *Quaternary International* **83:** 9–18.

Zhang W, Shea M. 2008. A new genus and species of land snail of the family Camaenidae from New South Wales. *Molluscan Research* **28:** 123–132.

Zwickl DJ, Hillis DM. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology* **51:** 588–598.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1. Maximum likelihood (ML) and Markov chain Monte Carlo (MCMC) 28S rDNA phylogenies. The 28S rDNA dataset was subjected to (A) Bayesian analysis using MrBayes v.3.1.2 (Ronquist and Huelsenbeck, 2003), (B) ML analysis using RAxML v.7.0.4 (Stamatakis, 2006); in both cases applying a GTR-G model. MrBayes used two runs by 5 million by four chains with standard heating, 1/500 sampling, 20% burnin, leaving a combined total of 16 000 samples for the majority rule posterior consensus. RAxML used the -f a function with 1000 bootstraps to provide a pool of trees for the final ML search. Results of a PAUP 4.10b ML heuristic search using an optimized GTR-inv are also shown (C).

Figure S2. All compatible consensus trees with support values of the 62 taxa that have all three genes, drawn from each taxa set analysis using the 3p model. A, 62t Markov chain Monte Carlo (MCMC) analysis; B, pruned from 147t MCMC analysis; C, pruned from the 327t MCMC analysis; D, RAxML fast bootstrap. Posterior probabilities calculated from pruned MCMC samples.

Figure S3. All compatible consensus trees with support values of the 147 taxa set, using the 3p model. A, 147t Markov chain Monte Carlo (MCMC) analysis; B, pruned from the 327t MCMC analysis; C, RAxML fast bootstrap. Posterior probabilities calculated from pruned MCMC samples.

Figure S4. Consensus trees for 62 taxa amongst Markov chain Monte Carlo (MCMC) analyses across models and taxasets. A, strict consensus amongst all compatible consensus trees for three models by three taxa sets (nine analyses). The next three trees are 95% posterior probability (PP) consensus of the 62 taxa set in MCMC analyses where the results of the three different partition models have been combined. B, 62t analyses; C, pruned from the 147t analyses; D, pruned from the 327t analyses. Nodes that are incompatible at the ≥ 0.95 PP level indicated by asterisk (*).

Figure S5. Consensuses for 147 taxa amongst Markov chain Monte Carlo (MCMC) analyses across models and taxa sets. A, strict consensus of three models $\times$ two taxa sets (six analyses); B, 95% consensus from 147t analysis

(three models combined); C, 95% posterior probability (PP) consensus from pruned 327t analysis (three models combined). Asterisk (*) indicates node where ≥ 0.95 PP support is influenced by presence of absence of taxon *Aslintesta camelus*.
Figure S6. Branch length comparisons amongst partition models.
Figure S7. Markov chain Monte Carlo (MCMC) distribution overlap amongst taxa set analyses as measured by the lnL score using the optimal 62t single partition GTR-G-inv model. The lnL score is calculated using both the topology and branch length of each MCMC sample, pruned to: A, the 62 taxa; B, 147 taxa.
Figure S8. Markov chain Monte Carlo posterior tree length (TL) parameter distributions.
Table S1. Discrepancies amongst Zoological Catalogue (AFD, 2007), museum, and genetic databases.
Table S2. Topological distances amongst posterior all compatible consensus trees across models and taxa sets.
Table S3. Shimodiara–Hasgawa tests among Bayesian and fast maximum likelihood trees.
Table S4. Branch length comparisons amongst taxa sets and models.
Appendix S1. Molecular laboratory procedures and PCR primers.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Supplementary Files

Appendix S1: Molecular laboratory procedures and PCR primers.

Table S1. Discrepancies among Zoological Catalogue (AFD 2007), museum and genetic databases.

Table S2. Topological distances among posterior all compatible consensus trees across models and taxasets.

Table S3. Shimodiara-Hasgawa tests among Bayesian and fast ML trees.

Table S4. Branch length comparisons among taxasets and models.

Figure S1. ML and MCMC 28SrDNA phylogenies. The 28SrDNA dataset was subjected to A) Bayesian analysis using MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003), B) ML analysis using RAxML v7.0.4 (Stamatakis 2006); in both cases applying a GTR-G model. MrBayes used two runs by 5 million by four chains with standard heating, 1/500 sampling, 20% burnin, leaving a combined total of 16,000 samples for the majority rule posterior consensus. RAxML used the -f a function with 1000 bootstraps to provide a pool of trees for the final ML search. Results of a PAUP 4.10b ML heuristic search using an optimized GTR-pinv are also shown (C).

Figure S2. All compatible consensus trees with support values of the 62 taxa that have all three genes, drawn from each taxaset analysis using the 3p model. **A** 62t MCMC analysis; **B** Pruned from 147t MCMC analysis; **C** Pruned from the 327t MCMC analysis; **D** RAxML fast bootstrap. Posterior probabilities calculated from pruned MCMC samples.

Figure S3. All compatible consensus trees with support values of the 147 taxa set, using the 3p model. **A** 147t MCMC analysis; **B** Pruned from the 327t MCMC analysis; **C** RAxML fast bootstrap. Posterior probabilities calculated from pruned MCMC samples.

Figure S4. Consensus trees for 62 taxa among MCMC analyses across models and taxasets. **A** Strict consensus among all compatible consensus trees for three models by three taxasets (9 analyses). The next three trees are 95% posterior probability (PP) consensus of the 62 taxa set in MCMC analyses where the results of the three different partition models have been combined. **B** 62t analyses; **C** pruned from the 147t analyses; **D** pruned from the 327t analyses. Nodes that are incompatible at the ≥0.95 PP level indicated by asterisk (*).

Figure S5. Consenses for 147 taxa among MCMC analyses across models and taxasets. **A** Strict consensus of three models x two taxasets (6 analyses); **B** 95% consensus from 147t analysis (3 models combined); **C** 95% PP consensus from pruned 327t analysis (3 models combined). Asterisk (*) indicates node where ≥0.95 PP support is influenced by presence of absence of taxon (*Aslintesta camelus*).

Figure S6. Branch length comparisons among partition models.

Figure S7. MCMC distribution overlap among taxaset analyses as measured by the lnL score using the optimal 62t single partition GTR-G-inv model. The lnL score is calculated using both the topology and branch length of each MCMC sample, pruned to: **A** the 62 taxa; **B** 147 taxa

Figure S8. MCMC posterior tree length (TL) parameter distributions.

Supplementary Figure S4A shows the nodes in the 62t analysis that have a posterior probability level of ≥0.95, combining the MCMC results (the .trprobs files) of the three different models (the three different partition strategies). Therefore support needs to be high in all three for the combined PP to be ≥0.95 (at worst ≥0.85 in one analysis if the other two are 1.00). Figures S4A and S4B are the same but for the 147t and 327t taxaset analyses pruned down to the common 62 taxa.

Among analyses, for matching taxasets pruned from larger taxasets, there was only one case of contradictory ≥0.95 PP (indicated on figures) but several nodes that were ≥0.95 in one analysis were <0.95 in others. For the 62 taxa comparisons there was one conflict between the 62t and 147t analyses - the position of *Pommerhelix duralensis.* For the 147 taxa comparisons there were no contradictory PP but there was fluctuating support for a key node of interest: including *Aslintesta camelus* this node has 0.89 PP but ignoring it, the support rises to 0.95 PP.

Supplementary files: Molecular Laboratory Procedures

PCR primers

Cytochrome Oxidase subunit II (COII) designed from GenBank data, giving a fragment of 559bp in Camaenid land snails.

L strand (coding) COIILA:
5' AAA TAA TGC TAT TTC ATG AYC AYG C 3'
H strand (non-coding) COIIHQ:
5' GCT CCG CAA ATC TCT GAR CAY TG 3'

Note the degenerate positions corresponding to 3rd codon positions.

16SrDNA primers based on the standard Palumbi 16S primers (Pp 205-247 in Hillis, Moritz and Mable Eds. Molecular Systematics, Sinauer), modified to fit Camaenids based on GenBank full mtDNA sequences and some Japanese Camaenids (Chiba, Evolution 1999 vol. 53: 460-471). Gives fragment of about 440bp. See also Shimizu and Ueshima 2000, Heredity vol. 85: 84-96.

L strand 16SAS:
5' CGA CTG TTT ATC AAA AAC ATA G 3'
H strand 16SBS:
CCG GTC TGA ACT CAG ATC ATG T 3'

12SrDNA primers designed to GenBank and some Japanese sequences (Chiba, Evolution 1999 vol. 53: 460-471). Amplify a fragment of about 500bp:

L strand PE12SF:
5' GTG CCA GCA GTC GCG GTC 3'
H strand PE12SB:
5' AGA GTG ACG GGC GAT TTG T 3'

Primer employed in 28SrDNA amplification and sequencing (from Jamieson et al. 2002. Zoosytema vol. 24: 707-734):

Primer 3' position	Sequence 5' to 3'	Gene
C1'	ACCCGCTGAATTTAAGCAT	28SrRNA 'coding'
D2	TCCGTGTTTCAAGACGG	28SrRNA
C2	TGAACTCTCTCTTCAAAGTTCTTTTC	28S rRNA
C2'	GAAAAGAACTTTGRARAGAGAGT	28S rRNA

DNA extraction

Most samples used a simple Chelex (Bio-Rad) extraction method. Chelex is a metal ion chelating resin and is used in conjunction with proteinase *k* to free the DNA from the tissue sample. Chelex does not try to purify the DNA, just release it to become available as template for PCR, but has the benefit of minimising the amount of work and manipulation and hence minimise cross contamination. Tissue from alcohol stored specimens is first soaked in STE solution for one hour (STE = 30mM Tris-HCl pH 8.5, 30mM NaCl, 3mM EDTA. Tissue is then added to the Chelex mixture consisting of 0.5 ml of 5-10% Chelex in pure water with 5-15 μl of 10mg/ml proteinase *k* depending the amount of tissue, and incubated at 50-55 ºC for a several hours with occasional remixing. The sample is then heated to 90 ºC for 10-20 minutes to inactivate any remaining proteinase *k* activity.

PCR and Sequencing

Between 2-5 μl of the chelex reaction is then used as template in PCR. For some old samples it may be necessary to do more than one round of PCR: the PCR is first run for 30 cycles, then more PCR mixture of primers, dNTPs and Taq DNA polymerase is added and PCR continued for another 25 cycles. The PCR is then checked by agarose gel. If the desired DNA fragment is visible but weak, and/or there is a lot of non-specific PCR product, a piece of the relevant band is cut out of the gel (a few microlitres), diluted with 40-100 μl of water and heated to 80-90 ºC to melt the gel. Then 1-3 μl of this is used as template in a new PCR of 25 cycles.

PCR template for sequencing is first gel purified, removing primer, buffer and nonspecific amplification products in one go. Twenty microlitres of PCR reaction is loaded onto a Tris-Acetate (TA) buffer gel of about 1.2-1.6% agarose. TA is a benign buffer for the subsequent PCR reaction: no EDTA or borate, as there is in TBE (stock 40xTA = 1.6M Tris and 0.8 M Na Acetate pH 8.0). The gel is stained with ethidium bromide and the band cut out using a long wavelength UV light source to see the DNA, giving a piece of gel about 15-40 microlitres in volume. This is placed into a 1.5ml eppendorf tube and frozen for least 20 minutes at -80 ºC. The eppendorf is then centrifuged at full speed in a microfuge (10k rpm) for 5-10 minutes and the free liquid pipetted off. This should contain enough DNA to sequence several times, after checking the concentration on a gel. Between 1-6 μl is used per ABI fluorescent labelled cycling sequencing reaction.

Table S2. Topological distances among Bayesian posterior consensus trees

Maximum

	62t	147t	327t
62t	3		
147t	8	2	
327t	10	9	9

	147t	327t
147t	8	
327t	22	17

Average

	62t	147t	327t
62t	1.5		
147t	8.0	0.7	
327t	9.1	5.8	6.7

	147t	327t
147t	5.3	
327t	16.9	13.7

Symmetric-difference distances/2 = nodes collapsed
Combination of three partition models by three taxaset analyses
Top pruned to 62t, bottom pruned to 147t

MCMC overlap as measured by summary lnL

	62t	147t	327t
62t	0.69		
147t	0.15	0.73	
327t	0.03	0.18	0.65

	147t	327t
147t	0.78	
327t	0.23	0.65

Based on 2000 MCMC samples
Top pruned to 62t, bottom pruned to 147t
lnL calculated in PAUP using combined data GTR-G-inv model
using both MCMC topology and branch length

Table S1. Discrepancies among AFD/Zoological catalogue and Museum databases for east coast taxa

Australian Faunal Directory (AFD) web site: www.environment.gov.au/biodiversity/abrs/online-resources/fauna/adf/index.html)
Based on the Zoological Catalogue (Smith 1992, with revisions 2002)
Museum databases: Queenslan Museum, Australian Museum, Museum of Victoria

AFD/Zoological Catalogue taxa missing

genus	species	author	year
Galadistes	alleni	Iredale	1943
Galadistes	bourkensis	Smith	1891
Galadistes	intervenens	Iredale	1938
Meliobba	shafferyi	Iredale	1940
Meridolum	exocarpi	Cox	1868
Noctepuna	poiretiana	Reeve	1852
Posorites	fucata	Pfeiffer	1853
Posorites	turneri	Shirley	1921
Neveritis	interna	Iredale	1943
Galadistes	marcescens	Cox	1868
Galadistes	stutchburyi	Pfeiffer	1857

AFD/Zoological Catalogue taxa not used in museum databases

genus	species	author	year
Galadistes	stutchburyi	Pfeiffer	1857
Sphaerospira	bala	Brazier	1878
Sphaerospira	challisi	Cox	1873
Sphaerospira	tomsoni	Brazier	1876
Spurlingia	darwini	Brazier	1872
Ventopelita	yatalaensis	Cox	1873
Xanthomelon	distractum	Iredale	1938
Xanthomelon	magnidicum	Iredale	1938
Xanthomelon	saginatum	Iredale	1938
Sphaerospira	mitifica	Iredale	1933
Trachiopsis	tuckeri	Pfeiffer	1846

Species names used in Museum database that are not used in AFD/Zoological Catalogue

genus	species
Bentosites	findera
Bentosites	starena
Figuladra	aureedensis
Figuladra	curtisiana
Figuladra	pallida
Figuladra	praetermissi
Sphaerospira	mulgravensis
Sphaerospira	sidneyi
Spurlingia	excellens
Spurlingia	portus
Spurlingia	tinarooensis
Thersites	darlingtoni

AFD/Zoological Catalogue taxa referred to as belonging with tip lineages

genus	species	author	year	327t_OTU
Discomelon	intricatum	Iredale	1938	Camaenidae_BL12
Hadra	semicastanea	Pfeiffer	1849	Hadra_bipartita_Cow
Sphaerospira	bayensis	Brazier	1875	Sphaerospira_reducta
Sphaerospira	cookensis	Brazier	1875	Sphaerospira_mitifica
Sphaerospira	mossmani	Brazier	1875	Sphaerospira_blomfieldi_Gurgeena
Sphaerospira	bala	Brazier	1878	Sphaerospira_rawnesleyi
Sphaerospira	challisi	Cox	1873	Camaenidae_BL47_challisi
Spurlingia	darwini	Brazier	1872	Spurlingia_forsteriana
Temporena	whartoni	Cox	1871	Camaenidae_BL61
Trachiopsis	tuckeri	Pfeiffer	1846	Camaenidae_CY5
Xanthomelon	magnidicum	Iredale	1938	Xanthomelon_pachystylum_Dawson

Table S3. Shimodaira-Hasegawa (SH) tests of Bayesian and fast ML trees

			PAUP single model			RAxML 3p model			RAxML 4pB model		
	Method	Analysis	-lnL	∆-lnL	SH-test	∆-lnL	SD:	Sig.	∆-lnL	SD:	Sig.
62t taxa set	MCMC	62t single	43367.16	(best)		1.902	11.127	No	4.971	10.924	No
		62t 3p	43368.44	1.2783	0.9590	0.188	10.159	No	1.993	9.925	No
		62t 4pB	43369.77	2.6117	0.8450	(best)			(best)		
		147t single	43381.61	14.4469	0.4910	12.069	16.948	No	16.864	16.315	No
		147t 3p	43381.44	14.2826	0.4920	11.382	17.545	No	17.675	16.849	No
		147t 4pB	43381.44	14.2826	0.4920	11.382	17.545	No	17.675	16.849	No
		327t single	43384.87	17.7109	0.3770	17.103	15.836	No	21.622	15.225	No
		327t 3p	43392.19	25.0363	0.2090	23.960	18.963	No	29.622	18.516	No
		327t 4pB	43385.67	18.5138	0.3720	12.967	16.530	No	14.957	16.291	No
	fast ML	62t 3p	43367.42	0.2577	0.9660	0.273	10.699	No	2.380	10.308	No
147t taxa set	MCMC	147t 4pB	70198.58	6.1400	0.8290	4.435	22.910	No	2.191	14.170	No
		147t 3p	70192.44	(best)		1.397	16.918	No	(best)		
		147t single	70192.85	0.4119	0.9080	4.740	19.365	No	1.858	9.441	No
		327t 3p	70214.63	22.1898	0.3880	19.593	21.276	No	24.094	16.862	No
		327t single	70208.97	16.5307	0.5360	18.489	21.740	No	16.894	15.166	No
		327t 4pB	70216.81	24.3679	0.3510	22.419	21.556	No	14.812	19.829	No
	fast ML	147t 3p	70206.45	14.0025	0.5830	(best)			4.723	15.626	No
327t taxa set	MCMC	327t single	89050.58	(best)		21.298	32.242	No	11.874	32.201	No
		327t 3p	89056.78	6.2080	0.7730	17.278	32.051	No	19.758	31.791	No
		327t 4pB	89073.16	22.5818	0.5230	37.225	35.897	No	13.126	34.434	No
	fast ML	327t 3p	89071.01	20.4367	0.5570	(best)	0.002	No	(best)		

All partitions used GTR-G-inv model. Base frequencies: PAUP estimated; RAxML empirical
All SH-test ∆lnL and test statistics based on tree with best score for that particular model
Analysis indicates taxaset and partition/model used to generate the tree
Larger trees have been pruned to the matching taxaset

Table S4. Branch length and tree length comparisons among models and taxasets.
Analyses use the same set of 2000 MCMC sample trees as used elsewhere.

	models		taxasets	
62 taxa	1p	4pB	147t_3p	327t_3p
correlation	0.9994	0.9992	0.9925	0.9857
outside 95%CI	0	0	1	11

147 taxa	1p	4pB	327t_3p
correlation	0.9989	0.9991	0.9940
outside 95%CI	0	0	5

327 taxa	1p	4pB
correlation	0.9889	0.9970
outside 95%CI	6	0

(327t1p all shorter than 3p)
(Larger taxaset subtrees all longer than 62t_3p)

Product moment correlation coefficient between analysis and the most data complete 3p model analysis.
Analysis based on all branches seen in the posterior all compatible consensus trees for each set of comparisons.
95%CI is based on the MCMC variance of the most data complete 3p model analysis.

Tree length (MrBayes TL parameter) comparisons

62 taxa		62t_3p	147t_3p	327t_3p	62t_1p	62_4pB
	median	20.25	20.44	21.53	19.26	21.57
	0.95	21.6	21.3	22.7	20.31	22.84
	0.05	18.9	19.6	20.9	18.25	20.47

(all within 3p 95%CI)

147 taxa		147t_3p	327t_3p	147_1p	147_4pB
	median	35.64	37.84	35.65	36.53
	0.95	37.0	39.5	37.35	38.25
	0.5	35.6	37.8	34.35	35.24

(327t is above 95%CI)

327 taxa		3p	1p	4pB
	median	59.67	58.85	58.63
	0.95	61.91	61.44	61.51
	0.5	57.55	56.47	55.86

(all within 95%CI)

Figure S1A. 28SrDNA model Bayesian majority rule posterior consensus tree using GTR-G model with posterior probablities

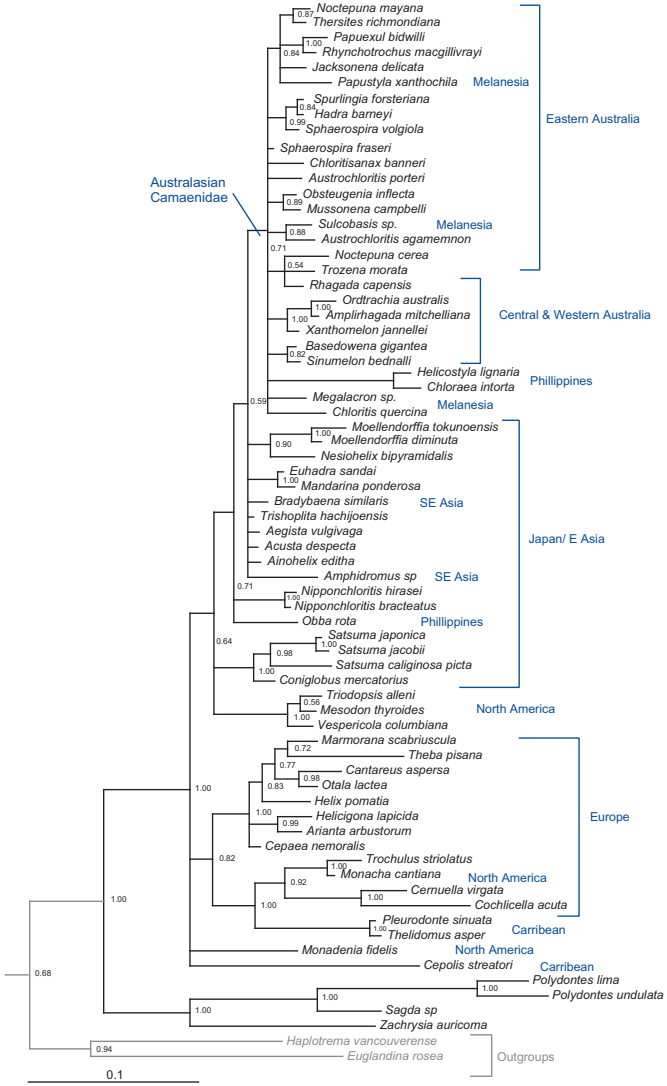


Figure S1B. 28SrDNA RAxML maximum likelihood tree using GTR-G model

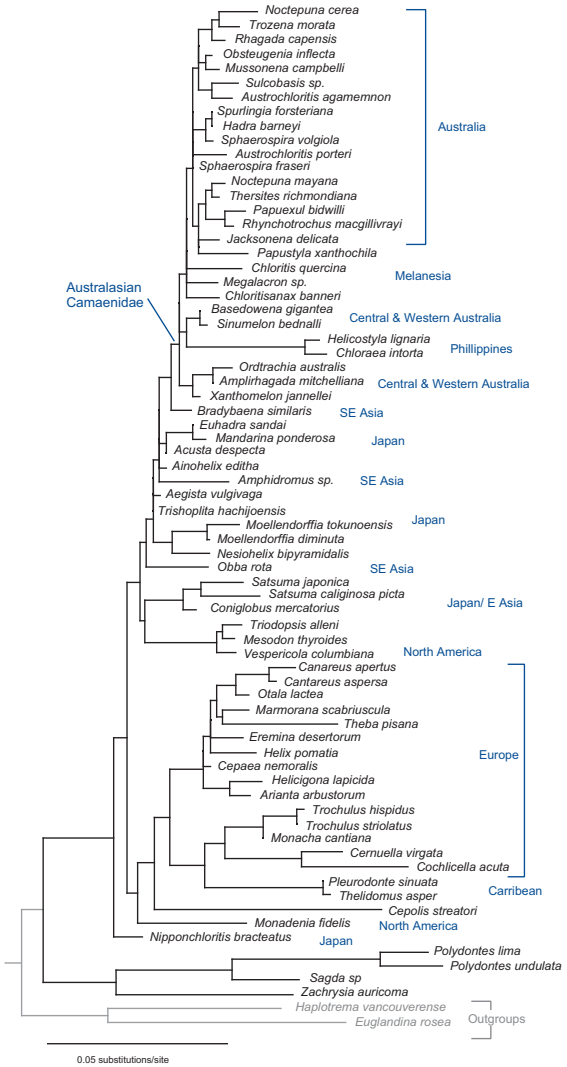


Figure S1C. 28SrDNA PAUP 4.10b SPR maximum likelihood tree using GTR-inv model

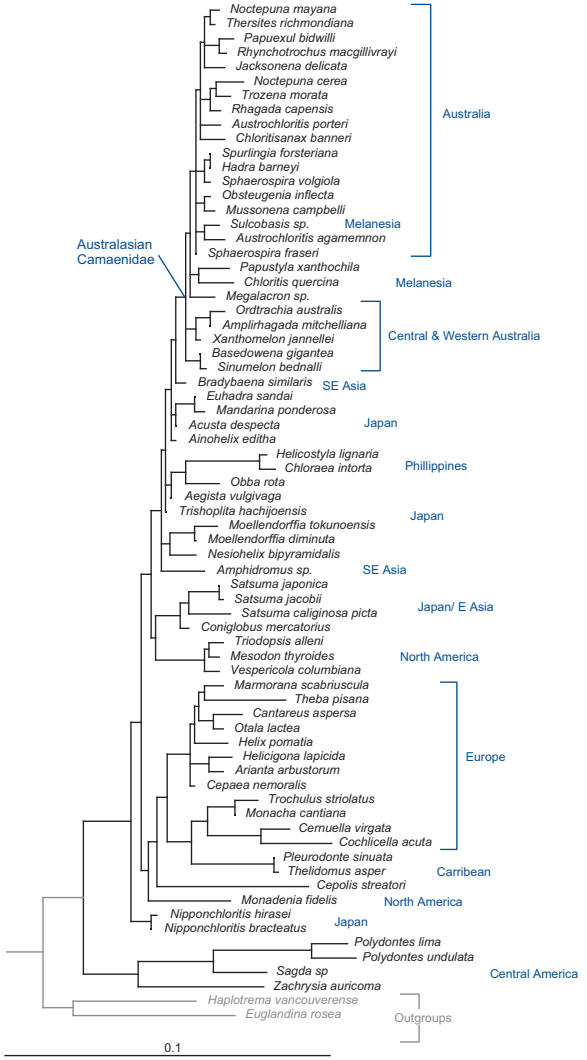
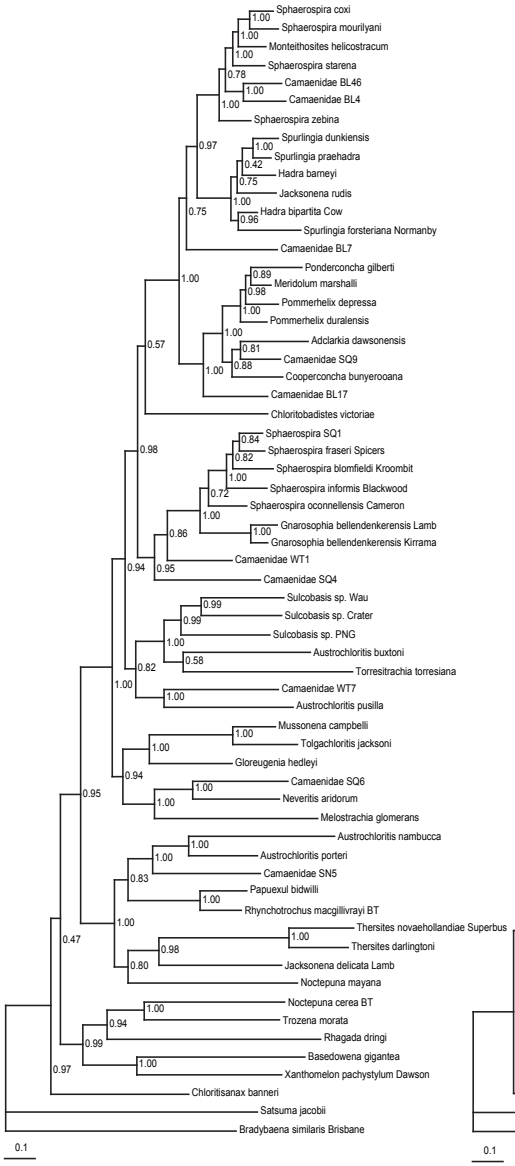
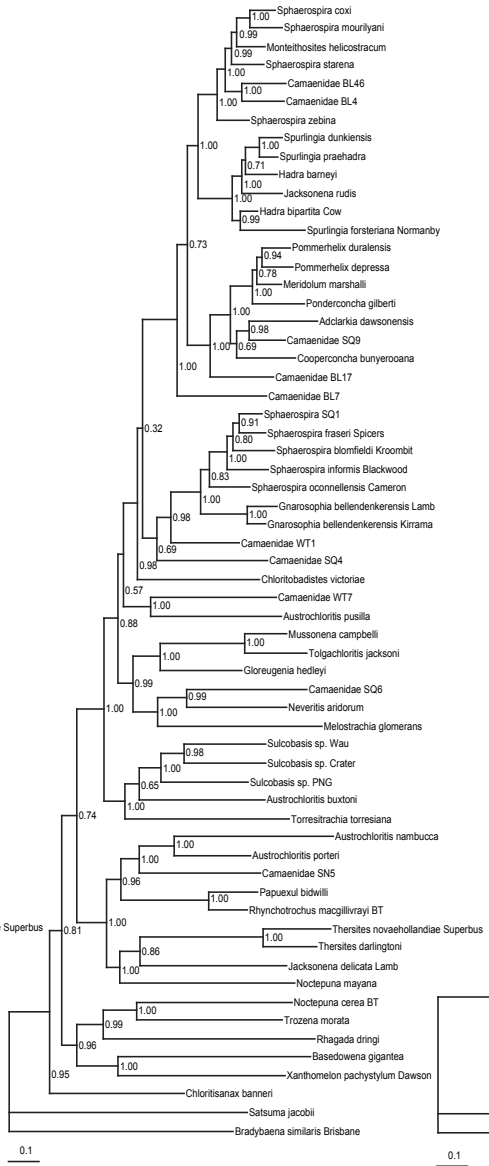


Figure S2. All compatible consensus trees for the 62t taxaset using the 3p model

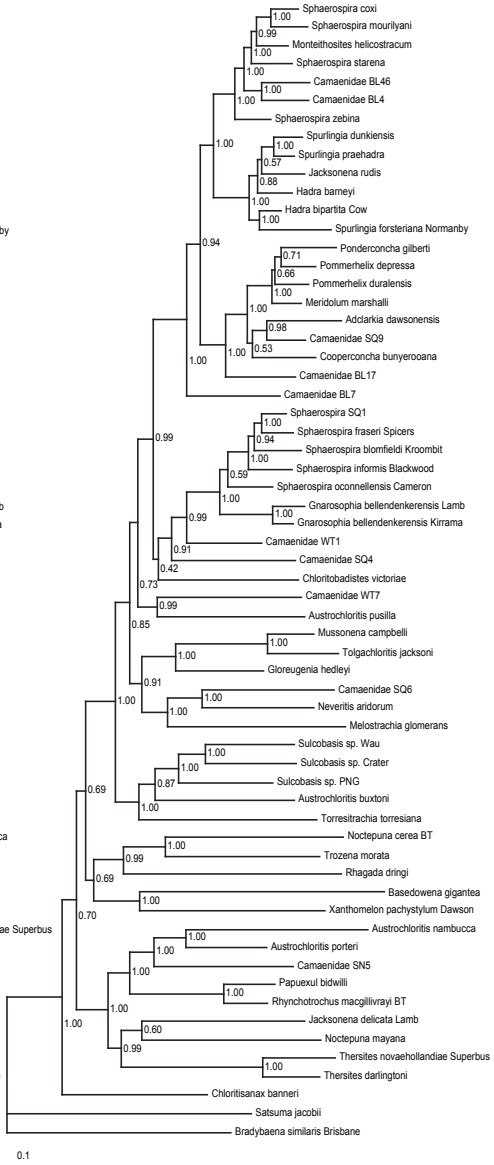
A 62t analysis (with posterior probablities)



B Pruned from 147t analysis



C Pruned from the 327t analysis



D RAxML (with bootstraps)

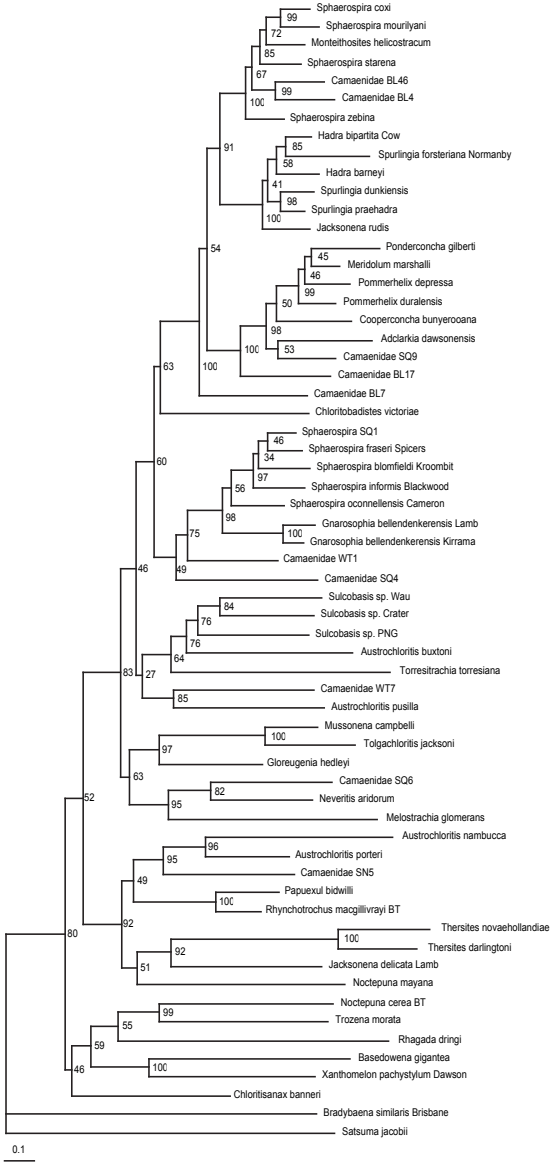


Figure S3. All compatible consenses for the 147t taxa set using the 3p model.

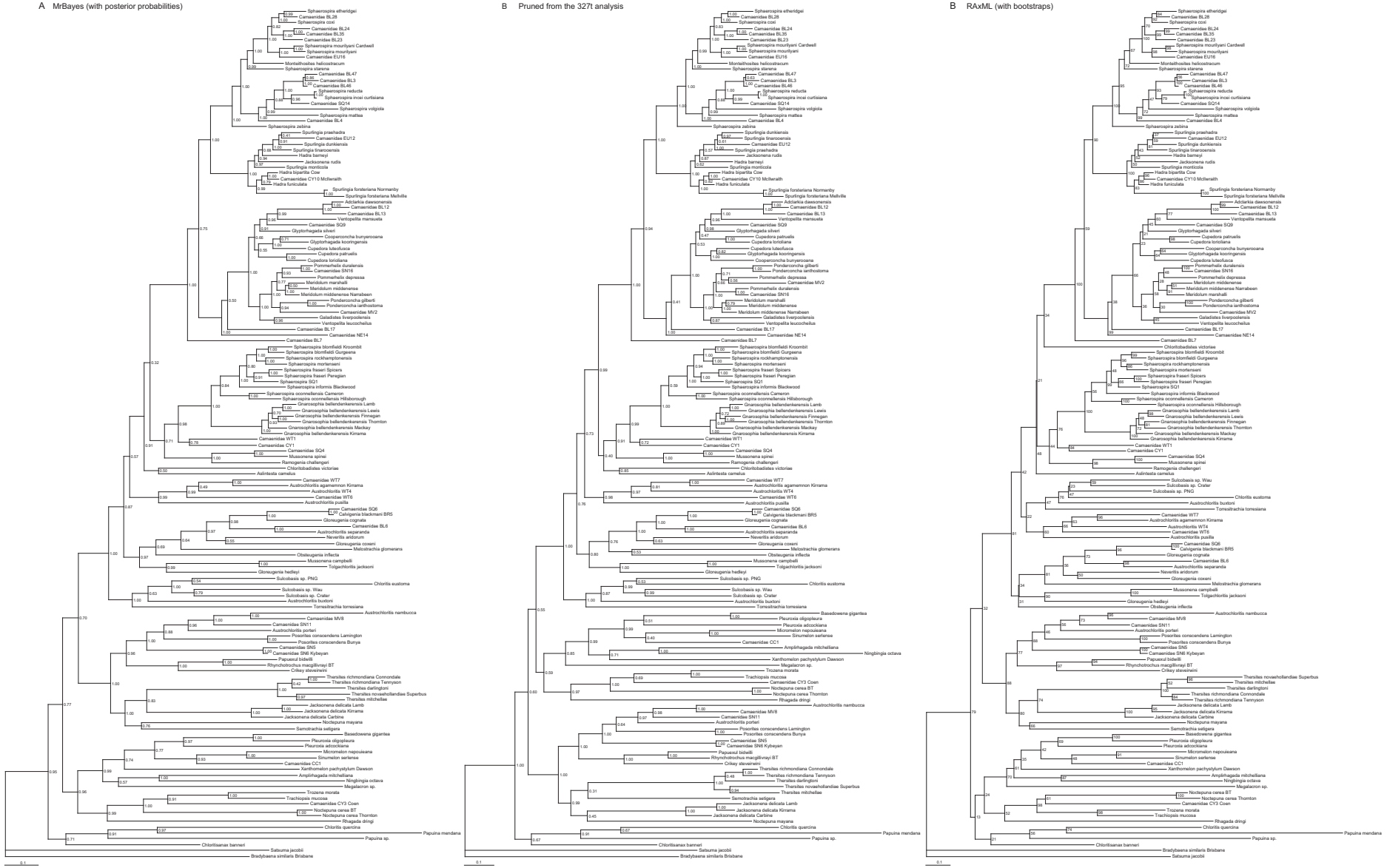


Figure S3. Consensus trees of 62 taxa with all genes among MCMC analyses

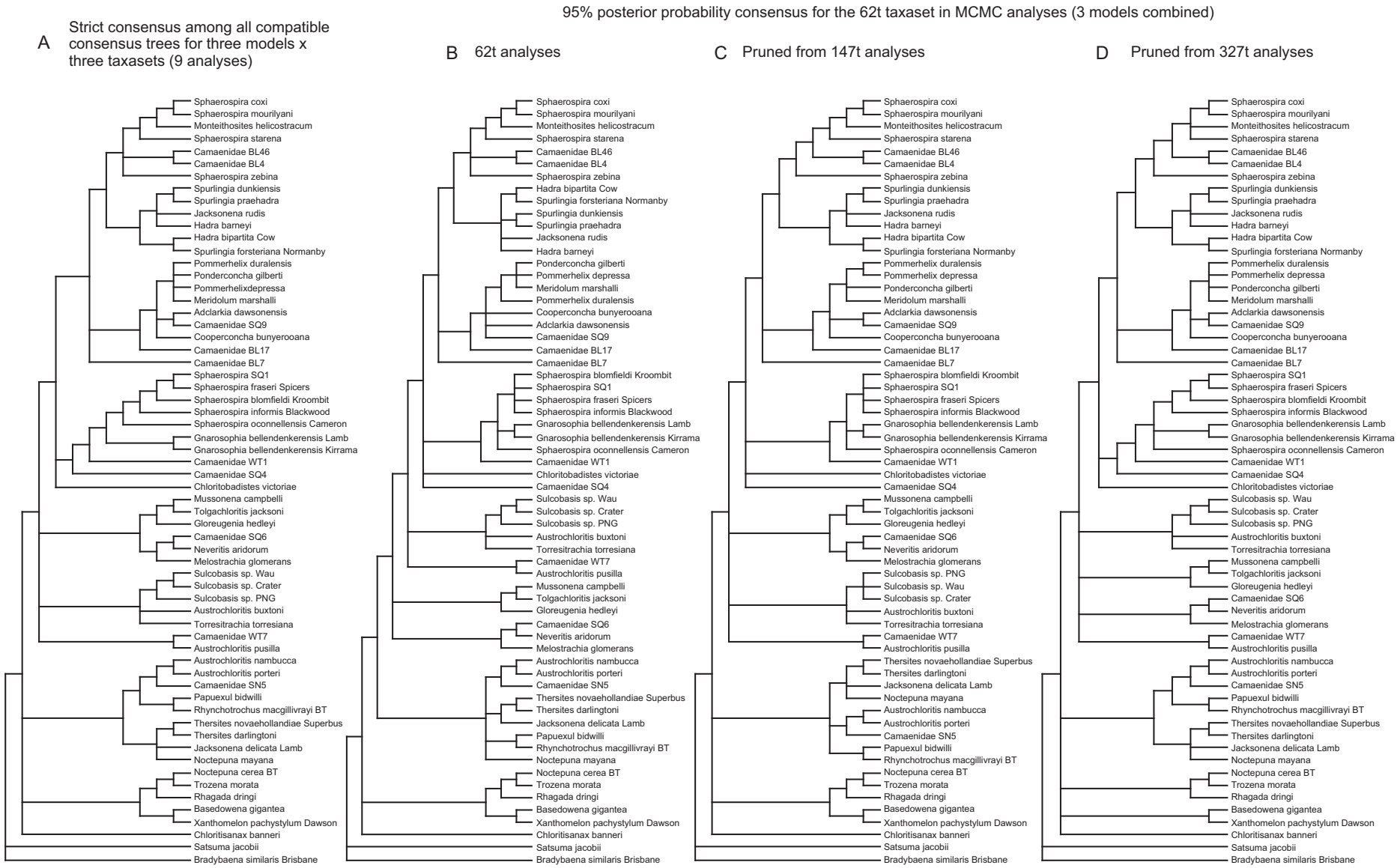


Figure S5. Consensus of 147t taxa set in MCMC analyses

A Strict consensus three models x two taxasets (6 analyses)

B 95% consensus from 147t analysis (3 models)

C 95% consensus from pruned 327t analysis (3 models)

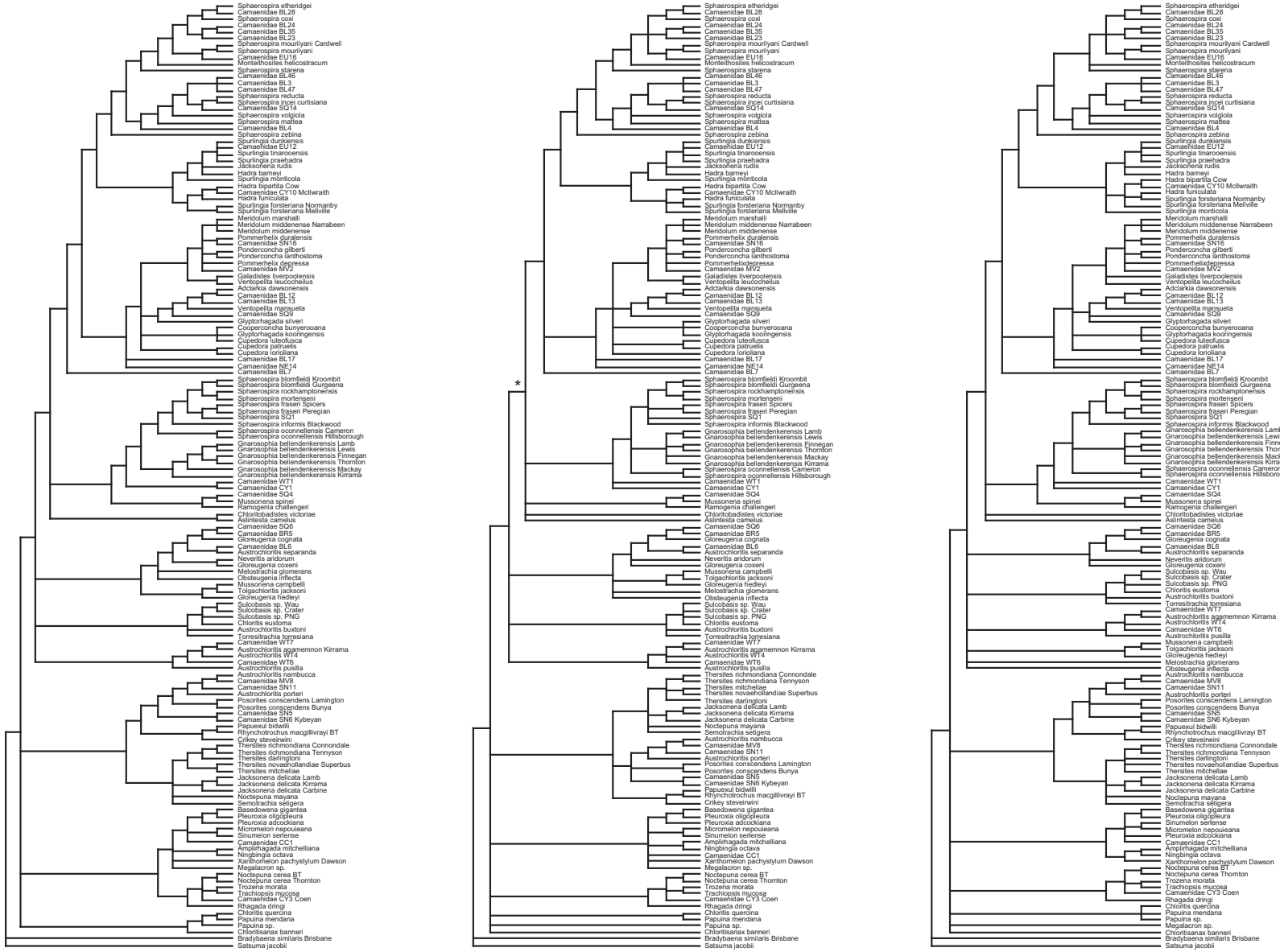


Figure S6. Branch length comparisons across models and taxasets . All common bipartition branch lengths from each of the taxaset analyses (pruned to the relevant number of taxa) plotted against the 3p model bipartition branch lengths. Confidence Intervals (CI) based on MCMC variation in the most data complete analysis CI lines are a visual guide only, results in supplementary Table S4 based on actual MCMC variation per individual branch; 121 branches for 62 taxa, 290 branches for 147 taxa.

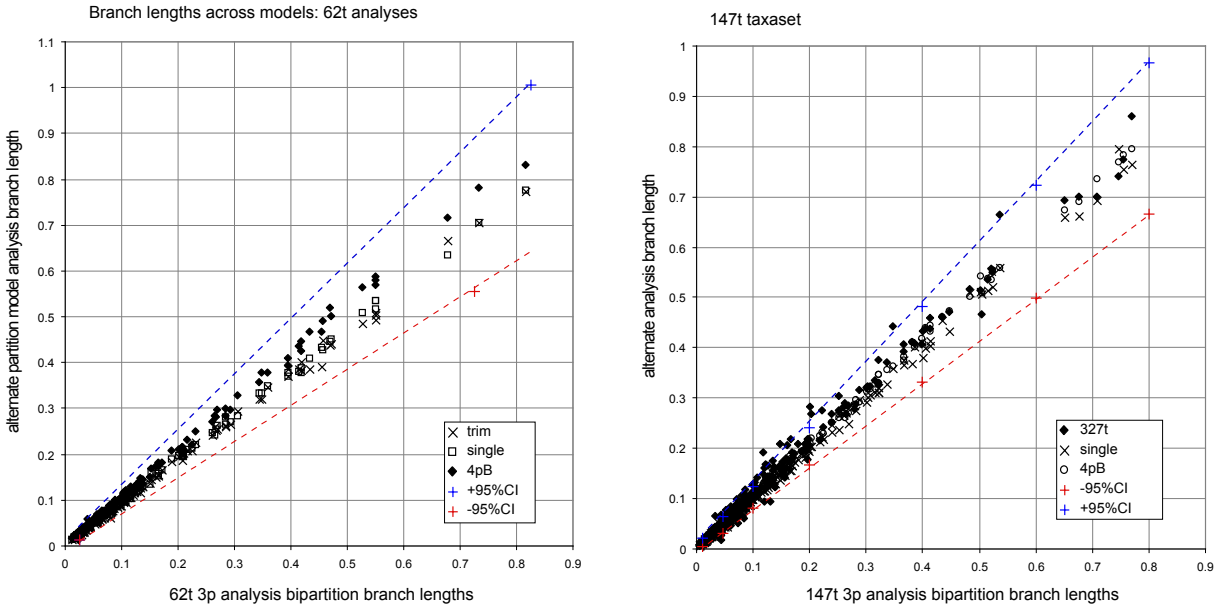


Figure S7. MCMC lnL distribution overlap

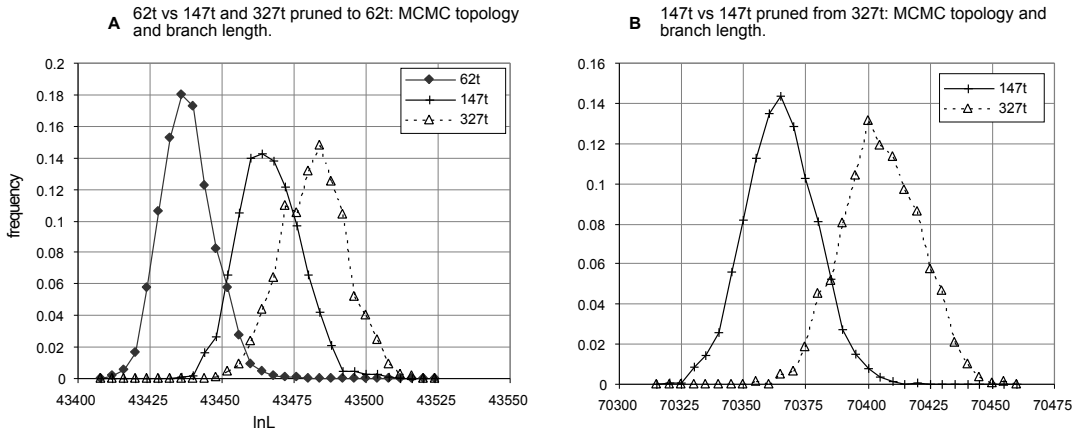
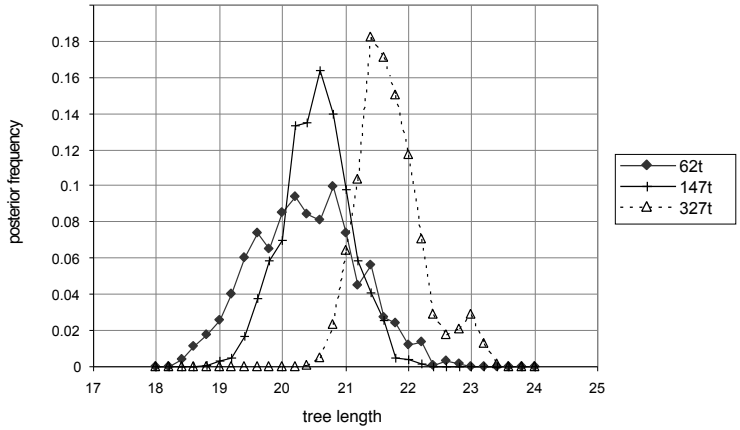


Figure S8. MCMC tree length (TL) distributions using the 3p model across taxasets, pruned to the 62t taxaset. Uses the same sample of trees as the branch length analyses.



Clade 1 all individuals mitochondrial COII (480 sites) HKY85 model Neighbour-joining bootstrap majority rule tree (1000 replicates) with support values shown for nodes with BS>70%
OTU code: genetic code_species_Mo number_locality
If Mo=QM specimen has no Mo number
For resolution among major clades see Chapter 2.

Clade 2 all individuals mitochondrial COII (480 sites) HKY85 model Neighbour-joining bootstrap
majority rule tree (1000 replicates) with support values shown for nodes with BS>60%
OTU code: genetic code_species_Mo number_locality
If Mo=QM specimen has no Mo number
For resolution among major clades see Chapter 2

Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875)

Andrew Hugall[†‡], Craig Moritz[§], Adnan Moussalli[†], and John Stanisic[¶]

[†]Cooperative Research Centre for Tropical Rainforest Ecology and Management, Department of Zoology and Entomology, University of Queensland, Brisbane, Qld 4072, Australia; [§]Museum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, CA 94720-3160; and [¶]Queensland Centre for Biodiversity, Queensland Museum, Brisbane, Qld 4000, Australia

Edited by David B. Wake, University of California, Berkeley, CA, and approved February 8, 2002 (received for review October 10, 2001)

Comparative phylogeography has proved useful for investigating biological responses to past climate change and is strongest when combined with extrinsic hypotheses derived from the fossil record or geology. However, the rarity of species with sufficient, spatially explicit fossil evidence restricts the application of this method. Here, we develop an alternative approach in which spatial models of predicted species distributions under serial paleoclimates are compared with a molecular phylogeography, in this case for a snail endemic to the rainforests of North Queensland, Australia. We also compare the phylogeography of the snail to those from several endemic vertebrates and use consilience across all of these approaches to enhance biogeographical inference for this rainforest fauna. The snail mtDNA phylogeography is consistent with predictions from paleoclimate modeling in relation to the location and size of climatic refugia through the late Pleistocene-Holocene and broad patterns of extinction and recolonization. There is general agreement between quantitative estimates of population expansion from sequence data (using likelihood and coalescent methods) vs. distributional modeling. The snail phylogeography represents a composite of both common and idiosyncratic patterns seen among vertebrates, reflecting the geographically finer scale of persistence and subdivision in the snail. In general, this multifaceted approach, combining spatially explicit paleoclimatological models and comparative phylogeography, provides a powerful approach to locating historical refugia and understanding species' responses to them.

Phylogeography seeks to reveal biogeographical history of species and the habitats they occupy via (*i*) qualitative spatial association of divisions between monophyletic clusters of alleles with biogeographic barriers, and (*ii*) quantitative estimates of historical population size (1–4). Much of this work has focused on mitochondrial DNA; however, stochastic variance limits our confidence in reconstructions of history from a single gene. One approach solving this limitation is to sample more genes (5). A more common approach is comparative phylogeography (6) in which sequence variation is surveyed at a single gene for multiple species across the same landscape. A limitation here is that histories of local extinction and recolonization may vary among species despite a common history of habitat fluctuation.

To improve inference of historical biogeography, we need to incorporate spatially explicit evidence from paleoecology into interpretation of species' phylogeography. Some recent studies have promoted the use of fossil evidence along with phylogeography to estimate historical distributions (7) or have examined sequence variation in the fossils themselves (e.g., refs. 8 and 9) However, appropriate fossils are sparse or nonexistent for most taxa. We explore a novel and more widely applicable approach that uses paleoclimatological models of species' distributions in conjunction with phylogeography.

Bioclimatic modeling (10) predicts potential distributions for species by deriving an environmental envelope from known distribution points and projecting this onto a spatially interpolated climate surface of an area. For paleomodeling, the same climate envelope for a species is mapped onto inferred paleoclimatic surfaces, these being based on analysis of local paleopalynology or other indices (11, 12). This method assumes that the species' physiological limits are constant over the time period concerned, an assumption that may not hold in some cases (8, 13). This approach is therefore most likely to be effective for species with climatically defined range limits in landscapes with steep climatic gradients, so that errors because of physiological evolution or from estimates of paleoclimate have less effect on predicted spatial distribution. These conditions hold for the fauna endemic to the Wet Tropics rainforests of north east Queensland, Australia.

The Wet Tropics (WT; Fig. 1) is the largest and most intact remnant of Gondwanan-derived wet forests that dominated the continent of Australia to the mid Miocene, declining in the Tertiary to its current state of fragmentation and isolation (12, 14, 15). This long-term decline, and particularly the Pleistocene climate oscillations, appears to have driven cycles of contraction and expansion of rainforest, probably structured around montane remnants surrounded by dry sclerophyll woodlands (16). These montane blocks are centers of endemism and have been proposed as Pleistocene refugia (17–19), and form the basis of the recognized subregions within the Wet Tropics (see Table 1 and Fig. 1). However, analysis of subfossil charcoal records have cast doubt on the integrity and existence of several putative refugia (20).

Congruence among vertebrate phylogeographies across these subregions support long-term restriction of rainforest-dependent fauna to at least two long-standing refugia, separated by the Black Mountain Corridor (BMC), across which rainforest was absent until ≈8 thousand years ago (Kya; refs. 21 and 22; Fig. 1). In addition, some vertebrate species show idiosyncratic disjunctions, perhaps indicating additional refugia in which these species persisted whereas the others did not (22). A more detailed record of historical rainforest habitat structuring might be preserved in the genetic pattern of taxa showing finer spatial scales of persistence and low vagility (6, 7). Such characteristics make land snails good candidates for recovering details of historical biogeography (23, 24).

Here, we use paleoclimatological modeling of the endemic land snail, *Gnarosophia bellendenkerensis* (Brazier 1875) to predict the location of refugia and to estimate the magnitude of change in the extent of suitable habitat in each biogeographic

---

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: WT, Wet Tropics; BMC, Black Mountain Corridor; Kya, thousand years ago; AMT, annual mean temperature; AMP, annual mean precipitation; PDQ, precipitation of the driest quarter; LGM, last glacial maximum; MT, Malbon-Thompson Range; BK, Bellenden Ker; AT, Atherton Tableland; FU, Finnegan Uplands; KU, Kirrama Uplands; AU, Atherton Uplands; TU, Thornton Uplands; CU, Carbine Uplands; LU, Lamb Uplands; WU, Windsor Uplands.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AY048376–AY048422).

[‡]To whom reprint requests should be addressed. E-mail: ahugall@zen.uq.edu.au.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

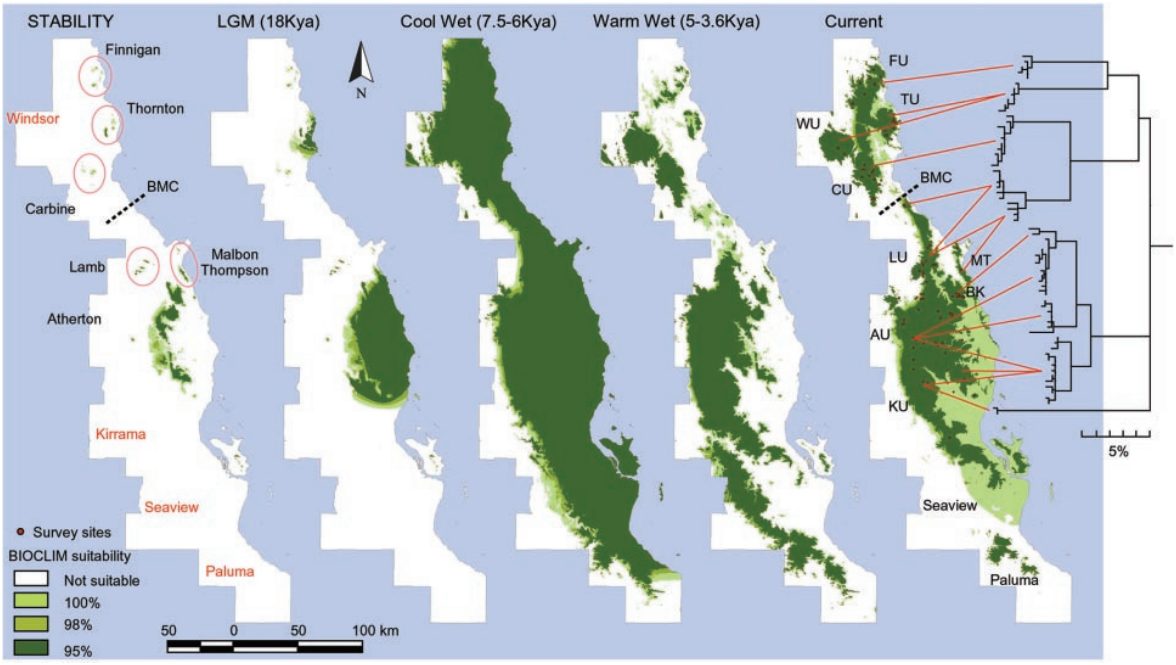


Fig. 1. BIOCLIM distribution models with mtDNA haplotype phylogeny, showing the geographical association of the major clades. Region names to the left, with codes used throughout on the right. Light to dark shades of suitability denote models by using respectively 100, 98, and 95% observed distribution climate parameter limits. Distribution sites for modeling shown in red. Eastern boundary is current coastline; the remainder, spatial data limits. Sequence divergence scale under phylogeny. Right to left: current climate model; then the three paleoclimate models warm-wet, cool-wet, and LGM; far left, the STABILITY surface—intersection of the other four models. On the STABILITY surface refuges for FU, TU, CU, LU, and MT are circled, and areas going extinct are named in red. The central part of the WT is shown again in Fig. 2, in more detail.

subregion and then use a mtDNA phylogeography of the snail to evaluate these predictions. We also compare the snail phylogeography to those reported previously for rainforest-restricted vertebrates endemic to the wet tropics (22, 25). The WT paleoclimates and modeling cover the last glacial maximum-Holocene period (last 20 thousand years). The palynological record also shows similar fluctuations back at least 100 Kya (16), whereas the vertebrate data suggests much older refugial patterns, more in keeping with Pliocene ages. Therefore, the modeling should allow direct quantitative comparison to within population genetic patterns (at the tips of the genealogy) and qualitative comparison to older spatial patterns. In particular, we

Table 1. Definition and attributes of nine Wet Tropics subregions

Region code	FU	TU	WU	CU	LU*	BK†	AU‡	KU	MT
Nucleotide diversity (π)									
G. bellendenkerensis	0.54	0.43	0.46	0.60	1.86	1.71	3.29	0.09	0.40
C. laevis	0.12	0.11	0.11	0.44	0.47	0.27	0.37	0.14	0.00
G. queenslandicus	0.71	0.26	0.19	0.19	0.46	0.00	0.13	0.16	0.43
Vertebrate average π§	0.36	0.12	0.13	0.25	0.33	0.24	0.19	0.24	0.10
C.l.-G.q. average π¶	0.42	0.19	0.15	0.32	0.47	0.14	0.25	0.15	0.22
Snail/vert. π ratio	1.30	2.34	3.07	1.91	4.01	12.67	13.15	0.59	1.88
Rainforest area‖									
Recent	28,090	23,770	25,620	39,430	20,824	29,260	170,489	59,093	6,192
LGM	2,399	31,504	0	1,869	2,034	29,260	334,099	27	6,192
STABILITY	2,400	3,121	0	1,870	2,034	26,957	76,975	27	4,484
Sampling									
Genetics									
Individuals	7	8	2	15	10	10	57	8	4
Haplotypes	5	5	2	10	7	5	30	2	3
Locations	4	4	2	6	6	5	20	3	3
Modeling									
Locations	10	13	3	14	12	15	29	3	3

*Includes Mcallister Range.
†Bellenden Ker and Bartle Frere.
‡Includes Herberton and Walter Hill Ranges, and BK.
§Average of the six species in ref. 22.
¶Average of *C. laevis* and *G. queenslandiae.*
‖BIOCLIM model, hectares.

are interested in evidence of multiple Pleistocene refugia and Holocene patterns of expansion from these.

Materials and Methods

Species and Modeling. *G. bellendenkerensis* is a moderately large (height = 33.08 mm, width = 40.19 mm, n = 97 adults) globose Camaenid land snail endemic to Wet Tropics rainforests. It is a leaf litter/log generalist and occurs predominantly in the upland (above 400 m) mesothermal rainforests extending from the Finnegan uplands (FU) in the north (15° 43′ 10′′ S) to the end of the Kirrama uplands (KU) in the south (18° 14′ 6′′ S; Fig. 1). We used either specimens from the Queensland Museum (stored in 70% ethanol) or live-caught animals. The museum database and field collections provide 102 records, accurate to 3 decimal degree places, for distribution modeling. Extensive molecular systematic surveys by the authors (unpublished data) show *G. bellendenkerensis* to be monophyletic. *G. bellendenkerensis* has recently been removed from the genus *Hadra* (26). Our definition of geographic subregions within the Wet Tropics (Table 1) follows previous biogeographical studies (18, 19, 21, 22, 27).

Climate-based distribution modeling followed the BIOCLIM procedure (10) and used spatially interpolated estimates (ANUCLIM, ref. 28) of annual mean temperature (AMT), annual mean precipitation (AMP) and precipitation of the driest quarter (PDQ) on an 80-m resolution digital elevation model for the wet tropics. These layers are the only ones for which paleoclimate scenarios are available; however, they are adequate to provide a good fit to current rainforest (12) and snail distributions (see below). The BIOCLIM procedure sets upper and lower boundary limits for each climate layer based on the upper and lower limits of the observed distribution points, allowing for trimming of outliers. Because the current distribution of the snail reaches the extremes of minimum temperature and maximum rainfall found in the wet tropics, we considered that these limits may not be intrinsic to the snail and therefore did not invoke these boundaries in the paleomodels. Boundary limits (100%) for the snail BIOCLIM model are AMT of 24°C (unused min. 15.5°), AMP of 1,398 mm (unused max. 6644 mm), and PDQ of 83 mm (unused max. 672 mm). Modeled distributions are restricted to the current coastline by the available digital elevation model. Whereas some refugia could have extended further east, the coastal platform was more likely gallery forest/savanna when exposed (29).

Paleoclimates have been estimated for the WT for late Pleistocene through the Holocene by examining overlap of current bioclimatic limits for species abundant in the respective sections of pollen cores (11, 12); these climate estimates have been used recently to model geographic shifts in vegetation types across the wet tropics (30). We predicted potential distributions of the snail for the three paleoclimates, corresponding to a cool-dry climate at the last glacial maximum (LGM, e.g., 18 Kya), a cool-wet period extending from ≈8–6 Kya, and warm-wet periods from about 5–3.6 Kya. Estimated shifts of the three climate parameters relative to present are as follows: LGM AMT −3.5°C, AMP 50%, PDQ 60%; Cool-Wet AMT −2.0°C, AMP 120%, PDQ 200%; and Warm-Wet AMT +2.0°C, AMP 125%, PDQ 150%. The STABILITY surface represents the intersection of all scenarios: presence in the current and three paleoclimate models. Areas included in the STABILITY surface therefore represent predicted refugia. For quantitative comparison with genetic estimates of expansion, we used the ratio of areas within the STABILITY surface vs. the CURRENT model as surrogates for historical fluctuation in population size.

Molecular Analyses. For most specimens, DNA was extracted by using the Chelex (Bio-Rad) method. All specimens were sequenced by Applied Biosystems automated sequencing, most with both strands. The final dataset used here comprises 121 individuals from 53 localities scored for 460 bases of COII giving 67 haplotypes. COII primers are coding strand: 5′-AAA TAA TGC TAT TTC ATG AYC AYG C-3′; H strand: 5′-GCT CCG CAA ATC TCT GAR CAY TG-3′. The snail sequences are lodged with GenBank, accession nos. AY048376–AY048422.

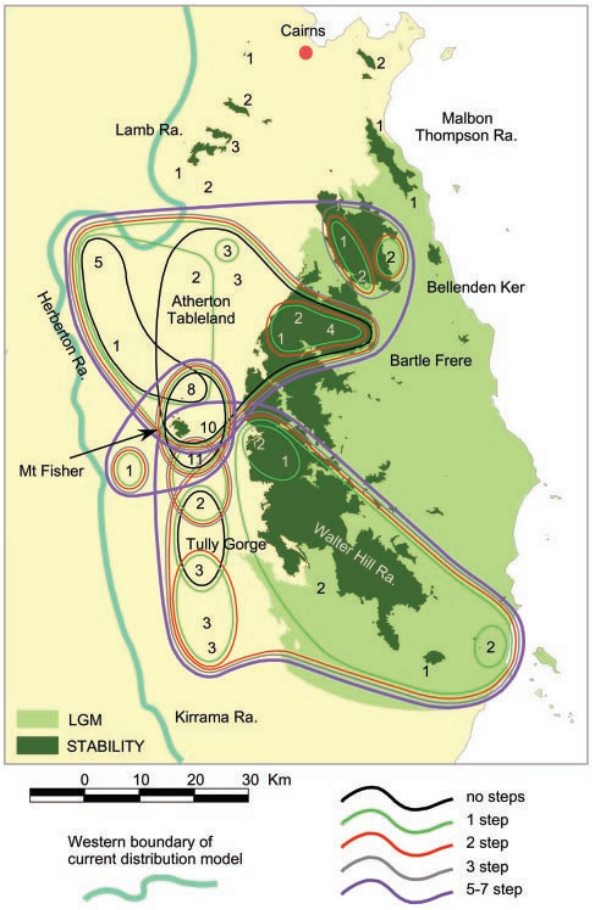


Fig. 2. MtDNA nested clade map of the Atherton Tableland and adjacent regions overlaying BIOCLIM LGM and STABILITY paleodistribution models (light and dark shades, respectively). Drawn for haplotypes shared across sites, sample sizes indicated for localities. The three 5- to 7-step clades (outermost lines) require another 8 steps to coalesce. Broad band to left indicates western boundary of current predicted distribution.

All analyses used the Tamura-Nei model to estimate sequence divergence. A phylogeny of haplotypes was inferred with PAUP* (31) by Neighbor-Joining, and nucleotide diversity and net nucleotide divergences within and among populations estimated with REAP (32). Simultaneous optimization and estimation of likelihood surfaces for theta and population growth rate parameters were done with the Metropolis-Hastings Markov Chain Monte Carlo genealogy sampler FLUCTUATE of Kuhner *et al.* (33). Number and length of chains run in the Metropolis-Hastings sampler accorded with guidelines by using a maximum likelihood starting tree. Quantification of population size and size change from theta and growth parameters estimated from sequence data requires estimates of mutation rate and time. The relationship of effective population size (N_e), diversity (Θ), growth parameter (g), time (t) in generations, and mutation rate μ per generation is given by $N_e = (\Theta/2\mu)e^{(-g\mu t)}$ (33).

Results

BIOCLIM Modeling of Current and Paleodistributions. The bioclimatic modeling of the snail (Figs. 1 and 2) predicts a substantial reduction

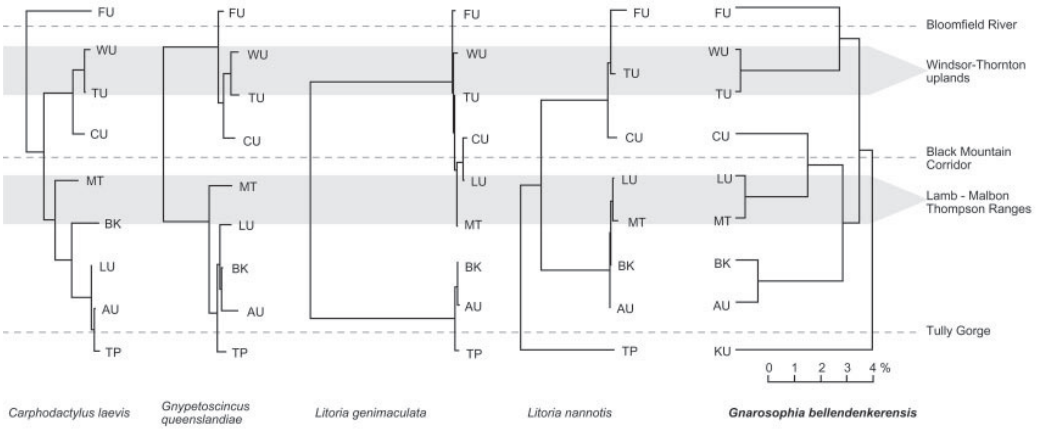


Fig. 3. Neighbor-joining trees of mtDNA net nucleotide divergence (*Da*) for snail and four vertebrate species, all drawn to the same nucleotide divergence scale, Tamura-Nei model. Snail; COII gene, lizards *Carphodactylus* and *Gnypetoscincus*: *Cyt* b, and *Litoria* frogs: COI. The southern subregion defined here as Kirrima Uplands (KU; Fig. 1) is to be compared with TP of Schneider *et al.* (22). *L. nannotis* not sampled in WU. Relationships between WU-TU and between LU-MT highlighted, dashed lines mark major genetic breaks in the vertebrates.

in range during the arid and cool conditions of the LGM, particularly south of the Atherton Uplands (AU), and severe fragmentation north of AU. Specifically, it predicts small discrete refugia for three northern subregions, Finnigan (FU), Thornton (TU), and Carbine (CU) Uplands, another on the central coast (Malbon-Thompson Range, MT), a major refugial area across the eastern AU and adjacent coast, and smaller areas just to the north (Lamb Uplands, LU) and the west at Mt Fisher (see Fig. 2 for additional detail). Conversely, the LGM model predicts the snail to be absent from the Windsor Uplands (WU) and nearly so from the KU. The modeled range is maximally expanded through the cool-wet period of the early Holocene (6–7.5 Kya), highlighting the potential for recolonization of all areas. During the warmer wetter period of the late Holocene (3.6–5 Kya), the modeling predicts further contractions, especially from the lowlands and eastern uplands (e.g., TU and AU in Fig. 1). Overall, these predicted ranges are very similar to fluctuations in mesothermal rainforest predicted by using similar methods (refs. 12 and 30). The STABILITY surface is essentially the LGM model with the Thornton and Atherton uplands (TU and AU) reduced by limitations of the late Holocene warm-wet scenario. These models predict changes in range across subregions varying from relatively minor to drastic [e.g., BK (Bellenden Ker and Bartle Frere) vs. others in Table 1].

Molecular Phylogeography and Comparison with Bioclimate Models. The 67 haplotypes detected fell into six divergent and geographically exclusive lineages; four with low and unstructured diversity (FU, TU/WU, CU, KU), one structured into two sub lineages (LU/MT), and the last a large complex (AU) of three sub lineages (Table 1, Fig. 1; all bootstraps >90%). Maximum sequence divergence among lineages was 12–15% and net divergence (*Da*) among areas (summarized in Fig. 3) ranged from 6.5% to 15.4% with average of 10.7%. These six major lineages showed almost complete geographical restriction to a single biogeographic subregion. Exceptions were identical or closely related (one step) haplotypes shared between WU and TU, AU and BK, and MT and LU.

This phylogeographic pattern can be compared qualitatively with the predictions of the paleomodeling: each of the predicted discrete LGM/STABILITY refugia of FU, TU, CU, LU, and AU hosted a discrete genetic lineage, with the latter showing greatest internal complexity (see below). The absence of a distinct genetic lineage for the WU matches the predicted elimination of suitable environment at the LGM; rather, the WU samples were a subset of the adjacent TU lineage, suggesting colonization from that source. However, the LU samples included, in addition to a locally endemic lineage, another shared with the adjacent coastal refuge (MT). The modeling also failed to clearly predict a refuge to match the highly divergent lineage found south of the Tully Gorge (KU, Table 1).

The relatively large AU-BK region is the best sampled for both genetic and spatial modeling analyses. Fig. 2 shows the geographical spread of nested haplotype clades superimposed on the LGM model and STABILITY surface (100% bounds). Again, the combination of molecular data and paleomodeling is highly informative. Key features are as follows: (*i*) the northern and southern limits of the AU/BK area represent contact zones with divergent lineages; (*ii*) several haplotypes are widespread (i.e., spread across at least half the area); (*iii*) there is substantial diversity in (western) areas predicted to be largely unsuitable at LGM but occupied in later phases; and (*iv*) diversity is structured into three sublineages (with bootstrap support >90%), one widespread with closely related haplotypes (0–1 step) across the Atherton Tableland (AT) and Bartle Frere and a sister group confined to Bellenden Ker, a second widespread with closely related haplotypes (1 step) across the Walter Hill Range and adjacent western areas, and the third confined to the western AT and centered around Mt. Fisher. Thus, the range of each lineage includes at least one predicted refuge (which equals STABILITY areas in Fig. 2), and all three lineages overlap in the vicinity of the predicted Mt. Fisher refuge.

For most regions, nucleotide diversity is low ($\pi < 0.6\%$; Table 1), three exceptions being LU, BK, and AU, each of which is structured into multiple sublineages. The AU region includes the three sublineages discussed above. The BK region is separated in the STABILITY surface into two areas corresponding to the two highest mountains in Queensland—Bartle Frere and Bellenden Ker, each of which hosts a separate three-step clade (Fig. 2). These observations indicate that the high diversity in each of these areas reflects admixture from separate refugia. The same reasoning can be applied to LU, which hosts two lineages, one shared with MT (another predicted refuge).

The distribution modeling predicts population expansion for some regions, defined as the area change between STABILITY and current models. This ratio can be compared with signals of population change in the genetic data, defined as the likelihood estimates of the exponential growth parameter, g (33). Only six of the nine subregions had sample sizes ($n > 7$) adequate for this purpose, including a sublineage of AU extending from Mt. Bartle Frere across the Atherton Tableland (AT-BF; see Fig. 3). This subset of the AU samples was included to represent a presumed expansion from a single source in keeping with the assumptions of the analysis. Table 2 gives the combined optimal Θ and g estimates, and likelihood support for $g > 0$.

There is a reasonable correlation across subregions between area ratio and estimates of g relationship ($r = 0.76$; $P = 0.04$): subregions with a large area increase also show substantial g, except LU being

Table 2. Optimal g/θ values and ΔlnL support for $g > 0$

Region	Area ratio*	θ	g	ΔlnL
AU	2.2	0.054	34	0.7
TU	7.6	0.018	611	1.0
LU	10.2	0.023	105	0.6
FU	11.7	0.046	934	1.1
AT-BF†	14.5	0.056	2266	13.1
CU	21.1	0.050	1639	4.9

*Ratio of modeled current distribution to STABILITY surface (100%).
†Bartle Frere-Atherton Tableland clade, see Fig. 2.

low. Likelihood ratio tests reject constant population size for CU and BF-AT. For CU, we get a relationship of $g = 1{,}639$ to area ratio = 21.1, with an overall area ratio/$g \approx 20/2000$. Similar results apply by using the LGM areas. Snail effective generation time of about one year (34)—the LGM time frame and clear subfossil evidence of rainforest expansion by 8 Kya (22)—suggests a time scale for expansion of the order of 10–20,000 t. For the genetic growth to quantitatively match the amount of spatial growth then requires a μ of around $0.7–1.3 \times 10^{-7}$, between phylogenetic and within population estimates of mtDNA mutation rate (26, 35). We note that some areas included here deviate from the assumptions of the coalescent expansion model in having substantial substructure (AU) or possible admixture (LU), and this result may underly the observed underestimate of g relative to the values predicted from the paleomodeling.

Comparison of Snail and Vertebrate Phylogeography. Geographic patterns of molecular diversity for the snail and some previously studied rainforest-restricted lizards and frogs from the wet tropics (22, 25) are summarized in Table 1 and the *Da* area phenograms in Fig. 3. The snail phylogeography is essentially a composite of the patterns for individual vertebrate species, reflecting idiosyncratic as well as common genetic breaks among the former, as well as admixture zones and inferred patterns of recolonization.

Proceeding from north to south, the divergence between FU and TU in the snail reflects the pattern seen in the gecko *Carphodactylus laevis*, and to a lesser extent, the skink *Gnypetoscincus queenslandiae*. The phylogeographic connection between WU and TU in the snail is common to all vertebrate species sampled from both areas (Fig. 3) where distinction allows. For the coastal MT range, vertebrates were a mixture of endemic (*C. laevis, G. queenslansdiae*) and immigrant lineages, the latter derived from either the north (*Litoria genimaculata, Carlia rubrigularis*) or the south (*Litoria nannotis*). The particular MT-LU connection seen in the snail occurs also in *C. rubrigularis* and *L. genimaculata*. In the center of the range, the association between AU and the adjacent BK is present in all of the vertebrates assayed except for *C. laevis* (Fig. 3), and the substructuring within AU-BK (Fig. 3) reflects patterns of mtDNA diversity within *G. queenslandiae* (36). At the southern limit of the snail, the divergent KU lineage reflects the pattern seen in *L. nannotis*.

Zones of admixture between major phylogeographic lineages of the snail have been observed at sites on the northern Atherton Tableland (AU/LU) and also near the Tully Gorge (KU/AU clades). Each of these corresponds to admixture zones seen in vertebrates, the former region having secondary contacts between major northern and southern lineages for the skink *C. rubigularis*, the frog *L. genimaculata*, and the bettong *Bettongia tropica* (ref. 37; B. Phillips, unpublished data) and the latter for the frog *L. nannotis* (ref. 22; M. Cunningham, unpublished data).

Turning to quantitative estimates, we compared nucleotide diversity (π) in the snail with that in the two most highly structured lizards (*C. laevis* and *G. queenslandiae*) as well as the average for all frogs and lizards (Table 1). In most areas, diversity in the snail is about 2- to 3-fold higher than vertebrates. Taking CU as a discrete area with the best individual estimates of π for both vertebrates and snails, we get 0.60% for the snail and 0.25–0.32% for vertebrate averages, a 2-fold difference as expected for mtDNA *Ne* in a hermaphrodite. Relative to vertebrates, snail diversity is exceptionally high in BK and AU. For each of these areas, the comparison of modeling and phylogeography suggests admixture from multiple refugia in the snail. Across four common and major phylogeographic divisions (FU-TU, LU-CU, AU-KU, and AU-CU) estimates of net sequence divergence (*Da*) for the snail and the more highly subdivided vertebrates are comparable (Fig. 3): snail *Da* ranges from 6.5% to 15.4% (avg. 9.7%), vertebrates from 4.2% to 14.8% (avg. 7.9%). Overall, the similarity of snail and vertebrate divergences and diversities suggests similar substitution rates, rather than the 10- to 20-fold higher rate postulated recently (38–40).

Discussion

Two concerns for comparative phylogeography are stochastic variance of loci, the "gene tree-species tree" problem, and differing responses among taxa because of varying habitat requirements or vagility. Modeling of paleodistributions may help to address both of these issues and can provide spatially explicit hypotheses about historical distributions for taxa lacking evidence from fossils.

In most natural environments, current patterns result from an amalgam of historical and current processes. For the WT fauna, extensive fluctuations in rainforest area and connectivity have led to episodes of differential contraction and extinction, followed by range expansion and recolonization, from multiple refugia. Paleomodeling can provide the explicit spatial hypotheses needed to untangle such complex histories and resultant phylogeographies. In this sense, paleomodeling provides an alternative to gathering more genetic data (i.e., more loci), and one that provides an independent demarcation of historical population structure. The WT provides an optimal testbed for this approach because species distributions are structured by steep environmental gradients into discrete geographic zones (12), because there are multiple detailed analyses of pollen sequence from which recent local paleoclimates can be estimated (11, 16), and because there is extensive information on species' distributions (19, 27) and phylogeography (21, 22, 25). In the present study, the combination of modeled paleodistributions and mtDNA phylogeography for the snail is highly informative, providing a spatial template of the size and distribution of mesothermal rainforest refugia against which the evolutionary and biogeographic history of other taxa can be compared.

Implications for Biogeography and Evolution in the Wet Tropics Rainforest Fauna. In general, attempts to locate and circumscribe historical refugia in tropical rainforest systems have proved contentious (41, 42), and the same is true for the Australian wet tropics where there is some discrepancy between inferences from paleoclimatic models (12), current phytogeography (17), and evidence for drier sclerophyll vegetation and burning within putative refugia (20). Although consistent phylogeographic divisions across the BMC in vertebrates support the division between the major northern and southern refugia predicted by modeling (22), further details are obscured for vertebrates by varying histories of local extinction and expansion, stochastic variance of gene trees, or both (43).

The BMC break evident in vertebrates is present in the snails (between LU/MT and CU clades), but, in the latter and the more subdivided of the vertebrates (Fig. 3), this result is just one of several deep divisions. For the snail, both the phylogeography and paleomodels indicate the presence of multiple refugia within the north (FU, TU and CU) and also across the eastern Atherton region (MT, LU, AU, and BK). Although the climate estimates necessary for paleomodeling are available for the LGM only through the Holocene, the substantial levels of sequence divergence seen among these areas for both the snail and some vertebrates are consistent with Kershaw's (16) suggestion that angiosperm-dominated rain-

forests were restricted to these refugial areas for much of the Quaternary and perhaps earlier.

Vertebrate phylogeographic analyses indicate some species survived the LGM in multiple refugia to the north and south of the BMC whereas others probably persisted in just one refuge in each area (Fig. 3; refs. 27 and 43). If we use the predicted refugia for the similarly distributed snail as a template, we can assess which areas were suitable for all vertebrate species and which for only some. The combined modeling and genetic analysis of the snail predict TU as the major northern refuge for vertebrates, with the related WU populations being recolonized from there. By contrast, FU and CU were probably too small or too fragmented (or both) to support refugial populations of most vertebrate species across the Pleistocene. Exceptions include *C. laevis* and perhaps *G. queenslandiae*, which appear to have persisted in the FU refuge (Fig. 3). Similarly, the AU/BK region was clearly the major refuge south of the BMC, although the same two lizard species may also have persisted with the snails in the small refuge on the coastal MT range. That most upland areas were too small for persistence of most endemic and rainforest-restricted vertebrates is also suggested in comparisons of local species richness and endemism (19, 44).

Although few vertebrate species seem to have survived within them, the presence of small, geographically disjunct refugia across the northern uplands (excluding WU) and on Atherton Tableland is consistent with analysis of taxa with a higher proportion of narrowly distributed species, e.g., low vagility insects and snails (27). For these species, areas such as CU, FU, KU, and MT have high species richness, including locally endemic species. Phylogenetic analysis of such groups may shed light on the extent to which long-term subdivision among the northern areas and across the Atherton region has contributed to speciation as well as phylogeographic structure within species. For example, sister species of flightless dung beetle (*Temnoplectron*) occur on adjacent refugial areas in the north (45).

In comparing taxa, a differing response because of ecology needs to be taken into account. This cross-ecology approach potentially can provide an additional perspective, but up to now has been done subjectively. In the present case, we have sufficient context to argue that the differences among taxa are largely due to different responses to a common history of changes in distribution of habitats. Paleodistribution modeling for each species may provide a more formalized way of expressing the ecological differences among taxa.

Limitations and Developments. The three key discrepancies between snail phylogeography and modeling (linkages between WU-TU and between LU-MT, and a KU refuge) are also evident in vertebrate phylogeographies, where resolution allows, indicating that the weakness lies with the modeling. In addition to the simplicity of the model used here, the climate estimates and spatial interpolations themselves have several limitations that may contribute to these discrepancies. Rainshadow effects may be behind the failure of the modeling to detect the WU-TU connection inferred from the genetic data, and whereas a KU refuge is not clearly within the BIOCLIM bounds, a larger area lays just below these limits. Other historical barriers such as rivers may also be important—MT currently is cut off from BK by the Mulgrave River. More sophisticated modeling systems that take into account climate parameter interactions, neighborhood size, and other nonlinear effects (reviewed in ref. 46) are effective in modeling snail current distribution (A.H. and A.M., unpublished results), and are being applied to modeling WT forest types (30); however, the main difficulty lies in developing paleoclimate estimates for these models. For reasons such as these, the combination of paleomodels and phylogeography is expected to be most informative at broad spatial scales and systems where simple parameters such as annual mean rainfall and temperature are highly informative.

Whereas our comparisons of modeling and genetics are largely qualitative, the results presented here are encouraging. Subregions predicted from the modeling (e.g., STABILITY surface, Fig. 1) to have contained refugia contained discrete phylogeographic lineages, whereas those from which the snail is predicted to have disappeared did not, but rather contained nested subsets of alleles as expected for recolonization from adjacent areas. There is also quantitative consistency between modeling and genetic estimates of the scale and tempo of population expansions from these refugia, in the correlation between area ratio and g, and plausible dependent parameters. However, each approach is laden with assumptions (e.g., area as a surrogate for *Ne* and panmixia and exponential growth in the genetic model), and the parameters obtained are not directly comparable. It would be far better to develop a framework in which the coalescent simulations on which genetic estimates of population parameters are based are combined directly with the spatially explicit predictions of the paleodistribution models.

We thank Michael Cunningham, Joanna Sumner, Conrad Hoskin, Chris Schneider, Keith McDonald, and Ben Phillips for collecting animals; Michael Cunningham, Stuart Baird, and Chris Schneider for discussion; and Catherine Graham, Leslie Rissler, Jim Patton, and David Wake for comments on the manuscript. This research was funded by grants from the Rainforest Cooperative Research Centre and the Australian Research Council.

1. Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. & Saunders, N. C. (1987) *Annu. Rev. Ecol. Syst.* **18,** 489–522.
2. Slatkin, M. (1987) *Science* **236,** 787–792.
3. Templeton, A. R., Routman, E. & Phillips, C. A. (1995) *Genetics* **140,** 767–782.
4. Hewitt, G.M. (2001) *Mol. Ecol.* **10,** 537–549.
5. Edwards, S. V. & DeBeerli, P. (2000) *Evolution* **54,** 1839–1854.
6. Avise, J.C. (2000) *Phylogeography: The History and Formation of Species* (Harvard Univ. Press, Cambridge, MA).
7. Cruzan, M. B. & Templeton, A. R. (2001) *Trends Ecol. Evol.* **15,** 491–496.
8. Hadly, E. A., Kohn, M. H., Leonard, J. A. & Wayne, R. K. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 6893–6896.
9. Leonard, J. A., Wayne, R. & Cooper, A. (2001) *Proc. Natl. Acad. Sci. USA* **97,** 1651–1654.
10. Nix, H.A. (1986) in *Atlas of Elapid Snakes of Australia*, ed. Longmore, R. (AGPS, Canberra, Australia).
11. Kershaw, A. P. & Nix, H. A. (1988) *J. Biogeography* **15,** 589–602.
12. Nix, H.A. (1991) in *Rainforest Animals: Atlas of Vertebrates Endemic to Australia's Wet Tropics*, eds. Nix, H. A. & Switzer, M. (ANPWS, Canberra, Australia), pp. 11–39.
13. Davis, M. B. & Shaw, R. G. (2001) *Science* **292,** 673–679.
14. Adam, P. (1992) *Australian Rainforests* (Clarendon, Oxford).
15. Truswell, E. (1993) *Aust. Syst. Botany* **6,** 533–557.
16. Kershaw, A. P. (1994) *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **109,** 399–412.
17. Webb, L. & Tracey, J. (1981) in *Ecological Biogeography of Australia*, ed. Keast, J. A. (W. Junk, The Hague, The Netherlands).
18. Nix, H. A. & Switzer, M. (1991) in *Rainforest Animals: Atlas of Vertebrates Endemic to Australia's Wet Tropics*, eds. Nix, H. A. & Switzer, M. (ANPWS, Canberra, Australia).
19. Williams, S. E., Pearson, R. G. & Walsh, P. J. (1996) *Pacif. Conserv. Biol.* **2,** 327–362.
20. Hopkins, M. S., Ash, J., Graham, A. W., Head, J. & Hewett, R. K. (1993) *J. Biogeogr.* **20,** 59–74.
21. Joseph, L., Moritz, C. & Hugall, A. (1995) *Proc. R. Soc. London Ser. B* **260,** 177–182.
22. Schneider, C. J., Cunningham, M. & Moritz, C. (1998) *Mol. Ecol.* **7,** 487–498.
23. Thomaz, D., Guiller, A. & Clarke, B. (1996) *Proc. R. Soc. London Ser. B* **263,** 363–368.
24. Douris, V., Cameron, R. A. D., Rodakis, G. C. & Lecanidou. R. (1998) *Evolution* **52,** 116–125.
25. Schneider, C. J., Smith, T. B., Larison, B. & Moritz, C. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 13869–13873.
26. Stanisic, J. (2000) *Mem. Qld. Mus.* **46,** 337–348.
27. Moritz, C., Richardson, K. S., Ferrier, S., Monteith, G. B., Stanisic, J., Williams, S. E. & Whiffin, T. (2001) *Proc. R. Soc. London Ser. B* **268,** 1875–1881.
28. Houlder, D. J., Hutchinson, M. F., Nix, H. A. & McMahon, J. P. (2000) ANUCLIM Users Guide (Australian National Univ., Canberra, Australia), Version 5.1.
29. Moss, P. T. & Kershaw, A. P. (2000) *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **155,** 155–176.
30. Hilbert, D. W. & Ostendorf, B. (2001) *Ecol. Modell.* **146,** 311–327.
31. Swofford, D. L. (2000) PAUP* (Sinauer, Sunderland, MA), Version 4.0b5.
32. McElroy, D., Moran, P., Bermingham, E. & Kornfield, I. (1992) *J. Hered.* **83,** 157–158.
33. Kuhner, M. K., Yamato, J. & Felsenstein, J. (1998) *Genetics* **149,** 429–434.
34. King, H. (1975) MSc. thesis (Univ. of Queensland, Brisbane, Australia).
35. Denver, D. R., Morris, K., Lynch, M., Vassilieva, L. L. & Thomas, W. K. (2000) *Science* **289,** 2342–2344.
36. Cunningham, M. & Moritz, C. (1998) *Biol. Conserv.* **83,** 19–30.
37. Pope, L. C., Estoup, A. & Moritz, C. (2000) *Mol. Ecol.* **9,** 2041–2053.
38. Chiba, S. (1999) *Evolution* **53,** 460–471.
39. Hayashi, M. & Chiba, S. (2000) *Biol. J. Linn. Soc.* **70,** 391–401.
40. Thacker, R. W. &. Hadfield, G. (2000) *Mol. Phys. Evol.* **16,** 263–270.
41. Brown, K. S. J. (1987) in *Biogeography and Quaternary History of Tropical America*, eds. Whitmore, T. C. & Brown, K. S. (Oxford Univ. Press, New York), pp. 175–196.
42. Colinvaux, P. A., De Oliveira, P. E., Moreno, J. E., Miller, M. C. & Bush, M. B. (1996) *Science* **274,** 85–88.
43. Schneider, C. J. & Moritz, C. (1999) *Proc. R. Soc. London Ser. B* **266,** 191–196.
44. Williams, S. E. & Pearson, R. G. (1997) *Proc. R. Soc. London Ser. B* **264,** 709–716.
45. Reid, C. A. M. & Storey, R. I. (2000) *Mem. Qld. Mus.* **46,** 253–297.
46. Guisan, A. & Zimmermann, N. E. (2000) *Ecol. Modell.* **135,** 147–186.

ANDREW HUGALL, JOHN STANISIC, AND CRAIG MORITZ

10

PHYLOGEOGRAPHY OF TERRESTRIAL GASTROPODS

The Case of the *Sphaerospira* Lineage and History of Queensland Rainforests

Phylogeography attempts to identify the mechanisms governing the geographic distributions of genealogical lineages (phylogenies) among and within closely related species (Avise 2000; Chapter 9, this volume), particularly the role of historical landscape processes. As such, it mirrors aspects of biogeography using species-level phylogenies (Nelson and Platnick 1981). Both disciplines require some way of describing the congruence or match to the historical landscape, typically using a comparative approach of identifying concordance among unrelated, codistributed taxa to infer common extrinsic historical forces. A next step is to link the spatial and temporal scales of intraspecific phylogeography to those of the underlying interspecific phylogeny. This can give insights into the geography and tempo of speciation (Barraclough and Nee 2001) and the connection between spatial patterns of phylogenetic diversity and species richness. For example, comparison of the age of phylogeographic groups with that of species has been taken as a measure of the tempo of allopatric speciation (Avise and Walker 1998). In this study, we go back to the case of a single phylogenetic series of taxa and endeavor to assess congruence or fit to landscape via bioclimatic distribution modeling, using a trans-species mtDNA phylogeny.

mtDNA PHYLOGEOGRAPHY IN LAND SNAILS

Terrestrial snails and slugs have strong potential to retain more historical signal of population structure—to be good "tracers of history"—because the combination of high persistence and low vagility results in numerous differentiated

populations at various levels of geographic and genetic isolation. What is the role of this isolation in the evolution of local and of regional species diversity? Here a phylogeographic framework spanning multiple species and including dense sampling of geographic diversity within species may enable partitioning of the relative importance of incidental divergence in allopatry, environmentally driven divergent selection, and parapatric processes in the development of regional and local diversity (Woodruff 1978; Schilthuizen and Gittenberger 1996; Moritz et al. 2000; Barraclough and Nee 2001; Pfenniger and Magnin 2001).

Numerous studies of presumed neutral genetic diversity in land snails have confirmed a strong link between geography and genetic diversity at very fine scales; examples include analysis of allozymes in *Albinaria corrugata* (Schilthuizen and Lombaerts 1994) and microsatellites in *Helix aspersa* (Arnaud et al. 1999, 2001) or *Cepaea nemoralis* (Davison and Clarke 2000). Such studies have also frequently encountered high levels of mtDNA diversity, often structured into several deeply divergent lineages (starting with Thomaz et al. 1996). Various hypotheses have been proposed about the source of this deeper structure; most of them focus on the role of historical vicariance processes in generating the structures and on the metapopulation structure necessary to retain them.

Untangling the more recent processes of the spread and sorting of ancestral diversity from those that created the diversity requires some reasonable alternative information about historical landscape processes. Therefore, more recent intensive regional-scale phylogeographic studies have used historical landscape frameworks for interpreting the distribution of the deeper level aspects of land snail mtDNA diversity. Studies include secondary contact after dispersal from centers of refuge, such as the postglacial expansion of European *Cepaea* (Davison 2000); river drainage patterns of *Discus* (Ross 1999); complex geological histories of the Kanto region of Japanese *Euhadra* (Hayashi and Chiba 2000; Watanabe and Chiba 2001); secondary contact between nascent *Helix* species in North Africa (Guiller et al. 2001); and habitat specialization in *Gyliotrachela* (Schilthuizen et al. 1999). However, some cases involving high levels of mixing, introgression, and possible cryptic species underscore the potential complexity of interpreting the diversity of land snail mtDNA (Ross 1999; Shimizu and Ueshima 2000; Pfenninger and Magnin 2001; Watanabe and Chiba 2001).

Scaling-up taxonomically, and looking for the link between geography and speciation, a lack of concordance between mtDNA phylogeny and species boundaries in *Partula* (Clarke et al. 1996; Goodacre and Wade 2001) and *Albinaria* (Schilthuizen and Gittenberger 1996; Douris et al. 1998) has been attributed variously to peripheral isolation, retention of ancestral polymorphism, and introgression among closely related species. Although these phylogenetic

studies have confirmed a geographic component in the gene phylogeny above the species level, they also show the folly of assuming the monophyly of island radiations (e.g., Crete: *Albinaria*, Douris et al. 1998; Society Island: *Partula*, Goodacre and Wade 2001; Hawaii: achatinellid tree snails, Thacker and Hadfield 2000).

VALUE OF BIOCLIMATIC MODELING FOR PHYLOGEOGRAPHY

Lest the genetic data be left to do all of the work—the gene tree versus species tree problem—it is best to reinforce molecular phylogeographies with independent, spatially explicit hypotheses about population structure; bioclimatic distribution modeling has some claim to providing this (Hugall et al. 2002). Both phylogeographic analysis and distribution modeling will work best in a situation where the climate governs the habitat, which is strongly structured by geography, and together these factors essentially govern the distribution of the organism. Bioclimatic distribution modeling also introduces the possibility of predicting the effects of historical climate change on the potential distribution—paleoclimate modeling (Nix and Kalma 1972; Nix 1991; Hilbert and Ostendorf 2001).

This combination of modeling and phylogeography is strongest in groups with a dense geographic distribution and a gradual dispersal, so that gaps in spatial and temporal distributions of lineages can be used to infer extinction. Because they can persist in small areas, phylogeographic structuring in land snails should also be informative about the spatial and temporal patterns of speciation. With phylogeography coupled to distribution modeling providing the spatial context, the molecular clock implicit in mtDNA phylogenies can provide the temporal aspect. This can be explored through analyses of lineage diversification models (Nee et al. 1994; Nee 2001) that can provide insights into the long-term balance between speciation and extinction. Altogether, these may allow identification of the speciation and extinction components in current species diversity.

BRIEF SUMMARY OF BIOGEOGRAPHIC STUDIES OF EAST AUSTRALIAN FORESTS

Hooker (1859) was the first to recognize the deep historical origins of the Australian biota as a mix of "Antarctic and tropical elements"—the Gondwanan element overlain by a more recent Southeast Asian component. Spencer (1896) began the organization of biogeographic regions, recognizing the northeastern "Torresian" region from the southeastern "Bassian." Snails were first applied to these bioregions by Iredale (1937), who identified a subtropical "Oxlyean" zone between the Torresian and the Bassian. Burbidge's (1960) landmark phytogeography recognized the special diversity of Wet Tropics and Border Ranges (the Iredale "Oxlyean"). In summarizing the distribution of birds, Keast (1961, 1981) identified four regions: Cape York (CY), Wet Tropics (WT), Mid-East Queensland (MEQ), and South East Queensland-Border Ranges (SEQ-NNSW) as centers of endemism and possible historical refuges (see Figures 10.1 and 10.2). Kikkawa and Pearse (1969) introduced analytical methods, identifying the Burdekin Gap separating the WT from southeastern Australia at species and genus levels. Schodde and Calaby's (1972) exegesis of birds cemented many of the names for the developing biogeography: Laura Gap, Burdekin Gap, Fitzroy Gap (also known as the St. Lawrence Gap), and the "Tumbunan" biota; the last recognizing the connection between eastern Australian and mid-montaine New Guinea rainforest avifauna. These pioneering works have been followed by cladistic approaches using various taxa (birds, lizards, frogs: Cracraft 1986; Phoracanthine beetles: Wang et al. 1996; plants: Crisp et al. 1995), but have also taken on a new perspective in the work of Henry Nix and colleagues in demonstrating the method and importance of a bioclimatic approach to the natural history of Australia. In 1972, Nix and Kalma introduced paleoclimate analysis to indicate the tenuous link between northeastern Australia and New Guinea rainforests. Nix (1982) formed a climate-based plant growth index classification and developed the concept of the mesothermal "archipelago" of closed forest along eastern Australia, culminating in the bioclimatic modeling of Pleistocene patterns of refugia in Wet Tropics rainforests (Nix 1991).

Joseph et al. (1993) and Joseph and Moritz (1994) provided the first molecular phylogeographic tests of these biogeographic scenarios, showing they were not just ecological, but also historical refuges. Intraspecific structuring in birds across the Burdekin Gap (the foremost biogeographic barrier) was dependent on ecology: rainforest birds showed a 5 to 7% divergence in the mtDNA cytochrome *b* gene, open forest birds did not. Subsequent studies of frog and lizard taxa introduced bioclimatic modeling (McGuigan et al. 1998) and comparative studies to phylogeography across parts of the region (James and Moritz 2000; Moritz 2000; Schäuble and Moritz 2001). These show a trend of increasing distinction among areas from south to north but with little overarching phylogenetic congruence, perhaps in part because of substantial differences in ecology (Schäuble and Moritz 2001). Interspecific phylogenetic studies of vertebrate taxa show old species with little regional diversification, which is interpreted as taxa dominated by extinction (reviewed in Moritz et al. 1997). An

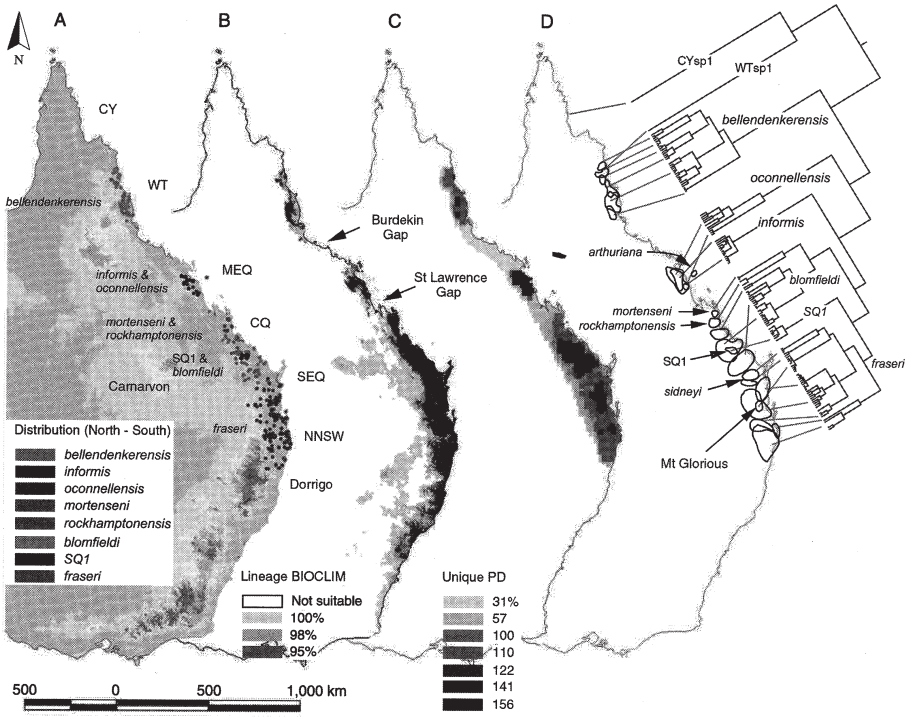


Figure 10.1. Species distribution points, bioclimatic model, and major phylogeographic units mapped onto eastern Australia. In this and other figures of eastern Australia, named biogeographic regions and barriers are indicated (for key, see text). (A) *Sphaerospira* lineage species known distribution points over DEM map. (B) *Sphaerospira* lineage bioclimatic model (current climate). (C) interpolated phylogenetic diversity (PD) from ultrametric tree of 24 major phylogeographic units (expressed as a percentage of tree height, 20-minute grid scale). (D) mtDNA haplotype tree linked to the approximate geographic distribution of the 24 phylogeographic units delimited by the marked 3% threshold. This tree is also shown in Figure 10.4.

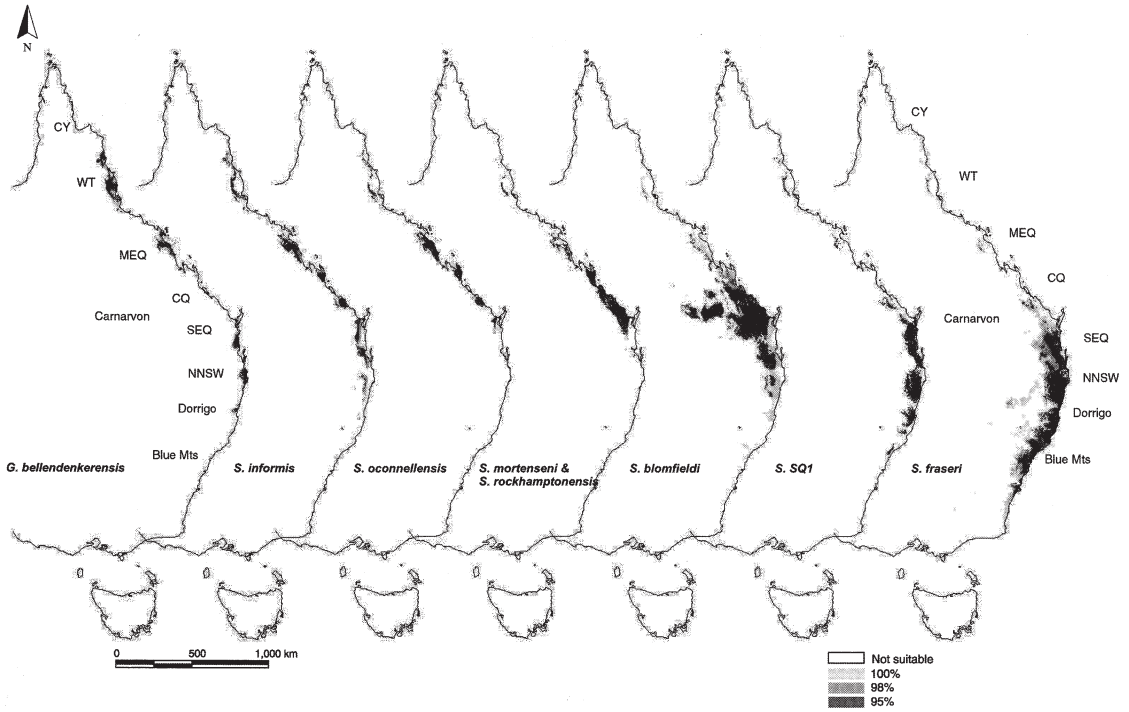


Figure 10.2. Bioclimatic models for seven species or groups based on three climate parameters (annual mean temperature, annual mean precipitation, and precipitation of the driest quarter) using minimum and maximum from the 100%, 98%, and 95% set of observed distribution points (see text). Minimum/maximum (100%) only for *S.* SQ1 and combined *S. mortenseni* and *S. rockhamptonensis* models. The models are arranged left to right for species from north to south.

exception to this pattern is the diversification of Carphodactyline geckos in the MEQ region (Couper et al. 2000); the comparative study of codistributed skinks by Stuart-Fox et al. (2001) strongly supports this pattern as regional endemicity driven by vicariance.

In summary, the history of the mosaic of forests along the east coast of Australia has been dominated by the long-term drying of the climate from the late Miocene on, leading to the decline and fragmentation of a formerly vast and rich rainforest habitat (Archer et al. 1991; Adam 1992; Truswell 1993). As they currently stand, major rainforest domains are restricted to the eastern areas of the Great Dividing Range, centered on a disjunct series of upland areas—an "archipelago" of cooler mesothermal climate. During the Pleistocene, these forests waxed and waned to an uncertain but considerable extent. For mammals and other large vertebrates, extinction caused by habitat contraction is a major theme, with most extant species being quite old (Moritz et al. 1997); however, to the land snail, this mosaic of forests may harbor multiple ongoing processes of vicariant and environmentally driven diversification.

The Australian Camaenidae are probably Miocene/post-Miocene immigrants from a large pool of Laurasian taxa—one of the Malesian elements (Solem 1979, 1997; but see Scott 1997), with the east coast dominated by the Camaeninae, including the Hadroid subgroup containing the "Sphaerospira facies." These large, often striped snails are distributed in a broad band down eastern Australia (see Figure 10.1). Here we apply biogeographic analyses to the trans-species mtDNA phylogeny of a snail lineage identified as mapping to eastern Australian rainforests. Specific questions for this *Sphaerospira* lineage include: Is the sympatric diversity due to filtering of ancestral diversity or in situ diversification? How does geographic population structure relate to speciation? What does this say about the biogeography of Queensland rainforests?

GNAROSOPHIA BELLENDENKERENSIS AS BENCHMARK

Gnarosophia bellendenkerensis is the Wet Tropics member of the *Sphaerospira* lineage. Previous detailed study of *G. bellendenkerensis* in the light of comparative phylogeography studies with vertebrates and bioclimatic modeling (Hugall et al. 2002) demonstrates that the distribution of mtDNA diversity maps a population structure formed by an interaction of climate and geography into patterns of contraction to and expansion from discrete refuges. The bioclimatic distribution modeling, by providing independent information on historical population structure, enables multiple processes that complicate the genetic data to be teased apart. Thus *G. bellendenkerensis* becomes our benchmark for understanding levels of genetic diversity down to a fine scale, and by reasonable extension, guides our interpretation of the larger-scale patterns observed across the entire *Sphaerospira* lineage.

TAXA INCLUDED

The taxa included in our discussion (including outgroups) are distributed across mesothermal, mesic forests in coastal eastern Australia, from Cape York in the north to the rainforests of northeastern New South Wales in the south. To describe the areas, we use a series of regional names established from previous biogeographic studies: Cape York (CY), Wet Tropics (WT), Mideast Queensland (MEQ), Central Queensland (CQ; Byfield and Shoalwater Bay area), Southeast Queensland (SEQ; Bulburin and Kroombit Tops to Main Range), and the McPherson (Border) Ranges and northeastern New South Wales (NNSW).

The lineage we studied, which included eight species and two subspecies, is defined as a monophyletic group in a larger molecular phylogenetic analysis of camaenid snails that encompasses 800 individuals from more than 200 taxa (A. Hugall, unpublished). This same analysis identified two undescribed taxa as outgroups: one taxa endemic to the isolated Iron Range rainforests of Cape York (CYsp1), and one taxa endemic to Thornton Peak within the Wet Tropics rainforests of northeastern Queensland (WTsp1). Sampling of this "*Sphaerospira*" lineage is complete in that all living species and subspecies are represented and geographic sampling is extensive for most taxa. Conversely, some species traditionally included within *Sphaerospira* are excluded because they lie phylogenetically outside the lineage. The taxonomy is from the latest zoological catalogue (Smith 1992) except where noted. All included taxa are terrestrial leaf litter/understory generalists.

The species concerned are *Gnarosophia bellendenkerensis* endemic to the WT rainforests; *Sphaerospira informis,* endemic to MEQ rainforest (the largest camaenid in Australia: 51.7 mm high; 52.1 mm wide; $n = 52$); *S. oconnellensis,* overlapping distribution and microsympatric with *S. informis,* and subspecies *S. o. arthuriana* on the adjacent Whitsunday Islands; *S. mortenseni* and *S. rockhamptonensis,* narrow endemics in central Queensland (CQ); *S. blomfieldi,* broadly distributed in vine scrub, wet sclerophyll, and rainforest across the CQ-SEQ region, and subspecies *S. b. sidneyi* at the southern limit; *Sphaerospira* sp., SQ1 endemic to the Dawes Range (Bulburin SF) in the SEQ region, microsympatric with *S. blomfieldi,* until recently considered as *S. fraseri* (Scott 1996); and *S. fraseri,* broadly distributed in mesic forest and rainforest across

SEQ-NNSW regions. *Sphaerospira fraseri* (Griffith and Pidgeon, 1833) is the type species of *Sphaerospira* (Mörch, 1867).

DISTRIBUTION MODELING

Locality records for bioclimatic modeling were obtained from the Queensland Museum database. The number of locations per species ranged from 315 distribution records for *S. fraseri* to 7 for the narrowly endemic *S. mortenseni* (Table 10.1). Because the number of records was small, *S. mortenseni* was combined with the geographically adjacent sister species, *S. rockhamptonensis,* for modeling (combined $n = 22$). The number of records per 20-minute grid was more consistent, ranging from 4.4 to 12.4. For modeling the total *Sphaerospira* lineage, 728 location records were used. Each species range was measured as the area within 0.25 degrees of each grid cell datapoint, falling within the coastline and within the 100% *Sphaerospira* lineage bioclimatic model.

Climate-based distribution modeling followed the BIOCLIM procedure (Nix 1986), which uses spatially interpolated estimates of annual mean temperature, annual mean precipitation, and precipitation of the driest quarter on a 36-second (approximately 1-km^2) resolution digital elevation model for the whole of Australia (ANUCLIM; Houlder et al. 2000). Bioclimatic models and map analyses were made using ARCVIEW GIS 3.2a ESRI, using only one record per 36-second grid cell per species.

The BIOCLIM procedure sets upper and lower limits for each climate layer based on the upper and lower limits of the observed distribution points, allowing for trimming of outliers. Distributions show absolute minimum and maximum (100%) and limits leaving out 2% and 5% of the observed distribution points at the upper and lower climate limits (98% and 95% models). This approach, although simple and based on few climate parameters, has proven highly effective for modeling current and paleodistributions of species and their rainforest habitats in eastern Australia (Nix 1991; McGuigan et al. 1998), including the Wet Tropics *G. bellendenkerensis* (Hugall et al. 2002).

Models were constructed for each species to estimate the extent of overlap of potential distributions and to visualize the biogeographic implication of physiological evolution. In addition, we combined models across species in two ways: (1) by summing the spatial overlap of individual models (±98% limits); areas present in all models representing a conserved distribution for the lineage; and (2) by modeling the distribution for all species simultaneously using the climate limits of the entire lineage. Quantitative comparisons of models and other map layers used map calculation functions in ARCVIEW.

Table 10.1
Genetic and distribution data for the *Sphaerospira* lineage bioclimatic modeling study

Species	Specimens	Locations /Species	Distribution Points	Max Dxy[a]	Π Species	Π Pops	Records per 20′ Cell	98% Model Area[b] (km^2)	100% Model Area[c] (km^2)	Area[d] (km^2)
G. bellendenkerensis	121	53	102	0.152/0.222	0.129	0.010	9.1	21,826	30,084	11,007
S. informis	17	9	75	0.037	0.024	0.008	6.3	27,268	27,268	7,391
S. oconnellensis	15	7	26	0.165/0.211	0.110	0.007	4.6	14,732	14,732	6,619
S. mortenseni	2	1	7	NA	NA	NA	12.4	24,270	24,270	2,898
S. rockhamptonensis	4	2	15	0.02	0.013	0.005	6.7	—	—	—
S. blomfeldi	27	10	116	0.143/0.249	0.091	0.011	5.4	86,790	108,828	27,436
S. SQ1	9	3	13	0.036	0.022	0.013	5.8	36,926	36,926	1,331
S. fraseri	43	22	315	0.162/0.233	0.113	0.016	7.2	86,955	121,293	55,283
S. b. sidneyi	5	2	50	0.11/	NA	NA	4.4	NA	NA	NA
S. o. arthuriana	1	1	9	NA	NA	NA	8.0	NA	NA	NA
Sister group	2	2	6	NA	NA	NA	NA	NA	NA	NA
Mean					0.072	0.011	6.3			
Total[e]	246	112	734					136,080	239,037	109,521

Note: Π Species = nucleotide diversity of entire species; Π Pops = average nucleotide diversity of single site populations within a species.

[a]Maximum divergence, by the COII Tamura-Nei model / alldata GTR-Γ MLK model.

[b]Modeled area using 98% percentile limits.

[c]Modeled area using 100% percentile limits, interpolated from points, 36-second cell = 1 km^2, *rockhamptonensis* and *mortenseni* combined.

[d]Best estimate of observed range, measured as area within 0.25 degrees from each datapoint, bounded by the coast and the full *Sphaerospira* lineage bioclimatic model.

[e]Total areas may not add because of overlap among models and ranges.

To predict potential distributions under paleoclimates, we used climate shifts estimated for the Wet Tropics region (Nix 1991; see also Hugall et al. 2002), and a modification of this (Nix, personal communication in Williams 1991) that allows for estimated latitudinal variation in paleoclimate parameters along the east Australian coast. To represent the climate ranges of the Pleistocene, we modeled two extremes: the restrictive cool and dry conditions of the last glacial maximum (18,000 years ago, annual mean temperature –3.0°C; annual mean precipitation 67% and precipitation of the driest quarter 75% of current monthly averages), and the more favorable cool-wet period (8,000 to 6,000 years ago: annual mean temperature –2.0°C; annual mean precipitation 120% and precipitation of the driest quarter 170% of current averages). The Williams model has the function –0.05°C per degree of latitude and precipitation layers reduced to 55 to 85% of current monthly averages weighted against summer rainfall. As in previous analyses, the modeled distributions are restricted to the current coastline.

MOLECULAR ANALYSES

DNA was extracted from either freshly collected snails or ethanol-preserved specimens from the Queensland Museum using the Chelex (Bio-Rad) method. All specimens were sequenced by ABI automated sequencing, most with both strands. The final ingroup dataset used here comprises 244 individuals from 110 localities (Table 10.1), for a total of 120 kb of sequence comprising COII, 16S rRNA, and 12S rRNA mtDNA. Sequencing was structured hierarchically. All individuals were sequenced for the most rapidly evolving gene, COII (480 sites), to identify tightly knit phylogeographic lineages. Nineteen individuals representing these phylogeographic lineages were also sequenced for 16S rRNA (450 sites), and a subset of nine of these was sequenced for 12S rRNA (520 sites) to further resolve branch lengths and relationships among the deepest lineages. Sequences were aligned in ClustalX (Thompson et al. 1994). Sequences are lodged with GenBank (accession numbers AY151055 to 151082 and AY151291 to AY151354).

We performed phylogenetic analyses using PAUP* 4.0b8 (Swofford 2000), lineage-through-time analyses using End-Epi v1.0 (Rambaut et al. 1997), tree manipulation using TreeEdit v1.0-a4.61 (Rambaut and Charleston 2000), divergences using the maximum likelihood (ML) model with ultrametric tree, and calculated diversities and divergences using REAP (McElroy et al. 1992). The ML model we used is derived from a larger tree of 66 taxa with all genes (1,450 sites), including the nine *Sphaerospira* lineage specimens with all genes. The model was chosen with reference to the gain in lnL as parameters were added, leading to an empirical base content, four substitution rate category, five median site rate category GTR-Γ model. The topology of this larger tree and the model framework provides a backbone for adding the further subset of taxa delimiting the major lineages within the group (19 taxa with COII and 16S), and then again for all haplotypes within the study lineage (164 haplotypes from 244 individuals, all with COII). With so many taxa, searching the ML tree space was not directly feasible. We therefore determined relationships among closely related haplotypes (within populations) with neighbor-joining and maximum parsimony. We then fixed these relationships and confined ML analyses to relationships between these phylogeographic groups. The wide divergence scale poses difficulties in developing an ultrametric tree. We used the complex site rate and substitution rate ML model to provide a best estimate of branch length for the all-taxa tree.

For lineage-through-time analyses, we generated an ultrametric tree using the ML tree and model, with clock assumption (MLK). This process was checked by comparing branch lengths of pruned trees with ML-modeled subset trees and comparing all gene trees with COII-only trees. The cost of imposing the clock is 88.9 ΔlnL—not considered significant given the number of nodes (using df = taxa – 2; Felsenstein 1981). Compared with COII-only trees, the all-data trees showed negligible difference at the tips of the branches and relative stretching of the tree length across the deeper nodes defining the position of the nearest outgroups and the primary splits within the ingroup.

We conducted lineage-through-time analyses both for all lineages and for recognized taxa only, including the subspecies *arthuriana* and *sidneyi*. These subset trees were pruned from the larger tree using PAUP* 4.0b8 and TreeEdit v1.0-a4.61. Log-linear plots were analyzed for deviation from linear with the Wilcoxon test in End-Epi v1.0, considering uncertainty in branch lengths estimated from nonparametric bootstrapping (Baldwin and Sanderson 1998).

PAUP* handles missing data in ML by calculating the lnL for all possible states. A state close to one for which there is data (a reasonably close relative with the additional genes) will score a higher likelihood, therefore contributing more than a state for which there is no close pattern—PAUP* essentially fills in the missing data with the most likely pattern. Therefore, such apparently large amounts of missing data—deadly for neighbor-joining distance methods—has little effect on likelihood estimations for both topology and branch length, unlike maximum parsimony analysis, which fails for the latter. The likelihood analysis of nested sequence data (incorporating several genes) provides a way of rationalizing the amount of sequence needed to be gathered for large trees with enough sites for accurate branch length estimation.

MOLECULAR DIVERSITY AND PHYLOGENY

Figure 10.1 provides an overview of the region, distribution, and three depictions of the genetic data: phylogeny, spatial phylogenetic diversity, and phylogeographic units. The phylogeny is shown again in Figure 10.3. The *Sphaerospira* lineage forms a discrete well-supported, well-resolved clade, within which each species is monophyletic (Figures 10.1 and 10.3), with the more widely distributed species composed of multiple, deeply divided phylogeographic lineages (see below for more detail). Nucleotide diversity within populations ranges from 0.001 to 0.032 (mean, 0.011; 24 populations with $n \geq 4$ individuals within a distance of 0.25 degrees), whereas nucleotide diversity within species ranges from 0.013 to 0.129 (mean 0.072), with maximum intraspecific mtDNA divergence up to 0.25 using the MLK model (Table 10.1).

Broadly, the *Sphaerospira* lineage forms a phylogenetically nested series from north to south spanning some 2,000 km and 15 degrees of latitude. Each of the few instances of sympatry combines one lineage with the northernmost representative of a more widely distributed sister clade; examples are WTsp1 (an outgroup) and *G. bellendenkerensis* in WT rainforests; *Sphaerospira oconnellensis* and *S. informis* in MEQ; and *S. blomfieldi* and *S.* SQ1 in SEQ. The northernmost member, *G. bellendenkerensis,* comprises a series of discrete phylogeographic lineages distributed across upland areas of the WT rainforests (see Hugall et al. 2002 for detail). Next to the south is MEQ in which *S. informis* and *S. oconnellensis* each show phylogeographic structure separating the Clarke Ranges (including Eungella) from the southern end of the MEQ region, the Connors Range. Although the divergence is shallow in *S. informis* (<4%), *S. oconnellensis* has a deeply divergent lineage (16%). The two widely distributed species from the south, *S. blomfieldi* (northern SEQ) and *S. fraseri* (southern SEQ to NNSW), consist of multiple, mostly discrete phylogeographic lineages, but in the latter we start to see more geographic overlap between lineages (Figure 10.1). Each of these species has allopatric sister taxa located to the north: *S.*

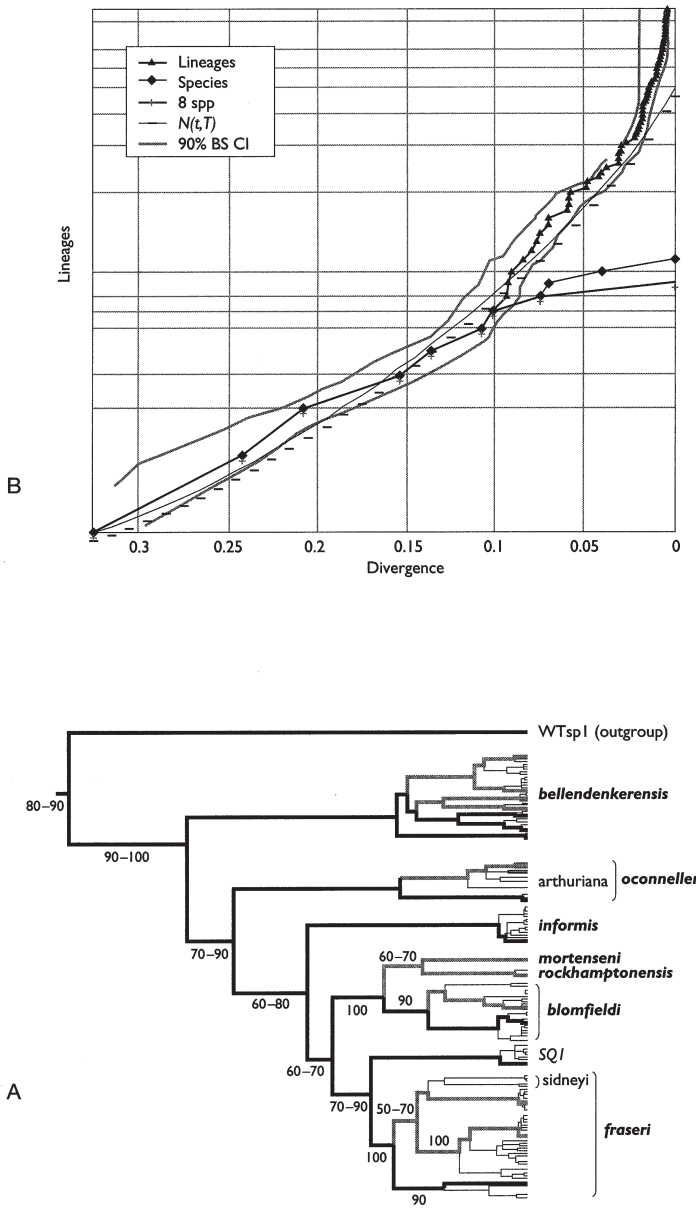


Figure 10.3. mtDNA haplotype molecular clock tree with lineage-through-time-plot. (A) Combined data likelihood model ultrametric tree of 139 ingroup haplotypes, with taxonomic designations. Heavy lines denote branches inferred from all three genes (1,450 sites); lighter lines denote two genes (930 sites); and black light lines denote COII only (480 sites). For clarity, only 139 of 164 haplotypes are shown, the remainder all fall within the most recent nodes. Range of bootstrap values from various methods and combinations of taxa are shown for some nodes. This is the same tree as in Figure 10.1. (B) Log lineage-through-time plot showing both all lineages (triangles, up to a ceiling of 100) and 10 recognized taxa (diamonds, pruned from lineage tree). Dashed lines are 95% CI bounds from bootstrap resampling; bars are *N(t,T),* the expected theoretical *b/d* equilibrium plot from estimate $b = 0.28$ and $d = 0.16$ (from Equation 2 in Harvey et al. 1994). The bar demarks 3% divergence phylogeographic threshold.

mortenseni and *S. rockhamptonensis* adjacent to *S. blomfieldi,* and *S.* SQ1 to the north of *S. fraseri*.

This intraspecific mtDNA diversity shows considerable geographic structuring down to small scales (e.g., *G. bellendenkerensis*); however, we will limit discussion of geographic patterns of mtDNA lineages to those above the level encountered within populations (samples from within a few grid cells), using phylogeographic structure in *G. bellendenkerensis* as a benchmark. Thus in these qualitative descriptions of the phylogeography, a threshold of about 3% divergence (along the molecular clock tree) delimits 24 geographically coherent groups, indicated on Figures 10.1 and 10.3. This phylogeographic scale is appropriate to the sampling we used and to the biogeographic scope of our investigations; study of finer-scale patterns below this level requires more targeted sampling.

BIOCLIMATIC MODELING

Not surprisingly, the potential ranges predicted by paleoclimatic modeling closely reflect the distribution of mesic mesothermal forests on the east coast (individual species in Figure 10.2, combined analyses in Figures 10.1 and 10.4). The model for each species predicts discrete zones, typically coastal wet forests, where other species are found. Most individual species models predict relatively small areas (Table 10.1), except those for *S. blomfieldi* and *S. fraseri,* which predict much larger ranges that include extensive areas not actually occupied to the south and the west (south to central NSW and west to the Carnarvon Ranges). The ratio of actual to modeled ranges for species ranges from 12 to 63% (within 98% bioclimatic limits) and is 80% for the whole lineage model. Quantitatively, model overlap ranges from zero (*S. blomfieldi* and *G. bellendenkerensis*) to 48% (mean 13%; Table 10.2). However, qualitatively, each species model predicts suitable habitat across all of the major biogeographic regions: WT, MEQ, CQ, and SEQ.

These models are spatial manifestations of subtle shifts in environmental envelopes spanning observed current distribution points, presumably reflecting a combination of physiological tolerances and available environmental space, throughout the evolutionary history of this complex. The general picture is one of conserved environmental limits, but with the southern species, *S. blomfieldi* and *S. fraseri,* occupying somewhat drier and cooler habitats, respectively. This overall conservatism, and the departures from it, is reflected in the combined analyses. The total lineage analysis encompasses the individual models and connects the coastal regions from CQ to NNSW, but the Burdekin and

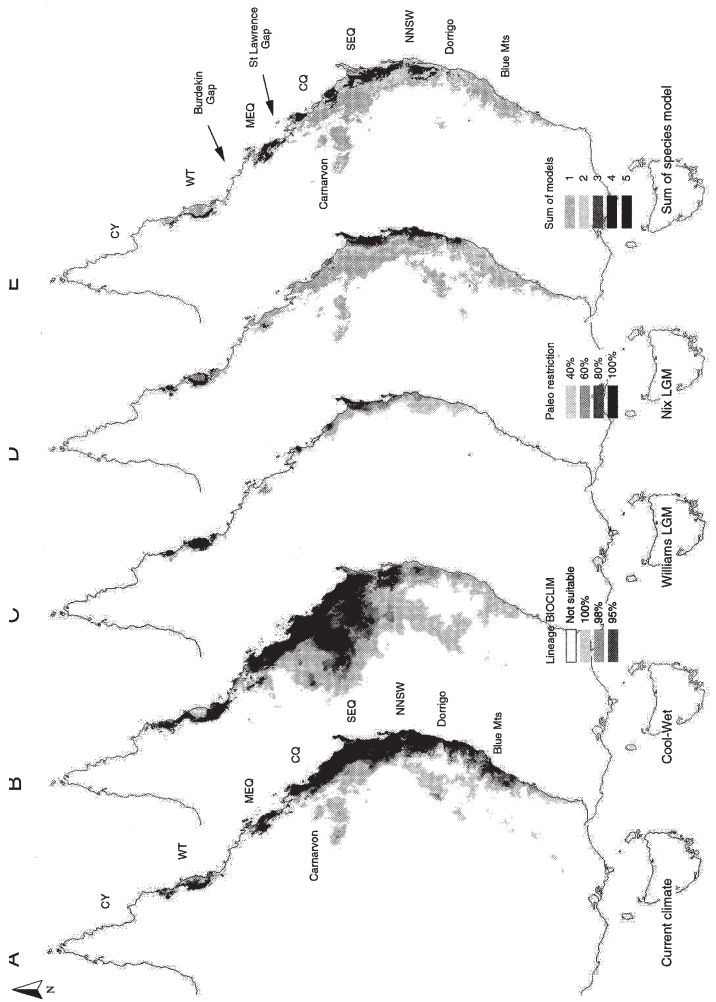


Figure 10.4. Combined lineage bioclimatic paleodistribution models, with sum of individual species model. (A) *Sphaerospira* current climate bioclimatic distribution model. (B) Cool wet scenario model (annual mean temperature [AMT] –2.0°C, annual mean precipitation [AMP] 120%, precipitation of the driest quarter [PDQ] 170%). (C) The Williams (1991) last glacial maximum (LGM) model (for SEQ this equates to AMT –3.0°C, AMP 67%, PDQ 75%). (D) The Nix (1991) wet tropics last glacial maximum model indicated as percentage of the full (100%) last glacial maximum scenario (AMT –3.5°C, AMP 50%, PDQ 60%). (E) Sum of the seven individual species current climate models shown in Figure 10.2.

Table 10.2
Model and range overlap for the *Sphaerospira* lineage bioclimatic modeling study

	bellendenkerensis	*informis*	*oconnellensis*	CQ taxa	*blomfieldi*	SQ1	*fraseri*
bellendenkerensis	**0.50**	—	—	—	—	—	—
informis	0.20	**0.27**	—	—	—	—	—
oconnellensis	0.14	0.48	**0.45**	—	—	—	—
CQ taxa[a]	0.04	0.21	0.25	**0.12**	—	—	—
blomfieldi	0.00	0.05	0.04	0.21	**0.32**	—	—
SQ1	0.11	0.23	0.03	0.07	0.07	**0.04**	—
fraseri	0.06	0.09	0.00	0.02	0.00	0.33	**0.64**

Note: Overlap between models is expressed as overlap-to-sum ratio. Actual-to-modeled area on diagonal (≥98%). Ranges measured as area within 0.25 degrees of each datapoint, bounded by the coast and the full *Sphaerospira* lineage bioclimatic model.

[a]*S. rockhamptonensis* and *mortenseni* combined.

St. Lawrence gaps remain. The sum of the models (Figures 10.1 and 10.4) highlights the conserved component of the bioclimatic profiles across the whole lineage, these being focused on the wet forests of WT, MEQ, CQ, and SEQ-NNSW, but excluding the wettest areas (eastern WT). When compared with the total lineage model, the sum model serves to highlight the major geographic disjunctions within the range of the *Sphaerospira* lineage, these being the Burdekin Gap between WT and MEQ and the St. Lawrence Gap between MEQ and CQ, in contrast to the relative continuity across SEQ-NNSW.

The last glacial maximum paleoclimatic models for the total lineage predict highly restricted distributions within each of WT, MEQ, CQ, and SEQ-NNSW, with separate predicted refugia in the north and south of WT and northern and central regions in SEQ-NNSW. The most obvious feature of the paleomodels is the elimination of the large extensions to the west and south, these representing substantial overpredictions in the current climate models of *S. blomfieldi* and *S. fraseri,* respectively. Relative to the Wet Tropics last glacial maximum paleoclimate scenario (Nix 1991), the model based on the Williams (1991) paleoclimate parameters predicts more extensive areas in WT and CQ, but greater restriction in SEQ-NNSW. The difference between the actual and modeled ranges for the lineage model corresponds to range contractions predicted in the last glacial maximum paleomodel; the modeled area is reduced by 79%, accounting for 94% of the difference between actual and (current) modeled ranges. At the other extreme, the cool-wet scenario (Figure 10.4) predicts increased connectivity between MEQ and southern areas and indicates that drier areas to the west became suitable, but does not extend the predicted range south of NNSW. Even with these permissive conditions, the Burdekin Gap remains. Preliminary explorations (not shown) suggest that it takes a further 2°C temperature decline and a marked reduction in seasonality to bridge this gap, and even more to connect the present CY rainforests.

The sum model and the last glacial maximum paleomodel are similar in picking out centers for WT, MEQ, two within CQ (Shoalwater and Bulburin/Kroombit), SEQ, and NNSW. Taken together, the lineage, sum, and paleomodels provide a map of biogeographic regions, their size and stability, and a type of hierarchy of the degree of connection among them: (WT, (MEQ, (SW, (CQ, (SEQ, NNSW))))). These correspond to the general pattern observed in the phylogeny.

RELATIONSHIP OF BIOCLIMATIC MODELS TO PHYLOGENY

Bioclimatic models of the whole clade, and the sum across taxa, highlight the major disjunctions in the phylogeny. The primary split between the observed ranges of *G. bellendenkerensis* and the other taxa corresponds to the Burdekin Gap and the secondary split between MEQ and south—the St. Lawrence or Fitzroy gap—separating the MEQ taxa *S. informis* from the southern group. This latter region, from Shoalwater south, is more or less continuous in the lineage bioclimatic model, but shows some structuring in the individual species, the sum model, and also in the more restrictive last glacial maximum paleoclimate scenario. The northern extremity of this region forms a narrowly connected bioclimatic domain containing the restricted endemics *S. mortenseni* and *S. rockhamptonensis,* sister group to the widely distributed *S. blomfieldi,* which itself is split into several phylogeographic units. The remainder is split into the CQ (Bulburin) endemic species *S.* SQ1, and its sister species *S. fraseri,* which is widespread further to the south in SEQ-NNSW, and is again dissected into multiple mtDNA sublineages. The bioclimatic models of *S. fraseri* and *S.* SQ1 are similar (33% overlap), with *S.* SQ1 occupying an isolated zone to the north of the main area of models correlating with the range of *S. fraseri* (Figure 10.2). Of the subspecies, *S. o. arthuriana* is distributed on the offshore island of the MEQ region (the Whitsunday Islands) and can be described as a sublineage within the phylogeographic structure of *S. oconnellensis* (but note the relatively limited sampling); and *S. b. sidneyi* forms two related sublineages within the *S. fraseri* mtDNA phylogeographic structure, distributed at the northern extremity of this range and at the edge of the core bioclimatic model.

DISTRIBUTION OF PHYLOGENETIC DIVERSITY

For a quantitative approach to the biogeographic distribution of genetic diversity encompassing the whole phylogeny, we turn to the analyses of the distribution of phylogenetic diversity (Faith 1994), shown in Figures 10.1 and 10.5. The phylogenetic diversity of individual areas is estimated as the sum of branch lengths of lineages within that area, expressed as the percentage of the total height of the tree (rooted with WTsp1 = 0.325). This is represented spatially in two ways: (1) a geographic map of the phylogenetic diversity interpolated from the 244 genetic data points onto 20-minute grids (Figure 10.1; here we consider only the phylogeographic units, i.e., lineages with less than 3% sequence divergence), and (2) as a latitudinal transect (Figure 10.5). At this scale, most grids have genetic data; however, where there is no genetic sampling but a species is (probably) present (known distribution and/or predicted in the bioclimatic model), the value of the surrounding points is interpolated. Where the lineage is absent (never found and/or outside modeled range), only the ancestral internodes can be interpolated (this assumes the presence of ancestors, i.e., steady fine-scale dispersal, some time in the past, a not unreasonable assumption for land snails at this spatial scale). A strict geographic pattern would have only one tip lineage at each point and so the phylogenetic diversity value is the height of the tree. Elevations above this level are due to (1) overlap of tip lineages (e.g., another species and secondary contact), and/or (2) internodes from the overlap of clades—this can be thought of as a measure of ancestral dispersal. Phylogenetic diversity values below the height of the tree imply extinction.

Although 44% of the actual geographic range has phylogenetic diversity greater than the basal value (tree height), only 10% of the total phylogenetic diversity is due to overlap (sympatry) of lineages. The tip component of overlap is due to sympatric species in MEQ and CQ (Bulburin); secondary contact (partial overlap) in WT, SEQ, and NNSW; and a Mount Glorious lineage of *S. fraseri*. Little overlap is due to internodes; across the whole, the ordering of relationships is such that only 3% can be ascribed to ancestral dispersal. In the more complex CQ-NNSW region (among some 14 phylogeographic clades), only 2% of phylogenetic diversity is due to historical dispersal. The principal components of this are due to relationships among *S. blomfieldi* and *S. fraseri* phylogeographic units (maximum overlap phylogenetic diversity 8% to 13%) and implied presence of SQ1 lineage (maximum 10%), but this is slight compared with the component due to tip lineages (presence of SQ1 at Bulburin; overlap of lineages in *S. fraseri* in SEQ, maximum increase 41, or 22%). With the caveat of limited sampling at the western edge of *S. fraseri,* the complete

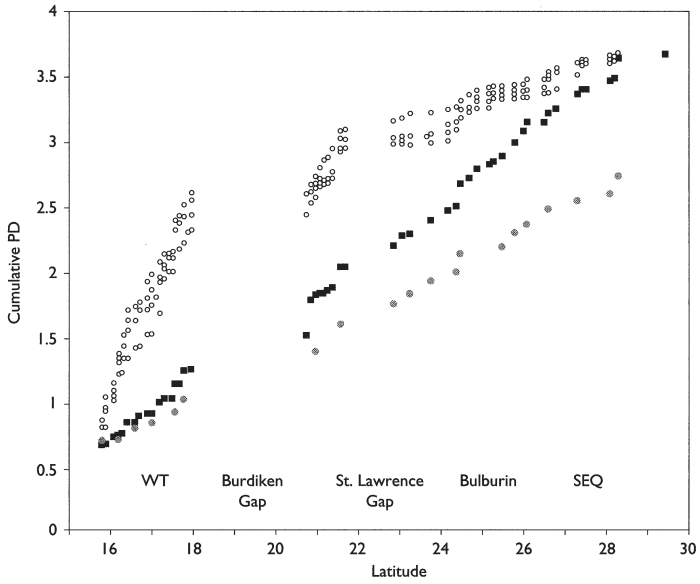


Figure 10.5. mtDNA phylogenetic diversity (PD) accumulation across a latitudinal transect from north to south. All haplotypes (black squares) and 24 phylogeographic groups (grey dots) are from haplotype MLK tree (rooted with WTsp1), per 0.1 degree of latitude. Major biogeographic areas and barriers are indicated. The randomized distribution plot is also indicated (white dots). See text for details.

overlap of lineages on the D'Aguilar Range (in particular Mount Glorious, see Figure 10.1) is less easily dismissed as secondary contact.

If vicariance fragments an otherwise fairly uniform distribution, uniform accumulation of divergence through time (geographically evenly spread) leads to a constant phylogenetic diversity-to-area relationship. If gaps in current geographic distribution represent local extinction of preexisting phylogenetic diversity, we should see departures from linearity, whereas if lineages dispersed to adjacent refugia (overlap), this phylogenetic diversity is maintained, preserving the underlying historical pattern. To examine the consequence of potential extinction of lineages in the climatically unsuitable areas between WT and MEQ (Burdekin Gap) and between MEQ and CQ (St. Lawrence Gap), we plotted cumulative phylogenetic diversity from north to south (Figure 10.5). Nominally the line can take any monotone, and random distributions should give hyperbolic curves. The observed pattern is close to linear (slope = 0.23

phylogenetic diversity per degree of latitude, $R^2 = 0.99$). Furthermore, using only the 24 phylogeographic clades produces a similarly linear relationship. This linearity has two aspects: (1) the phylogenetic diversity needed to span the large Burdekin Gap is found in MEQ as two sympatric lineages, and the phylogenetic diversity gained in the sympatric species of northern SEQ corrects for a slight slump crossing the St. Lawrence Gap (from regression residuals not shown). (2) The within-region intraspecific phylogeographic component of the phylogenetic diversity has the same function as that between regions (= species). This linearity reflects the uniform accumulation of diversity across an essentially intact ancestral geographic system; the implication here is that there has been no large-scale removal of preexisting phylogenetic diversity (i.e., extinction of local clades within the current large gaps in the range).

TEMPO OF DIVERSIFICATION: LINEAGE-THROUGH-TIME ANALYSES

The modeling provides a perspective on climatic limitations on distributions under current conditions and across the climatic extremes of the Quaternary, but does not directly address the longer time scale of speciation implied by the levels of sequence divergence in the molecular phylogeny. Here we further explore this temporal aspect of diversification, in particular, the relationship of phylogeographic groups to speciation, using lineage-through-time analyses (Harvey et al. 1994; Nee et al. 1994). This method uses an ultrametric tree to plot the log of lineage diversity through time and compares the observed trends with those expected with a constant birth-death (*b–d,* speciation-extinction) process or a constant birth (speciation) process (Nee 2001). This approach requires dense sampling across lineages. Our mtDNA phylogeny traverses the species level. Among species and major phylogeographic groups, sampling is complete; among local populations, the sampling becomes more sparse. Molecular phylogenies only show extant lineages. A key consequence is that lineage accumulation patterns reflect both birth *and* extinction rates. Put simply, the slope from the base of the phylogeny represents net diversification (*b–d*), whereas the most recent lineages can have a steeper slope dominated by birth rates alone. It is in this framework that changes in speciation/extinction rates (e.g., episodic speciation or mass extinction) must be evaluated.

In the following analysis, we contrast the temporal pattern of lineage accumulation considering only recognized taxa (species only, $n = 8$; or species + subspecies, $n = 10$) versus lineage accumulation that includes the 24 phylogeographic lineages as well as recognized taxa. Inclusion of the latter is

prompted by the hypothesis (Avise and Walker 1998) that phylogeographic units represent "species-in-waiting." Fine-scale analysis in *G. bellendenkerensis* indicates that Holocene scales of population expansion from refugia affect only the extreme tips of the tree (Hugall et al. 2002), well within the 3% threshold, and do not contribute to the shape of the lineage-through-time plot.

The tempo of reconstructed cladogenesis along our ultrametric tree is shown in Figure 10.3B. Two levels of diversity have been plotted: all of the observed lineages (the "haplotype tree"), and a phylogeny of recognized taxa only, pruned from the haplotype tree. For the taxon tree, we included both 8- and 10-taxa plots, the latter includes the two subspecies. The 3% divergence threshold is shown. Qualitatively, the plot appears triphasic. From left (basal) to right (recent) the following elements are represented: (1) a region from the root up to around the split between phylogroup and taxon plots (corresponding to maximum intraspecific mtDNA divergence), which is linear and corresponds to a net cladogenesis rate of one lineage formed per 8% divergence (or 0.12 births per lineage per percentage divergence); (2) an intermediate section, between the point of maximum intraspecific divergence and approximately the 3% threshold broadly separating phylogeographic units and within-lineage diversity, which has a steeper slope (Wilcoxon sign test, $p < 0.05$, including bootstrap variance) corresponding to a rate of one lineage formed per 3.4% sequence divergence (or 0.28 births per lineage per percentage divergence); and (3) the most recent segment of less than 3% divergence corresponding mostly to within-population diversity and which is expected to conform to coalescent expectations of population demographics—a birth-minus-death (*b–d*) process, in this case consistent with a constant population size (endemic transform Wilcoxon test, End-Epi v1.0).

Using the concomitant species/lineage line as a measure of *b–d* and estimating *b* from the slope of the midsection (phylogroups) yields the following estimates of lineage birth and death rates: $b = 0.28$, $d = 0.16$ (per lineage per percentage divergence), the birth-to-death (b/d) ratio is 1.7 (56% extinction rate), and a net *b–d* rate of one birth per lineage per 0.08 divergence (±10% from bootstrap variance in slope). At levels of sequence divergence <3%, the slope (*b*) trebles and gives a b/d ratio approaching 1, corresponding to intraspecific genetic drift/mutation equilibrium.

The divergence of lineage-through-time plots around the boundary between species and phylogroups indicates a change in tempo. The species curve is linear up to 8 or 9 lineages, then tapers somewhat. Given the hypothesis that "speciation" will take time, we do not necessarily expect new taxa to arise within a short time period relative to this net "speciation" rate. Therefore, including the subspecies *S. o. arthuriana* and *S. b. sidneyi,* the taxon curve is very close to the

mtDNA phylogeographic unit *b/d* expectations. Excluding the drift/fixation process within phylogroups (i.e., below the 3% threshold), we have 24 lineages, 10 species or subspecies, with equilibrium expectation of 11 taxa. In this model interpretation, the phylogenetic structure of species diversity (including sympatric species) can be maintained by 20% of phylogeographic units becoming fully independent lineages or recognized species, at a tempo of one per lineage per 0.08 mtDNA divergence time.

IMPLICATIONS OF MOLECULAR CLOCK RATES ON THE AGE OF PHYLOGEOGRAPHIC BREAKS

Although we do not have a reference for absolute time, the relative ages of the different process levels in the tree can be reconciled with interpretations of rainforest history. This in turn gives us some limits to clock rates for the MLK tree. If the ultrametric transform is effective, the depth of the tree, and depth at various levels within it (e.g., the local phylogeographic groups, species), constrains relative times of these levels and forms a temporal and biogeographic reference. For the group to have been shaped by the biogeographic forces outlined in the beginning of this chapter, the tree is bounded by Pleistocene/Holocene ages toward the tips and, more approximately, by a mid-Miocene limit for the deeper nodes, this last representing the time at which the major east coast rainforest blocks are thought to have become isolated (Adams 1992).

Applying recently postulated molecular rates of 10% per million years (Chiba 1999; Hayashi and Chiba 2000; Thacker and Hadfield 2000) places the 3% phylogroup at 0.3 million years, the deeper phylogeographic units (the speciation rate gauge of 0.08 divergence) at 0.8 million years, and the base of the group at 2.4 million years. Given the persistence of the Burdekin Gap to climatic ranges postulated for the Quaternary, and its presence in supraspecific biogeographic analyses, these dates strain credibility. Alternatively, a rate of 2% per million years (per lineage) corresponds to 2, 4, and 12 million years. The latter scenario is more consistent with the time scales usually considered for major events in the biogeography of these wet forest communities.

In biogeography, the issues of how to define regions—the geographic scale at which speciation and extinction shape biodiversity (Rosenzweig 1995; Hubbell 2001)—and how to account for these being different for different taxa (Nelson and Platnick 1981; Crisp et al. 2001)—may be quantitatively addressed by integrating distributional modeling and molecular phylogeny. The modeling and the spatial analysis of phylogenetic diversity (spatial phylogenetic diversity) provide readily integrated perspectives on ecogeographic structure and endemicity. For lineages with high levels of allopatric diversity (e.g., terrestrial snails), different spatial and temporal scales of diversification can be linked via a trans-species phylogenetic analysis.

What are the spatial and temporal scales of extinction and speciation in this system? Here the spatial scales are derived from modeled paleoclimate and actual ranges combined with intraspecific phylogeography, and the temporal scales are derived from lineage-through-time analyses and consilience of phylogeny with preexisting biogeographic inference. Spatial phylogenetic diversity is an attempt to integrate the spatial and temporal information. For this group, the phylogenetic diversity map is relatively simple. With more lineages, more complex phylogenetic diversity landscape patterns can be expected in this temporal endemicity index, thus allowing identification of associated landscape characteristics via quantitative comparisons of phylogenetic diversity and actual or modeled species diversities using GIS. For example, a "museum" of ancestral diversity could have a large phylogenetic diversity–to–species diversity ratio, whereas regions of geographic or adaptive radiations with many young endemics would have a low phylogenetic diversity–to–species diversity ratio. In the current analysis, CQ (Shoalwater/Byfield area) and Kroombit/Bulburin (northern SEQ) are both bioclimatic (ecological) refuges (Figure 10.4) and both contain two species (CQ, *S. mortenseni* and *S. rockhamptonensis;* Kroombit/Bulburin, *S.* SQ1 and *S. blomfieldi*). However, CQ has a much lower phylogenetic diversity (Figure 10.1) and, correspondingly, is less clearly identified as a historical refuge (Figure 10.4).

SHAPE OF VICARIANCE: PHYLOGENY AS BIOGEOGRAPHY

The monophyletic *Sphaerospira* lineage maps to and elaborates on the previously proposed biogeography of the mesic forests of the east Australian mesothermal archipelago. Distribution modeling highlights the discrete nature of these environmental zones and identifies core areas and connections, encapsulated in the lineage and sum species models. The spatial phylogenetic diversity analysis highlights the match of this environmental domain with the phylogeny. The distribution of sister lineages and taxa, in context with the distribution models, suggests that the principal mode of development of these sister lineages is via peripheral vicariant isolation. In each case of the more recently diverged taxa (i.e., *S. rockhamptonensis* and *S. mortenseni, S.* SQ1, *S. o. arthuriana,* and *S. b. sidneyi*), narrowly endemic taxa represent geographically and bioclimatically peripheral isolates of a more widely distributed sister lineage.

Allopatric divergence and vicariance are most clearly evident at the intraspecific phylogeographic level, but what of the entire phylogeny? Broadly,

there are two interpretations of the phylogeny: (1) static, comprising a series of historically unique species with essentially ephemeral phylogeographic diversity, or (2) an emergent property of a continuous process of diversification. The first would interpret the lineage-through-time dynamic as different birth rates for events associated with the formation and distribution of the primary lineages (= species resident in the core refugia) and for phylogeographic structuring within these dating from a later era of fragmentation. The second treats it as indicating the balance of a continuous environmentally driven vicariant birth/death process. It would be difficult to argue either way were it not for the additional information in the form of the spatial context.

The lineage-through-time model can be taken as either high birth rates *or* high extinction rates for phylogeography relative to species. The *b/d* model unifies these in one process: gradual allopatric divergence with recognizable species emerging from phylogeographic lineages at a rate of approximately one per 8% mtDNA sequence divergence. Not all phylogeographic units of a species survive to form separate species (only 20%); most fade away either through physical extinction or genetic swamping, dwindling down over time to only the one lineage. *G. bellendenkerensis, S. oconnellensis,* and *S. informis* are good examples; each has a very long internode. Most if not all extinction is at the tips, but this does not mean that there was no extinction in the past, just as it does not mean there were no phylogeographic groups in the past.

In lineage-through-time methods, the more clade-specific the extinction, the less visible it is in the reconstructed phylogeny; the more evenly distributed across the phylogeny extinction is, the more visible (Harvey et al. 1994). In *Sphaerospira* this would mean, respectively, the geographically localized elimination of an old lineage versus global contraction and elimination of sublineages within regions. In turn, these scenarios could equate to (1) the deep divisions between lineages separated by the localized Burdekin and St. Lawrence gaps, and (2) strong phylogeographic structure within regions. Although the lineage-through-time analyses may not detect the extinction processes in the first, localized loss of basal lineages should cause sharp reductions in phylogenetic diversity in regions away from historical refugia. However, in the present case, the consistency of phylogenetic diversity down the east coast transect (Figure 10.5) suggests that there has been little loss of ancestral diversity, and that the underlying pattern of phylogenetic diversity is similar to the within-region phylogeographic pattern.

Although the distribution of diversity is the result of a series of specific events—the actual history—the feature of the *b/d* process is that it accounts not just for the recent phylogeography but also for the deeper lineages and the spatial pattern of phylogenetic diversity. Our conclusion is that the *Sphaerospira* lineage is as near to being a perfect vicariant system as nature can devise. In this view, the formation of the phylogeny is contemporaneous with the environmental history; the geography, as embodied in the distribution models, continuously prunes the phylogeny so that it cannot have anything other than a precise fit. This dominating *b/d* process includes the sympatric species: historically *S. oconnellensis* and *S. informis* were sister taxa and *S.* SQ1 and *S. blomfieldi* were sister taxa—that is, their ancestral manifestations were; and to account for filling the gaps in the spatial phylogenetic diversity analysis, *S. oconnellensis* probably started as a phylogeographic unit somewhat to the north of its current distribution, as did *S. blomfieldi*.

BIOGEOGRAPHIC IMPLICATIONS

The potential of snails to provide a fine-grained perspective on the historical biogeography of their habitats is well realized in this study. The combination of spatial phylogenetic diversity and bioclimatic models produces an integrated assessment of historical biogeography, emphasizing more than most previous studies have the endemicity contained within MEQ and historical subdivisions of the wet forests to the south. Of particular interest is the Bulburin/Kroombit area, the separation of NNSW, and the phylogenetic diversity richness of the central SEQ region. The phylogeny and lineage-through-time patterns highlight the considerable age of this fine-grained biogeography.

The phylogeny has spatial and temporal consilience with the proposed biogeography of the region (e.g., the influence of the Burdekin Gap) and provides the first substantial phylogenetic evidence for an effective St. Lawrence (also known as the Fitzroy) Gap, thus the distinction between MEQ and wet forests to the south. Indeed MEQ has never been recognized in any analytical biogeography (cf. Cracraft 1986; Crisp et al. 1995). The MEQ rainforests are supposed to have suffered severe reduction with substantial extinction of taxa through the Pleistocene, as evidenced by the absence of several rainforest-specialist vertebrates that are found to the north and south (WT and SEQ/NNSW; reviewed in Joseph et al. 1993; Low 1993). For the *Sphaerospira* lineage, the high phylogenetic diversity and phylogeographic congruence with endemic leaftail geckos (*Phyllurus;* Couper et al. 2000) suggests that long-term persistence of multiple (albeit small) refugia within MEQ has promoted, as well as preserved, diversity (Stuart-Fox et al. 2001).

Although effectively no previous phylogeographic analysis addresses historical subdivision across the St. Lawrence Gap, some studies traverse the extended SEQ bioregion (including CQ to NNSW as defined here). Of these, the consistent feature is of deepening structure/isolation northward, with three of five species showing a major break between Kroombit/Bulburin and the rest

(Moritz 2000). The pattern in the *Sphaerospira* lineage shows that, like the MEQ region to the north, the CQ (Shoalwater, Bulburin/Kroombit) region is not just a refuge, but is also a source of endemic diversification, and so can be argued to be a biogeographic region in its own right. Further south, the combination of overlapping tips and internodes gives a substantial peak in PD in the D'Aguilar Range. Could this be a sign of nascent speciation and/or admixture in the *S. fraseri* complex, a parallel to the processes (sympatric species) causing the PD peaks in CQ and MEQ?

SPECIATION

The phylogeny and distribution patterns (both actual and bioclimatic models) provide spatial and temporal support for the proposal that local peripheral isolation is the primary process driving diversification. The current intraspecific phylogeography provides a calibration of this ongoing process that has shaped virtually the entire phylogeny. Combined with limited evolution in ecology (inferred from climatic profiles), this indicates that ecological divergence comes after allopatric divergence, with limited local (alpha) diversity obtained secondarily. Although subsequent evolutionary processes leading to speciation, ecological divergence, and lineage overlap (alpha diversity) are not analyzed here, this framework provides avenues for further investigation. For example, we can ask what governs the tempo of morphological change: in particular, the relative importance of age of a refuge, its degree of geographic isolation, and stability of environment.

In addition to drift in allopatry, two other possible mechanisms for the evolution of ecomorphological divergence and sympatric species diversity are (1) environmentally driven divergent selection among isolated units, and (2) reinforcement or displacement processes leading up to secondarily derived sympatry (Rice and Hostert 1993). These processes may be assessed by a comparative analysis of morphological environmental variables across the phylogeny in the context of genetic divergence, spatial overlap, and modeled environmental change, thus encompassing population and species categories available in the *Sphaerospira* trans-species phylogeography.

ACKNOWLEDGMENTS

Thanks to Adnan Moussalli and Catherine Graham for modeling advice, Stuart Baird for discussion of likelihood analyses, the Museum of Vertebrate Zoology for GIS resources and general hospitality, and Mike Lee for comments on the manuscript as well as financial support. This research was supported by a grant from the Australian Research Council to C. Moritz and by Rainforest CRC.

REFERENCES

Adam, P. 1992. Australian Rainforests. Clarendon Press, Oxford.

Archer, M., S. J. Hand, and H. Godthelp. 1991. Riversleigh: The Story of Animals in Ancient Rainforest of Inland Australia. Reed Books, Balgowlah, NSW, Australia.

Arnaud, J. F., L. Madec, A. Bellido, and A. Guiller. 1999. Microspatial genetic structure in the land snail *Helix aspersa* (Gastropoda: Helicidae). Heredity 93:110–119.

Arnaud, J. F., L. Madec, A. Guiller, and A. Bellido. 2001. Spatial analysis of allozyme and microsatellite DNA polymorphisms in the land snail *Helix aspersa* (Gastropoda: Helicidae). Molecular Ecology 10:1563–1576.

Avise, J. C. 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge, MA.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics 18:489–522.

Avise, J. C., and D. Walker. 1998. Pleistocene phylogeographic effects on avian populations and the speciation process. Proceedings of the Royal Society of London B 265:457–463.

Baldwin, B. G., and M. J. Sanderson. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proceedings of the National Academy of Sciences, U.S.A. 95:9402–9406.

Barraclough, T. G., and S. Nee. 2001. Phylogenetics and speciation. Trends in ecology and evolution 16:391–399.

Burbidge, N. T. 1960. The phytogeography of the Australian region. Australian Journal of Botany 8:75–210.

Chiba, S. 1999. Accelerated evolution of land snails *Mandarina* in oceanic Bonin islands: Evidence from mitochondria DNA sequences. Evolution 53:460–471.

Clarke, B., M. S. Johnson, and J. Murray. 1996. Clines in the genetic distance between two species of island land snails: How "molecular leakage" can mislead us about speciation. Philosophical Transactions of the Royal Society of London 351:773–784.

Couper, P. J., C. J. Schneider, C. J. Hoskin, and J. A. Covacevich. 2000. Australian leaftail geckos endemic to eastern Australia: A new genus, two new species, and other new data. Memoirs of the Queensland Museum 45:253–265.

Cracraft, J. 1986. Origin and evolution of continental biotas: Speciation and historical congruence within the Australian avifauna. Evolution 40:977–996.

Crisp, M. D., S. Laffan, H. P. Linder, and A. Monro. 2001. Endemism in the Australian flora. Journal of Biogeography 28:183–198.

Crisp, M. D., H. P. Linder, and P. H. Weston. 1995. Cladistic biogeography of plants in Australia and New Guinea: Congruent pattern reveals two endemic tropical tracks. Systematic Biology 44:457–473.

Davison, A. 2000. An East-west distribution of divergent mitochondrial haplotypes in British populations of the land snail, *Cepaea nemoralis* (Pulmonata). Biological Journal of the Linnean Society 70:697–706.

Davison, A., and B. Clarke. 2000. History or current selection? A molecular analysis of "area effects" in the land snail *Cepaea nemoralis*. Proceedings of the Royal Society of London. B 267:1399–1405.

Douris, V., R. A. D. Cameron, G. C. Rodakis, and R. Lecanidou. 1998. Mitochondrial phylogeography of the land snail *Albinaria* in Crete: Long-term geological and short-term vicariance effects. Evolution 52:116–125.

Faith, D. P. 1994. Phylogenetic pattern and the quantification of organismal biodiversity. Philosophical Transactions of the Royal Society of London 345:45–58.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. Journal of Molecular Evolution 17:368–376.

Goodacre, S. L., and C. M. Wade. 2001. Patterns of genetic variation in Pacific island land snails: The distribution of cytochrome *b* lineages among Society Island *Partula*. Biological Journal of the Linnean Society 73:131–138.

Guiller, A., M. Coutellec-Vreto, L. Madec, and J. Deunff. 2001. Evolutionary history of the land snail *Helix aspersa* in the Western Mediterranean: Preliminary results inferred from mitochondrial DNA sequences. Molecular Ecology 10:81–87.

Harvey, P. H., R. M. May, and S. Nee. 1994. Phylogenies without fossils. Evolution 48:523–529.

Hayashi, M., and S. Chiba. 2000. Intraspecific diversity of mitochondrial DNA in the land snail *Euhadra peliomphala* (Bradybaenidae). Biological Journal of the Linnean Society 70:391–401.

Hilbert, D. W., and B. Ostendorf. 2001. The utility of artificial neural networks for modeling the distribution of vegetation in past, present and future climates. Ecological Modeling 146:311–327.

Hooker, J. D. 1859. Introductory essay. Pp. i–cxviii in Botany of the Antarctic Voyage of H.M. Discovery Ships *Erebus* and *Terror* in the Years 1839–1843. Vol. 3, Flora Tasmaniae. Reeve, London.

Houlder, D. J., M. F. Hutchinson, H. A. Nix, and J. P. McMahon. 2000. ANUCLIM Users Guide v. 5.1, CRES, ANU, Canberra.

Hubbell, S. P. 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Monographs in Population Biology 32, Princeton University Press, Princeton and Oxford.

Hugall, A. F., C. Moritz, A. Moussalli, and J. Stanisic. 2002. Reconciling paleodistribution models and comparative phylogeography in the wet tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). Proceedings of the National Academy of Sciences U.S.A. 99:6112–6117.

Iredale, T. 1937. A basic list of the land Mollusca of Australia. Australian Journal of Zoology 8:287–333.

James, C. H., and C. Moritz. 2000. Intraspecific phylogeography in the sedge frog *Litoria fallax* (Hylidae) indicates pre-Pleistocene vicariance of an open forest species from eastern Australia. Molecular Ecology 9:349–358.

Joseph, L., and C. Moritz. 1994. Mitochondrial DNA phylogeography of birds in eastern Australian rainforests: First fragments. Australian Journal of Zoology 42:385–403.

Joseph, L., C. Moritz, and A. Hugall. 1993. A mitochondrial perspective on the historical biogeography of mideastern Queensland rainforest birds. Memoirs of the Queensland Museum 34:201–214.

Keast, J. A. 1961. Bird speciation on the Australian continent. Bulletin of the Museum of Comparative Zoology Harvard 123:303–495.

Keast, J. A. 1981. The evolutionary biogeography of Australian birds. Pp. 1585–1635 in Ecological Biogeography of Australia (J. A. Keast, ed.). W. Junk, The Hague.

Kikkawa, J., and K. Pearse. 1969. Geographical distribution of land birds in Australia: A numerical analysis. Australian Journal of Zoology 17:821–840.

Low, T. 1993. Last of the rainforests: Rainforest refuges of the Mackay region. Wildlife Australia 4:18–21.

McElroy, D., P. Moran, E. Bermingham, and I. Kornfield. 1992. REAP: The restriction enzyme analysis package. Journal of Heredity 83:157–158.

McGuigan, K., K. McDonald, K. Parris, and C. Moritz. 1998. Mitochondrial DNA diversity and historical biogeography of a wet forest-restricted frog (*Litoria pearsoniana*) from mid-east Australia. Molecular Ecology 7:175–186.

Moritz, C. 2000. A molecular perspective on the conservation of diversity. Pp. 21–34 in The Biology of Biodiversity (M. Kato, ed.). Springer-Verlag.

Moritz, C., L. Joseph, M. Cunningham, and C. Schneider. 1997. Molecular perspectives on historical fragmentation of Australian tropical and subtropical rainforest: Implications for conservation. Pp. 442–454 in Tropical Rainforest Remnants: Ecology, Management, and Conservation of Fragmented Communities (W. F. Laurance and R. O. Bierregard, eds.). University of Chicago Press, Chicago.

Moritz, C., J. L. Patton, C. J. Schneider, and T. B. Smith. 2000. Diversification of rainforest faunas: An integrated molecular approach. Annual Review of Ecology and Systematics 31:533–563.

Nee, S. 2001. Inferring speciation rates from phylogenies. Evolution 55:661–668.

Nee, S., E. C. Holmes, R. M. May, and P. H. Harvey. 1994. Extinction rates can be estimated from molecular phylogenies. Philosophical Transactions of the Royal Society of London B 344:77–82.

Nelson, G., and N. Platnick. 1981. Systematics and Biogeography: Cladistics and Vicariance. Columbia University Press, New York.

Nix, H. A. 1982. Environmental determinates of biogeography and evolution in Terra Australis. Pp. 47–66 in Evolution of the Flora and Fauna of Arid Australia (W. R. Barker and P. J. M. Greenslade, eds.). Peacock Publications, Frewville, South Australia.

Nix, H. A. 1986. A biogeographic analysis of Australian elapid snakes. Pp. 4–15 in Atlas of Elapid Snakes of Australia (R. Longmore, ed.). AGPS, Canberra.

Nix, H. A. 1991. Biogeography: Patterns and process. Pp. 11–39 in Rainforest Animals: Atlas of Vertebrates Endemic to Australia's Wet Tropics (H. A. Nix and M. Switzer, eds.). ANPWS, Canberra.

Nix, H. A., and J. D. Kalma. 1972. Climate as the dominant control in the biogeography of northern Australia and New Guinea. Pp. 61–92 in Bridge and Barrier: The Natural and Cultural History of Torres Strait (D. Walker, ed.). Australian National University, Canberra.

Pfenninger, M., and F. Magnin. 2001. Phenotypic evolution and hidden speciation in *Candidula Unifasciata* spp. (Helicellinae, Gastropoda) inferred by 16S variation and quantitative shell traits. Molecular Ecology 10:2541–2554.

Rice, W. R., and E. E. Hostert. 1993. Laboratory experiments on speciation: What have we learned in 40 years? Evolution 47:1637–1653.

Rambaut, A., and M. Charleston. 2000. TreeEdit: Phylogenetic Tree Editor, v. 1.0-a4.61, University of Oxford.

Rambaut, A., P. H. Harvey, and S. Nee. 1997. End-Epi: An application for inferring phylogenetic and population dynamical processes from molecular sequences. CABIOS 13:303–306.

Rosenzweig, M. L. 1995. Species diversity in space and time. Cambridge University Press, Cambridge.

Ross, T. K. 1999. Phylogeography and conservation genetics of the Iowa Pleistocene snail. Molecular Ecology 8:1363–1373.

Schäuble, C. S., and C. Moritz. 2001. Comparative phylogeography of two open forest frogs from eastern Australia. Biological Journal of the Linnean Society 74:157–170.

Schilthuizen, M., and E. Gittenberger. 1996. Allozyme variation in some Cretan *Albinaria* (Gastropoda): Paraphyletic species as natural phenomena. Pp. 301–311 in Origin and Evolutionary Radiation of the Mollusca (J. D. Taylor, ed.). Oxford University Press, Oxford.

Schilthuizen, M., and M. Lombaerts. 1994. Population structure and levels of gene flow in the Mediterranean land snail *Albinaria corrugata* (Pulmonata: Clausiliidae). Evolution 48:577–586.

Schilthuizen, M., J. J. Vermeulen, G. W. H. Davison, and E. Gittenberger. 1999. Population structure in a snail species from isolated Malaysian limestone hills, inferred from ribosomal DNA sequences. Malacologia 41:283–296.

Schodde, R., and J. H. Calaby. 1972. Climate as the dominant control in the biogeography of northern Australia and New Guinea. Pp. 257–300 in Bridge and Barrier: The Natural and Cultural History of Torres Strait (D. Walker, ed.). Australian National University, Canberra.

Scott, B. J. 1996. Systematics, phylogeny and biogeography of the larger camaenid land snails of eastern Queensland rainforests (Pulmonata: Stylommatophora: Camaenidae). Ph.D. thesis, Department of Zoology, James Cook University, Townsville, Australia.

Scott, B. J. 1997. Biogeography of the Helicoidea (Mollusca: Gastropoda: Pulmonata): Land snails with Pangean distribution. Journal of Biogeography 24:399–407.

Shimizu, Y., and R. Ueshima. 2000. Historical biogeography and interspecific mtDNA introgression in *Euhadra peliomphala* (the Japanese land snail). Heredity 85:84–96.

Smith, B. J. 1992. Non-marine Mollusca. Pp. 110–174 in Zoological Catalogue of Australia (W. W. K. Houston, ed.). Australian Government Printing Service, Canberra.

Solem, A. 1979. Land-snail biogeography: A true snail's pace of change. Pp. 197–221 in Vicariance Biogeography: A Critique (G. Nelson and D. E. Rosen, eds.). Columbia University Press, New York.

Solem, A. 1997. Camaenid land snails from Western and Central Australia (Mollusca: Pulmonata: Camaenidae) VII. Taxa from Dampierland through the Nullabor. Records of the Western Australian Museum Supplement 50:1461–1899.

Spencer, W. B. 1896. Summary. Pp. 137–199 in Report on the Work of the Horn Scientific Expedition to Central Australia Part 1 (W. B. Spencer, ed.). Dulau, London.

Stuart-Fox, D. M., C. J. Schneider, C. Moritz, and P. J. Couper. 2001. Comparative phylogeography of three rainforest-restricted lizards from mid-east Queensland. Australian Journal of Zoology 49:119–127.

Swofford, D. L. 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.

Thacker, R. W., and M. G. Hadfield. 2000. Mitochondrial phylogeny of extant Hawaiian tree snails (Achatinellinae). Molecular Phylogenetic and Evolution 16:263–270.

Thomaz, D., A. Guiller, and B. Clarke. 1996. Extreme divergence of mitochondrial DNA within species of pulmonate land snails. Proceedings of the Royal Society of London B 263:363–368.

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

Truswell, E. 1993. Vegetation changes in the Australian Tertiary in response to climatic and phytogeographic forcing factors. Australian Systematic Botany 6:533–557.

Wang, Q., I. W. B. Thornton, and T. R. New. 1996. Biogeography of the Phoracanthine beetles (Coleoptera: Cerambycidae). Journal of Biogeography 23:75–94.

Watanabe, Y., and S. Chiba. 2001. High within-population mitochondrial DNA variation due to microvicariance and population mixing in the land snail *Euhadra quaesita* (Pulmonata: Bradybaenidae). Molecular Ecology 10:2635–2646.

Williams, J. E. 1991. Biogeographic patterns of three sub-alpine eucalypts in south-east Australia with special reference to *Eucalyptus pauciflora* Sieb. Ex Spreng. Journal of Biogeography 18:223–230.

Woodruff, D. S. 1978. Evolutionary and adaptive radiation of *Cerion:* A remarkably diverse group of West Indian land snails. Malacologia 17:223–239.

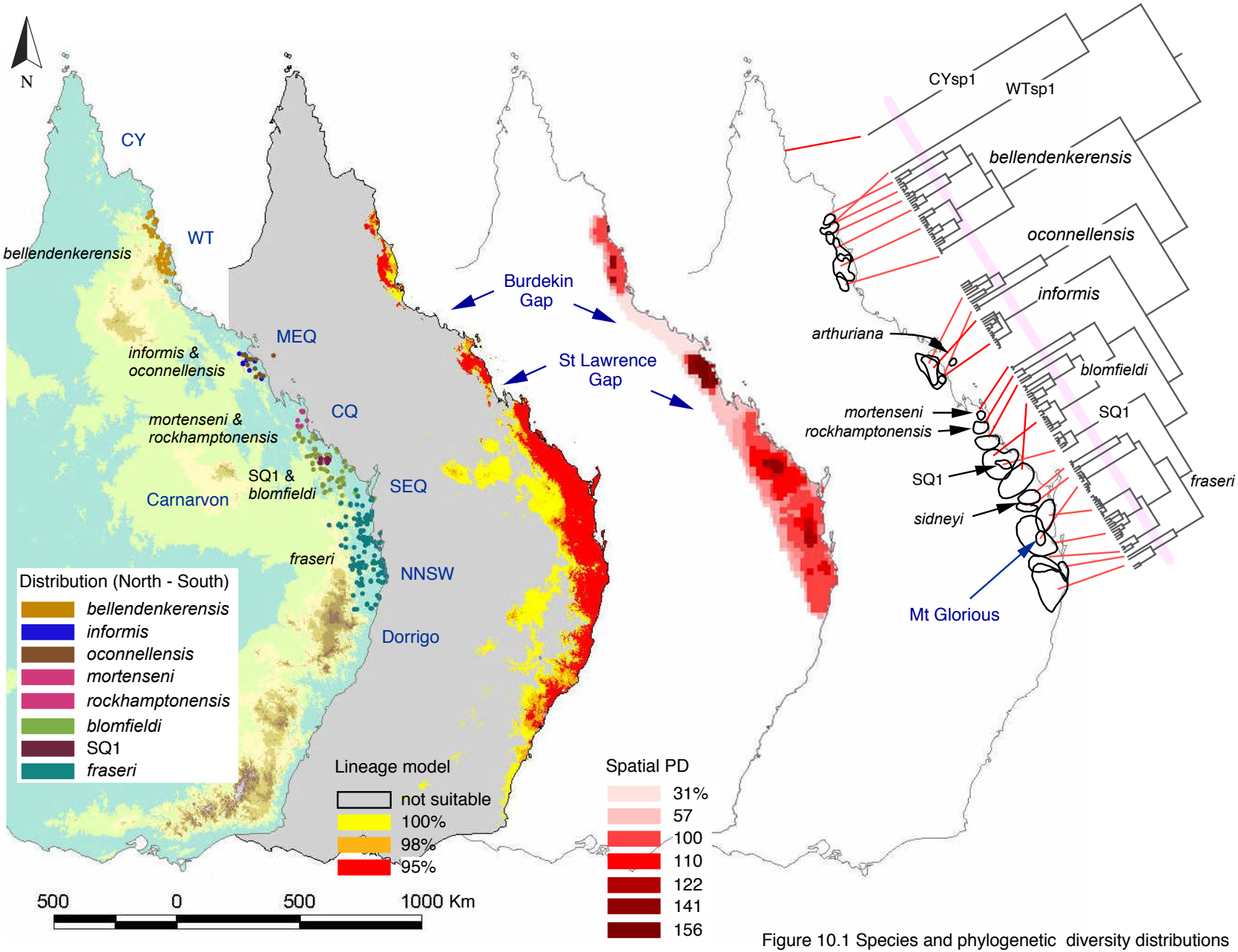


Figure 10.1 Species and phylogenetic diversity distributions

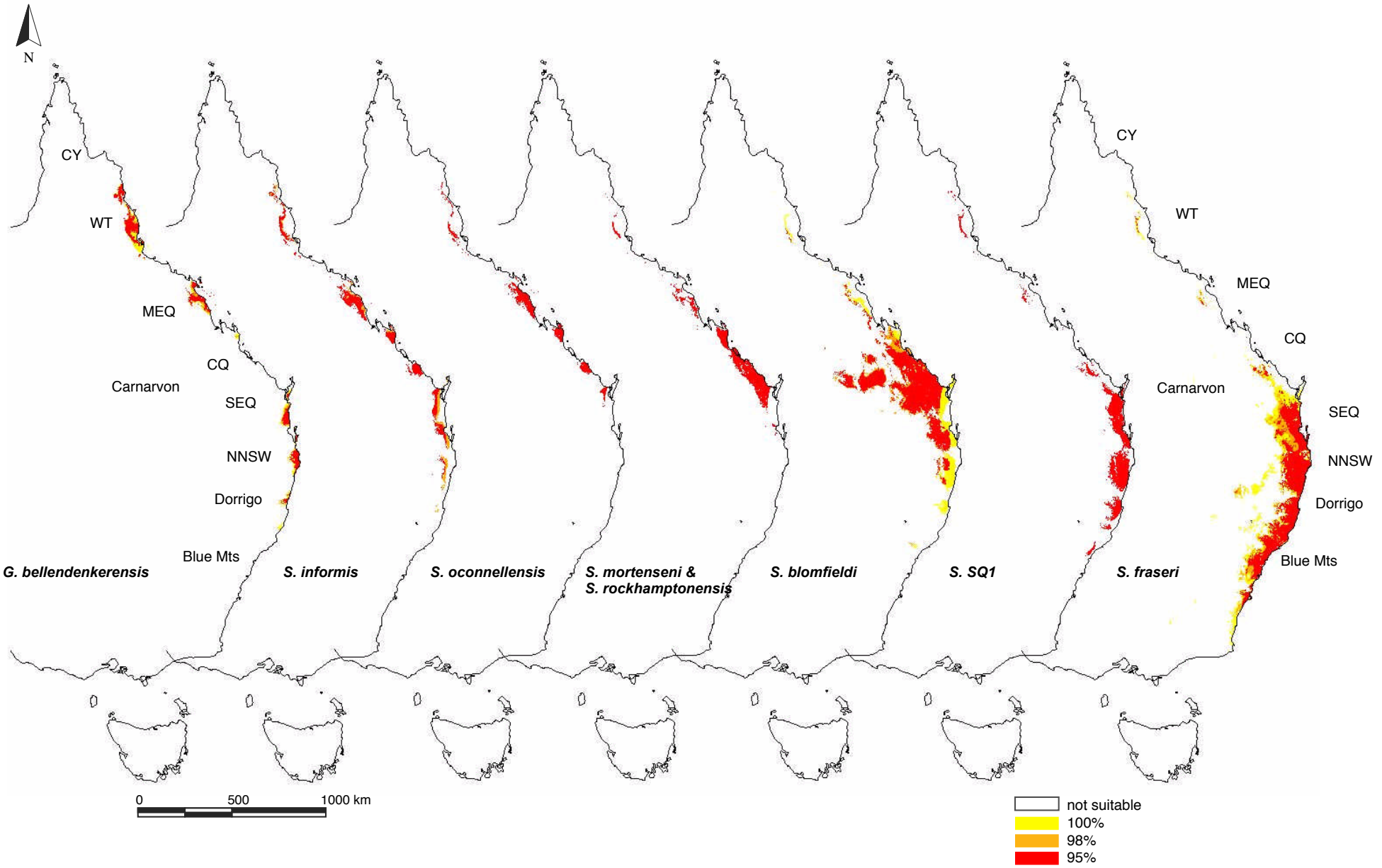


Figure 10.2 BIOCLIM models for seven species/groups based on AMT, AMP and PDQ. 36" DEM grid scale.

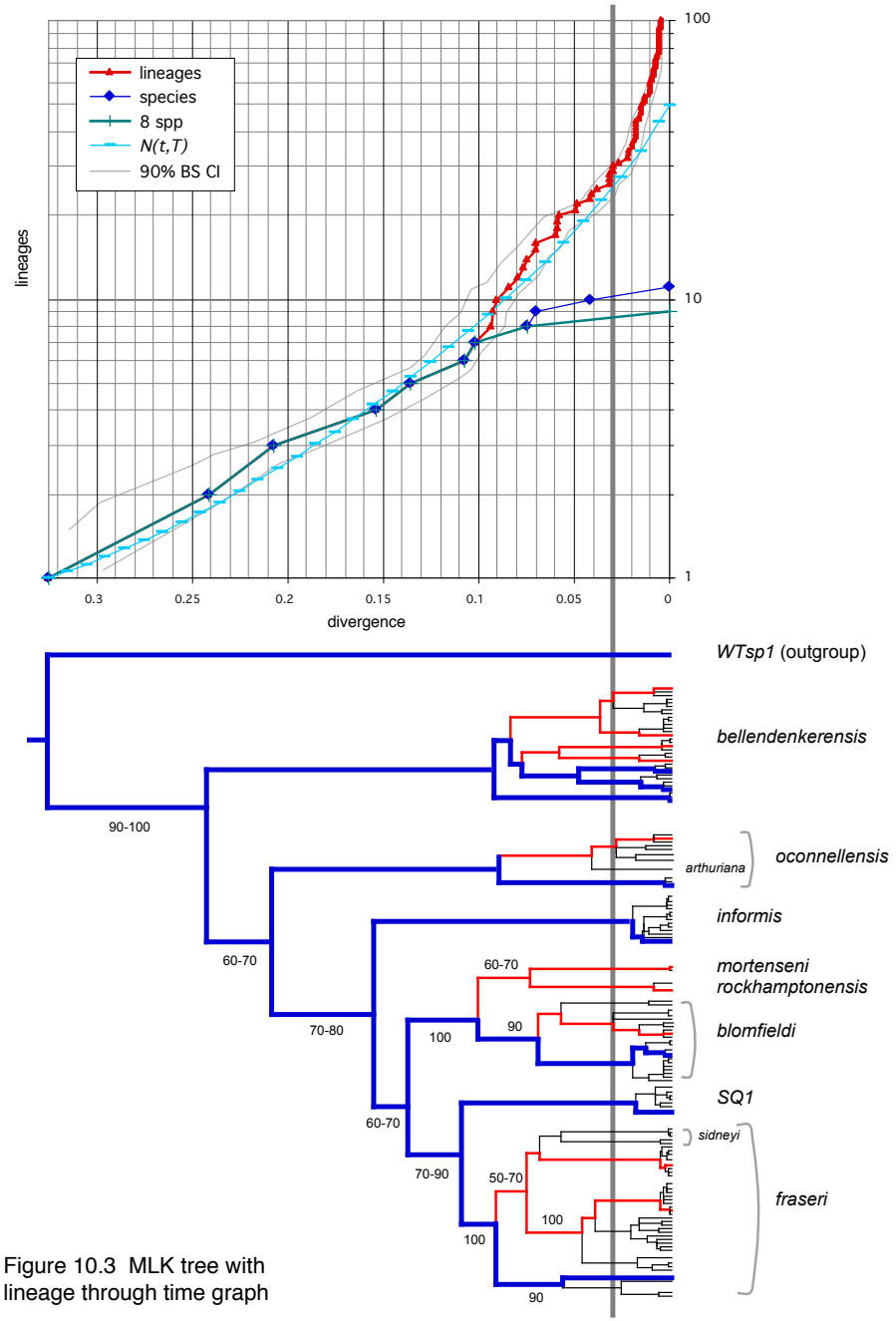


Figure 10.3 MLK tree with lineage through time graph

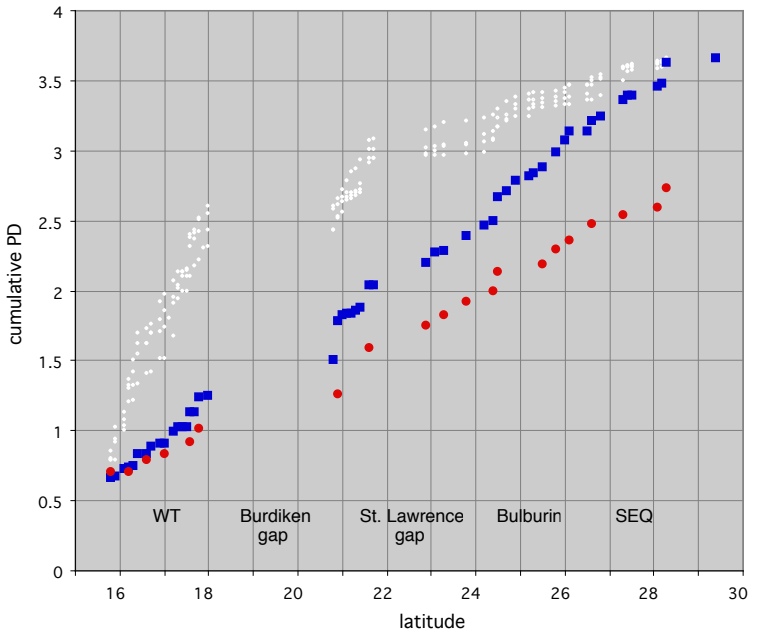


Figure 10.5 Transect of phylogenetic diversity

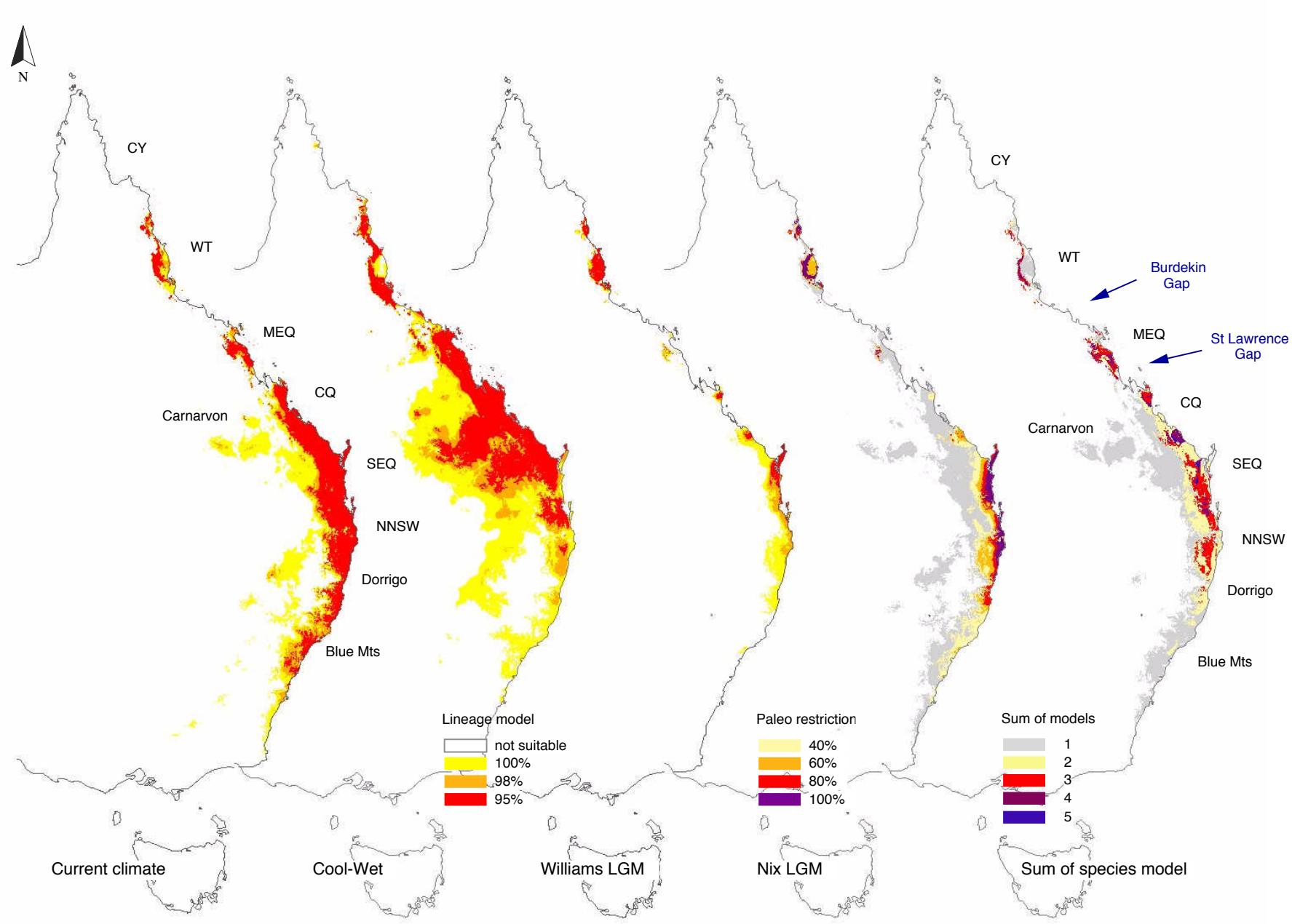


Figure 10.4 Combined lineage with paleomodels and sum of species model

Chapter 5: Spatial and temporal distribution of diversity

This chapter returns to some of the themes introduced in Chapter 1 and the biogeography touched upon in Chapter 2 to marry the near species-complete molecular phylogeny to the detailed distribution database. This first requires some description of methodology, then an overview of the biogeography, followed by more speculative analyses of spatial and temporal patterns of biodiversity. Many analyses are presented and much discussed, and the reader is directed to previous chapters for additional background information. Many – most – of the analyses are variations on themes, on ways of presenting data, ways of interpreting data; they do not necessarily amount to independent evidence but explore the data, the methods, and together hopefully the patterns and processes behind the biodiversity.

Given adequate sampling, the phylogeny can be combined with the distribution data to generate maps describing patterns of diversity, endemicity and turnover delimiting biogeographic regions (e.g. Bickford et al. 2004; Davies et al. 2008). Phylogenetic diversity (PD) is the sum of the branch lengths of a set of taxa (Figure 1) and here the branch lengths are essentially a measure of relative time. Indices relating species diversity (SD) and PD may provide measures of relative diversification, highlight regions 'laboratories' of recent endemic diversification versus 'museums' retaining ancestral diversity, and provide a basis for comparative analysis of environmental correlates of biodiversity and community assembly (e.g. Webb et al. 2002; Davies et al. 2007; Weir and Schluter 2007; Graham et al. 2009). Much recent discussion on diversity has revolved around historical processes of speciation, extinction and age of lineages, over and above environmental and ecological factors that have been the traditional focus of macroecology (e.g. Ricklefs and Schluter 1993; Cardillo 1999; Blackburn and Gaston 2004; Wiens and Donoghue 2004; Ricklefs 2006; Butlin et al. 2009). In the case here of the camaenid snails of eastern Australia, the primary data provides a description of broad patterns of diversity delineating biogeographic regions or domains. This framework then provides a basis for more complex inference on questions of processes of diversification and the determinants of biodiversity. Broad themes include:

Can we differentiate standing diversity *per se* into the component aspects of generating, accumulating and retaining diversity?

Is diversity governed by speciation or extinction, or something else for which these are merely proxies?

Can diversity be broken down into components due to historical and ecological sorting of ancestral diversity versus local endemic diversification? What is the relationship between speciation and the emergence of sympatric diversity?

Are there signs of the historical ecosystem change proposed from palaeontology and earth sciences? If so, how much of the diversification is a radiation – an increase in diversity, versus a turnover of diversity? Is it possible to infer extinction and replacement as opposed to net diversification?

Are distributions governed by environmental constraints or historical contingencies?

How do biogeographical regions compare? For example the Border Ranges, MEQ and the Wet Tropics are similarly species rich but does this reflect the same processes?

How do different habitats compare? Rainforests have the highest species diversity but does that mean that mesic regions show higher speciation rates than xeric ones. Or is speciation rate associated with transitional zones, zones that have fluctuated through historical climate changes (Smith et al. 1997; Moritz et al. 2000)?

Mapping and quantitative analysis of phylogenetic diversity is a developing field with a variety of approaches being attempted, with some duplication and uncertainty. A number of procedures presented below are novel. However the essential concepts are straight-forward and borrow from established procedures using entity counts (i.e. species diversity analyses; Rozenswieg 1995). Originally PD was conceived for use as a measure of biodiversity value for conservation purposes (e.g. Barker 2002; Faith et al. 2004), however here I use it for investigating evolutionary dynamics (e.g. Davies et al. 2007; Weir and Schluter 2007). In addition to incorporating PD, the following study uses spatial units that make up a continuous complete geographic coverage, rather than select a few widely spaced exemplars (c.f. Rozenswieg 1995; Lomolino et al. 2006).

The methods here extend from those described in Chapter 1 and where relevant refer to Chapter 1 for details. Some of the analyses are developments of approaches taken in Chapter 4. All distribution data management and diversity index calculations used dedicated Microsoft Excel v2003 spreadsheets and statistical analyses done in either Excel or JMP 3.1.5 (SAS Institute Inc.). Geographic and bioclimatic attributes and maps were done in ArcView 3.2 with ANUCLIM data (Houlder at al. 2000). Phylogenetic trees were managed in PAUP v10.4b (Swofford 2000), BBEdit Lite v6.1.2 (Bare Bones Software, Inc.) and TreeEdit v1.0a (Rambaut and Charleston. 2000).

Phylogenetic and spatial units

Tree branch lengths are estimated from the primary sequence data using a stochastic model, with divergences among the taxa due to a combination of selection, substitution rate and time (e.g. Felsenstein 2004). As the dominant signal is of time, and time is the only signal that we can meaningfully interpret for biogeographical analyses, ideally the idiosyncratic affects of selection and rate variation on observed branch length ought to be ameliorated. Therefore an ultrametric transform is desirable. While this does require additional assumptions, there is some theoretical and empirical basis to justify this (e.g. Wertheim et al. 2010), it limits confounding PD with idiosyncratic effects, and allows for a range of additional types of analysis.

Throughout, PD indices are based on the 327 tip phylogenetic analysis of Chapter 2. Trees were first pruned back to a set of 279 taxa that covered all the east coast taxa (and extra phylogeographic lineages) plus 20 extra-limital taxa to define nodes interspersed among the east coast lineages. Trees were then converted to ultrametric using Penalized Likelihood Rate Smoothing (PLRS; Sanderson 2002) with the program r8s v1.7 (Sanderson 2003) using the ADD penalty function, optimal smoothing factor=50, and the root calibrated to return a mean rate equal to the original unconstrained tree, so that branch lengths reflect the original divergence units of the input trees. The effect on indexing PD of uncertainty in topology, branch length, and clock assumptions was assessed (see Appendix 5A). While individual sub-clades can show some notable differences, by the time the data has been converted to measures of grid cell PD such differences have an unimportant effect on broad patterns of PD. It is possible to conduct PD analyses incorporating this uncertainty (e.g. by using multiple trees) but as this appears to be a minor concern, the subsequent analyses all use the pruned PL rate-smoothed version of Chapter 2 Figure 7 (see Appendix 5N).

As mentioned in Chapter 2 there is no definitive timescale for the phylogeny, however, as elaborated in Chapter 4, the *Sphaerospira* lineage shows an excellent fit to the vicariant biogeography of the mesothermal archipelago and so represents a temporal benchmark for that history. There is limited published dating studies for co-distributed non-vagile mesic forest endemic taxa across the Burdekin Gap but the few (e.g. *Hypsilurus* ca. 17 mya as a maximum; Hugall et al. 2008; *Saproscincus;* Moussalli et al. 2005) are consistent with the paleo-environmental scenarios of a Miocene timeframe of 10-20 Mya (Bowler 1982; Truswell 1993; McGowran 2004; Greenwood and Christophel 2005). The upper 0.30 divergence level of the snail tree, which corresponds to this timeframe, accounts for ~80% of the total tree length and therefore the vast majority of PD and variation in the lineage-through-time patterns. Some of the deep PD may pre-date diversification within Australia but a small confounding of ancestral with endemic diversity should only have a minor effect on recent endemic patterns.

A key aspect in analysis of diversity is choice of spatial scale: one approach is to analyse at several scales (Crisp et al. 2001; Bickford et al. 2004) but here the single grid scale developed in Chapter 1 is used. This allows a good combination of phylogenetic and distribution data coverage, and range of species diversity and endemicity, for (almost) the entire biota across the entire range. There are two versions of these grid cells: a 50km scale equal area grid and a ½ degree grid. The former is the more correct for quantitative analysis, the latter easier to map. Empirically there is little difference across eastern Australia and so they are used interchangeably (see also Chapter 1). There are 266 ½ degree grid cells in the included set, with SD ranging from 1-18 (quartile range 3-8). Fifteen grid cells have SD=1, 23 have SD=2. Some indices have trivial values when SD=1, and while others can be calculated for cells with low taxonomic diversities of 1-3, at these low values methodological distortions can be magnified, and in practice the methods, and the informative power, are best realized at higher levels of diversity (Legendre and Legendre 1998; Hardy and Sentierre 2007). Hence, for some indices grids with less than 2-3 taxa were excluded. Further some grids of low land area (overlapping coastline) were excluded for the biodiversity analyses (see Appendix 5E).

Recalling Chapters 1 and 2, of the named species and informal codes across the east coast region (east of 141ºE) there is phylogenetic information for 244/321 (76%; 92% of formally named species), with most of the remainder assignable to clade level. Of all possible hadroid taxa 166/215 (77%; 84% disregarding South Australian species) are sampled, and for east coast taxa 152/181 (84%; 95% of named species). For the purposes of the diversity indices and maps used in this chapter, grid cell diversity is based on the set tips in the phylogeny assignable to the species listed in that grid cell according to the distribution database. Some notable intraspecific phylogeographic lineages are counted as individual units (e.g. see Chapters 3 & 4). As described in Chapter 1, some grid cells are excluded due to lack of suitable information. The set of 266 included grid cells encompass 304 taxa of which 167 are hadroids; there is phylogenetic information on 79% and 88% of these respectively. To represent this taxonomic diversity the phylogeny has 255 east coast tips, 152 of which are hadroids. Thus included grid cell SD and PD should be adequately represented.

Indices of Phylogenetic Diversity (PD)

There are numerous ways to define PD, each with different properties, and different limitations, which have been discussed in a number of papers (Faith 1994; 2002; Faith et al. 2004; Barker 2002; Sechrest et al. 2002; Lewis and Lewis 2005; Hartmann and Steel 2007; Rosauer et al. 2009). Below is a brief description of various indices, referring to Figure 1 schematic.

Root PD is the sum of all branches in the sub-tree defined by the set of taxa down to the designated root.

Crown PD is the sum of branches down to the most recent common ancestor of the set.

Stem PD includes the stem branch.

Unique PD counts only the tip branches of each species (equivalent to the exclusive PD of Sechrest et al. 2002; Lewis and Lewis 2005).

A special consideration arises when there is only one species in the set: here crown PD is undefined, while stem and unique PD become identical.

Further elaborations are endemic PD, relative PD or phylogenetic clustering (e.g. Webb et al. 2002) and turnover in PD (phylogenetic beta turnover: Graham and Fine 2008); these are now described in more detail.

Endemic PD

The functions used to characterize endemic species diversity (SD) can be applied to PD (D. Faith pers. com.). The amount of endemic PD for each branch in the phylogeny is a function of the number of grid cells that contain that branch: the union of the ranges of the taxa subtended by that branch (Figure 1C). The total endemic PD is the sum of the endemic PD of each of the branches in the sub-tree spanning the set of taxa (see also Rosauer et al. 2009). Both the inverse and the sigmoidal functions can be used (see Chapter 1). The deeper in the tree the branch the wider it is distributed and hence the less endemic it becomes. For many lineages the vast majority of endemic PD is contained in the tip branches; overall ~65% of the PD is in the tip branches, ~87% for endemic PD. However, it should be added that there is more than 100-fold range in tip branch lengths and hence a large variation in unique PD per taxon.

Phylogenetic turnover

Turnover in diversity has been a subject of much investigation (e.g. Koleff et al. 2003; Qian et al 2009; and confusion: Tuomisto 2010a,b) and has recently been applied to phylogenetic diversity using a variety of metrics (e.g. Hardy and Sentiere 2007; Graham and Fine 2008). The widely used Sørensen pairwise dissimilarity function, as used for SD in Chapter 1, can be adapted to PD (S. Ferrier pers. comm.): the sum of the unique elements divided by the total: $\beta t = (b+c)/(2a+b+c)$, where a = units shared by both grid cells, b = units unique to one cell, c = units unique to the other. Thus the parameters a, b and c are the sum of the lengths of the relevant branches in the sub-tree of the combined set of taxa (see Figure 1B). This relative index is sensitive to both differences in species composition and to differences in species diversity (e.g. Baselga 2010). Absolute change in PD can be calculated as the sum of the different branches between two sub-trees: all the branches found in one and not the other, $\Delta PD=b+c$, analogous to that for SD. Like endemicity, turnover is dependent on spatial scale and analyses could incorporate multiple scales but here only the one scale is used. For mapping, ΔPD and $PD\beta t$ were calculated between each grid and the (up to) eight grids that are immediately adjacent. Thus each grid gives rise to a set of (up to) nine point values: one for each of the pairwise comparisons with the centre point assigned the average. Hence turnover appears mapped out at a finer scale (Figure 7).

The properties of species turnover investigated in Chapter 1 can be investigated for PD, shown in Figure 2. This uses the same set of grids and adjacent grids, environmental differences (ΔE) and methods – see Chapter 1 and Appendix 5F for details. PD and PD turnover is contingent upon SD: diversity is first assigned an SD value, which for the following analyses is represented by number of tips in the tree assigned to that grid cell. The additional intra-specific phylogeographical units in the tree are regarded as taxa. The same taxon is always the same tip and therefore grids with identical taxa will always have identical sub-trees and hence identical PD. Thus as expected PDβt tends to be smaller than SDβt and never larger (Graham and Fine 2008). Considering adjacent grids only, in 98% of comparisons it is less, typically around the half the numerical value (mean 0.280 vs 0.507).

PDβt shows a good fit to the logarithm of distance, in fact such a good fit it appears to actually be a logarithmic function (Figure 2A,B): using 50km scale binned averages the relationship has an R^2 =0.994, over a large scale out to 2,000km, which is a much better fit than that for species βt. However, high values towards the limit (=1) inevitably degrade any relationship (e.g. Legendre and Legendre 1998), whereas the lower values of PDβt may allow a more continuous fit. Empirical evidence and neutral theory indicate that species βt tends to broadly follow a logarithmic relationship with geographic distance but this has been little explored for phylogenetic diversity (Hubbell 2001; Condit et al. 2002; Ricklefs 2006; Soininen et al. 2007; Hu et al. 2009). Two aspects of phylogenetic diversity may contribute to its more exact fit: 1) the nested shape of the phylogeny (which is largely log-linear or Yule) allowing a lower range of values; 2) the relationship between geography and genetic distance. The further the distance apart, the more distantly related taxa tend to be, and hence relatively more turnover. If we eliminated the component due to the nested shape of the tree - if the tree was a star phylogeny - all species would be similarly different from one another and the phylogenetic turnover function would converge on that of species βt (Figure 2C). Alternatively, if we eliminated the relationship between genetic distance and geography by randomizing the names on the phylogeny (retaining the SD and species βt patterns and the nested shape of the total phylogeny), we get a poor fit curve that rises to a maximum of the overall random average (PDβt of ~ 0.65; Figure 2C). Therefore we can conclude that the exact fit reflects the profound relationship between phylogenetic divergence and geographic space that is at the heart of phylogeography (c.f. Chapter 4), and in the spirit of Hubbell (2001) speak of the log function scalar (0.16) as the fundamental phylogeography number of east coast camaenids.

Like SDβt, PDβt shows a strong correlation with absolute diversity levels, either measured as total species or as total PD ($a+b+c$): for the camaenids rich areas tend to have lots of endemics and hence high levels of relative turnover. Similarly PDβt shows a clear relationship with environmental difference (Figure 2D). Using adjacent grid cells only, in a multivariate framework factoring for total diversity and latitude, PDβt shows a significant correlation with ΔE ($p<0.0001$) but perhaps slightly less strong than that for SDβt (R^2 = 0.129 vs 0.150; see also Chapter 1 Figure 13). PDβt shows a weak non-significant trend with latitude, opposite to that of SDβt: SDβt increases with latitude while PDβt decreases. Factoring for ΔE and total diversity, while both turnovers show significant (but opposite) correlation with latitude, they only have very weak predictive power ($R^2<0.1$). As with SDβt, the major patterns are at the regional scale (see Figure 7).

The βt index used combines both change in composition and change in total diversity (Baselga 2010) and these may have different properties for SD and PD due to the intrinsic difference that unit values can take. This can be investigated by considering turnover for cases where there is only change in composition (both grids have the same number of species) and those where one grid is a subset of the

other (difference in number, or nestedness). Intrinsic differences can be explored by randomizing species names on the phylogeny, keeping the underlying distribution of branch lengths and species diversities. This suggests that there is little intrinsic (theoretical) difference between SD and PD versions of βt (Appendix 5B) but some observed difference (Figure 2F) discussed in a later section.

Relationships between PD, SD and null expectations

While individual indices based on PD are of interest in their own right, perhaps the real value (especially for evolutionary inferences) is in the relationship between SD and PD indices. Here there are various issues to consider: what might they measure; what are their properties; and how are they related to each other? An obvious choice is the ratio of SD/PD, analogous to average genetic distance among species (c.f. Bickford et al. 2004; Hardy and Sentierre 2007). The ratio uPD/SD could be regarded as the average age from the MRCA of the set of species. Another index worth considering is the ratio of endemic SD to endemic PD as a measure of local endemic diversification, filtering out diversity that is ancestral or extra-limital to the local region (D. Faith pers. com.). Both inverse and sigmoidal functions can be used. As endemic diversity can take on extremely low values the log of this index is used.

Some of these indices can be expected to have an intrinsic relationship to SD due to the hierarchical nature of the phylogeny. For example, the ratio of SD/PD is not expected to be constant because as more species are added they will necessarily be more likely to more closely related to some species already in the set and hence there is an intrinsic declining relationship between SD/PD and SD that reflects the shape of the tree (Figure 3). For certain purposes it is desirable to have indices that are not intrinsically related to SD. Therefore it is useful to determine the intrinsic - or null - relationship between index and SD, which can be developed via randomizations from a null pool of diversity. This approach is increasingly being used in community ecology as a measure of phylogenetic clustering and raises a number of aspects concerning just what is an appropriate null pool, and corresponding interpretation of the statistic (Tofts and Silverton 1999; Webb et al. 2002; Hardy and Sentierre 2007; Davies et al. 2007; Hardy 2008; Vamosi et al. 2009; Cavender-Bares et al. 2009). A particular sample may be more or less clustered than the null depending on the scale of the sample with respect to biogeographical sub-structuring, affecting interpretation of the statistic as a test value. However, as a means to provide a metric free of any intrinsic correlation to SD for quantitative analysis, such concerns may be less of a problem (c.f. Davies et al. 2007). Key issues are what phylogenetic pool and type of randomization are appropriate? The simplest way is to use the given tree and randomize the list of taxa. Another approach is to weight the probability of randomly selecting a taxon based on abundance or distribution: more widely distributed taxa, found in more regions would be weighted to be more likely to be selected in the random sets. This will give a null distribution more centred with respect to the observed range than equal weighting (c.f. Hardy 2008).

For the analyses here, null expectations are based on 300 random samplings for each SD level (2-20) from the pool of east coast species in the 327 tip tree (=255 east coast taxa). Figures 3 and 4 show some aspects of the results. For each random draw a species is selected from the pool (without replacement) with either equal probability or probability weighted in proportion to the observed range (number of ½ degree grid cells). Thus species with wider distributions are more likely to be included in a set of species and endemic measures will be lower than selecting with equal probability. Correction functions can be either ratio, or difference to null averages, or regressions. The latter assume a function whilst the former do not. The discrete integer values for SD allow the use of correction by averages, avoiding

the problem of complex (non-linear/unknown) relationships. Relative indices have the (apparent) appeal of giving values that range from more to less than expected (such as more or less clustered than expected), thus providing a value judgment but a probability test statistic has a complex sigmoidal relationship with the relative index metric (Figure 4 bottom) Another, computationally simple, approach to highlighting areas with relatively high or low PD is to use the residuals from a LOESS (or spline) regression of the observed PD and SD indices (e.g. Davies et al. 2008).

Divergence Rank Lineages DRL

Using an ultrametric tree (with or without an absolute time scale) allows for a standardized system of higher ranks based on divergence level. Like species, these can be mapped and measured for distribution and diversity. For example, at the 0.1 divergence level (similar to *Gnarosophia bellendenkerensis* and *Sphaerospira fraseri* phylogeographies; see also Chapters 3 & 4) there are some 143 east coast lineages with average range of 330 km and average 50km grid diversity of 5.3; at 0.21 divergence there are 74 lineages with average range of ~500km and 50km grid diversity of 4.6 (Figure 5D,E & Appendix 5C). Patterns of the change in the number of lineages between divergence levels may provide an alternative view of diversification, detecting the accumulation of diversity within a local (grid) scale due to diversification of sibling lineages. Comparisons can be between species (alpha taxonomy; tree tips) and divergence level units, or between two divergence levels. The latter can be thought of as a way of looking at deeper divergence patterns or as an alternative parallel analysis somewhat removed from relying upon the alpha taxonomy (Isaac et al. 2004; Hey 2004; Monaghan et al. 2009). The phylogenetic approach measures diversity that has persisted over evolutionary time spans. While there are processes occurring at shorter time scales that are overlooked, either the pattern that such processes engender will be mirrored at slightly longer timescales (detectable with phylogenetic methods) or they are ephemeral, do not lead to longer term diversity, and in any case require different miro-evolutionary methods to investigate.

Therefore another diversification index is explored, one based on the difference in the number of lineages between two divergence levels (ranks; DRL) for each grid cell (Figure 1D). This can be either a difference or a ratio. Figure 6B shows an example with the ratio of species (= tree tips) to 0.1 divergence level lineages. There is a wide range of species subtended by these lineages (from one to 17) and the index measures the spatial distribution of this difference: where sibling lineages are co-distributed (at the scale of the grid cell) DRL ratio is greater than one. Lineage accumulation in the east coast camaenid phylogeny is roughly log-linear with a doubling time of ~0.1 divergence. This trend provides a point of reference for evaluating the DRL index: values above this have within region sympatric diversification greater than the general average (which includes the allopatric component). At low to moderate divergence levels sibling lineages have small ranges relative to the total east coast and are largely regionally endemic (three-quarters of 0.1 lineages have distributions of less than 15 grid cells and maximum linear ranges of less than 450 km). Therefore the index maps something like accumulation of local (at the given grid scale) diversity due to regional endemic diversification. Where sibling lineages are distributed among grid cells, no increase is measured, effectively filtering out the allopatric component. For example, most of the phylogeographic units in *G. bellendenkerensis* fall between grids and therefore do not contribute to this signal of diversification.

Various indices such as SD/PD and DRL show strong correlations with SD either linear or non-linear, others such as endemic diversity ratio do not (Figure 3A). For this snail data, the choice of equal or range weighting randomization makes little difference for relative PD but does have a noticeable effect

on endemic PD (Figure 3B). LOESS regressions are shown in Figure 4A; in the case of PD most cells are very similar to the relative PD calculated using randomizations.

Maps of diversity, turnover and diversification indices

Figures 5-9 show various examples of indices derived from using PD. The visual appearance and hence emphasis of patterns can depend on the colour scale used. Colour coding for the maps are (mostly) an orange-green-blue 'precipitation' colour spectrum for absolute values, and blue/cyan-red colour contrasts for more relative and differential indices. The scales used are (mostly) either equal interval or by standard deviation as optimized by ArcView 3.2. Relative indices have the attraction of a scale where diversities can be above or below a mean but it should be noted that this can depend on the null pool. Where colour contrasts are used this is either based on the null mean or on the observed mean across all grid cells.

No one map embodies all characteristics of the indices and no one index is completely independent of another; they are various ways of looking at the diversity, each emphasizing somewhat different aspects and together they provide information that can identify the different aspects that go to making up the total diversity: diversification, accumulation and retention. No one analysis is unambiguous or definitive; each may have several ways of being interpreted; some may be redundant, others unhelpful. To some extent the known biogeography of eastern Australia provides a framework to assess the merits of these indices. A number of these indices have not been used before and their characteristics remain to be explored, so at the very least the results here go some way towards this.

Figure 5A-C shows SD, PD and relative PD. This is similar to the LOESS regression in Figure 6D. As PD is measured from the root of an ultrametric tree, all cases of a single species will have the same PD, equal to the height of the tree, and hence the same average relative PD (in this case = 0). Therefore grids with SD<2 are either excluded or ignored in such analyses. In any case, investigations of diversity generally avoid using regions of extremely low diversity, for a variety of reasons. While SD and PD are broadly similar, relative PD emphasizes the contrast, highlighting areas with relatively more or less PD per species. For example the Wet Tropics and mid-east Queensland have similarly high SD (and endemic SD: Figure 5F) but MEQ has less PD per species than WT. Compared to the overall null average, WT has high levels of PD per species – on average species are divergent from one another – while MEQ (particularly the northern part) has low levels of PD per species – the species are relatively closely related. Figure 6A shows endemic PD. As with PD, endemic PD can be measured relative to SD (Figure 6E). This is calculated as endemic PD relative to the average endemic PD of random sets of the same number of species, with probability weighted by range, thereby preserving the distribution of endemicity (c.f. Hardy 2008; Vamosi et al. 2009). A LOESS residual can also be used, and like PD is reasonably similar to the null adjusted index (not shown). Figure 7 shows maps of turnover, both absolute ($b+c$) and relative (βt), and the difference between species and phylogenetic βt. The Euclidean distance among plant growth indices (ΔE) as described in Chapter 1 is also shown.

These maps are a view of biogeographical domains, partially highlighting 'museums' and 'laboratories'. Two other indices intended to be more specifically focussed on diversification rate are the endemic SD/PD ratio and the Divergence Rank Lineage ratio. The endemic SD/PD ratio (Figure 6C) is the natural log of endemic SD divided by endemic PD (such as Figures 5F & 6A). Here regions of SD =1 can take on different values. The endemic SD/PD ratio is something of a mirror of relative endemic PD. The DRL ratio (Figure 6B) is the species diversity (Figure 5A) divided by the 0.1 lineage

diversity (Figure 5D). The index has a minimum value of one and is only relevant to areas with SD>1. Across the whole phylogeny there are about twice as many species (= tips) as lineages of 0.1 divergence depth but ~60% of the 50km grid cells show no increase at all, with less than 5% of grid cells near the trend level – the rest of the diversification is distributed among grid cells. Of the 143 0.1 lineages, only 24 contribute species that co-occur at the 50km grid scale. As DRL ratio has a strong intrinsic correlation to SD a corrected or relative version can be derived, shown in Figure 6F. The alpha taxonomy of the east coast camaenids is not exactly know (c.f. birds) and includes many informal codes: some of these will certainly be described (Stanisic et al. in press) some probably, and some perhaps never. Here DRL can be used between two divergence levels to investigate if patterns are heavily dependent upon alpha taxonomy. In particular two areas, Sydney Basin and northern parts of MEQ, appear to have high levels of recent diversification. However, DRL ratio analysis between two levels 0.04 and 0.14 produces similar patterns to using the alpha taxonomy (both SB and MEQ record high signal), indicating the pattern is robust to pooling the tips into half the number of lineages (Figure 8).

The various indices relating SD and PD bear some resemblances (summarized in Appendix 5D), with relative PD reasonably correlated to DRL and endemic diversity ratio (Pearson product-moment coefficient: -0.63, -0.52) and of course to SD/PD (-0.70), while endemic diversity ratio and DRL and relative endemic PD are somewhat less correlated (0.39, -0.26). A final index that may capture some interesting aspect of diversification is the difference between species turnover and phylogenetic turnover (Figure 7F), the distribution of which partial resembles that of relative PD and the diversification indices.

Endemic diversity ratio has theoretical appeal as a measure of diversification but in practice may be very sensitive to how endemicity is defined and to scale, and as endemic values can approach zero, can behave erratically. Relative PD (or phylogenetic clustering) also has an intuitive appeal but is dependent on a computationally intensive null, and what it signifies is a matter of interpretation. Primarily it is a biogeographic measure of relative average species age, demarking of areas with relatively old or young species (e.g. Davies et al. 2007). In community ecology discussion of relative PD emphasizes competitive exclusion and habitat filtering (e.g. Webb et al. 2002; Vamosi et al. 2009), where inferring the effect of niche evolution on community assembly may best focus on the more recent regionally endemic component of diversity. In the case here relative PD shows a reasonable correlation to measures of local recent diversification rate, consistent with the intention. The DRL is straightforward but depends on an *ad hoc* choice of divergence depths but a doubling is a simple rule. Turnover in PD has obvious appeal for biogeography alone but again it is the interplay between species and phylogenetic turnover that may have evolutionary interest. As they are both (nominally) adjusted to the same relative scale βt may be the best choice for this. The difference between species and phylogenetic βt emphasizes areas where species composition changes relatively more than the underlying phylogenetic diversity, for example where there is recent diversification across a gradient or geographical feature (Graham and Fine 2008). I return to further interpretation of some of these indices later.

Delineating biogeographical domains

Having dealt with some of the methodological matters we are now in a position to recall the general overview of regional diversity presented in Chapter 1 and place the phylogenetic results of Chapter 2 in this context to delineate a biogeographical framework. Figure 10 maps the known distribution of the

ten large clades discussed in Chapter 2 comprising virtually all of the east coast camaenid diversity. Clades 5, 6 and 7 have been grouped for simplicity – they may form a clade but are better thought of as an assemblage. These distributions underlay the distribution of phylogenetic and species and lineage diversities shown in Figures 5-9.

Some patterns are worth noting. Two assemblages (clade 10 and clades 5-7) which have (or probably have) northern extra-limital (PNG, NT, Kimberleys) members show a similar northern distribution; *Xanthomelon* (not shown) also has a similar pattern. The Wet Tropics (and fringes) has members of all but one (clade 3), and a trio (clades 8, 9 and 4) show a disjunct distribution between a WT component and more southern regions following the 'mesothermal archipelago'. Clades 8 and 9 show the 'leapfrog' distribution suggestive of extinction across MEQ region (see Chapter 4 and Norman et al 2002). Along with the endemic clade 5.5 '*Austrochloritis*' as well as a few other rarities of possible Melanesian origins, the WT contains elements of all of the deep phylogenetic history of eastern Australian Camaendiae (e.g. the Bassian connection with southern Australia in clades 8, 9). Large ranges but little overlap among many clades; for example sister clades 1, 2 and 3 have a series of adjacent distributions with little overlap spanning the entire east coast region. The combination of distributions highlights the turnover across the Border Range region from SEQ to NENSW. Clade 2 extends into SEQ and is replaced to the south and west by clade 3, clade 5 (chloritids) reach as far south as the D'Aguilar Ranges and adjacent Brisbane River Valley, while the clade 9 chloritids go as far north as the Border (and Main) Ranges – distribution north of this is due to the papuinid *Posorites*. There is a parallel pattern in clade 8 with one sub-lineage of *Thersites* extending to the D'Aguilar and Conondale Ranges, and a reverse pattern in clade 4 with *Sphaerospira fraseri* phylogeography extending just into NENSW to about the Clarence River (latitude 29.8 ºS).

Considering the additional information from the endemicity, turnover and relative diversification maps, some of the major features can be described in more detail differentiating aspects of retention, accumulation and diversification.

The small but discrete Cape York rainforests have moderate SD and endemism but more substantial PD and phylogenetic turnover. There is little endemic community diversification but there are barriers to dispersal – especially to the south (the Laura Gap). Some of the pattern in PD may be due to Melanesian elements resulting in an over-estimation of endemic PD, however if we consider the hadroids alone (Figure 9) there is still a strong endemic signal. Three Cape York endemics in three different lineages (*Hadra barneyi*, *Sphaerospira* lineage CY1 and '*Hadra*' CY10) are all relatively divergent from their respective southern sisters.

The Wet Tropics has it all: high diversity and endemism of all types, and turnover both between surrounding areas and within the WT (c.f. Chapter 3). Across the WT there is roughly a north-east to south-west variation in 'diversification' indices from low to moderate: eastern parts show a pattern of retention of diversity, old species (see relative endemic PD), while towards the fringe there is more sign of younger endemic diversification. The WT shows evidence of retention in that it contains members of all the deep ancestral lineages. It shows evidence of retention – of maintenance of diversity or low extinction - in the density of relative old regionally endemic diversity (old sibling species). It shows evidence of accumulation of diversity in the presence of several Melanesian elements, and in signs of recent diversification at the adjacent western and southern fringes – across the steep environmental gradients. The WT is a remarkable 'island': a refuge and an engine of diversity.

Mid-east Queensland shows a distinct boundary to the north, less so to the south. There is moderate phylogenetic diversity, endemic PD and turnover but low relative PD and subregions of high endemism suggesting recent radiation in fragmented areas adjacent to the main centre of phylogenetic diversity in the more upland mesic forests of the Clarke and Conway Ranges. Here, there is some evidence of mesothermal refuge in the relative endemic PD index (see Chapter 4 and Moussalli et al. 2005). The northern part, roughly Conway Range to Bowen, contains a patchwork of fragmented lowland megathermal mesic forests and vine scrubs and so contains many populations that are differentiated to various degrees but irrespective of taxonomic subjectivity, underlying this is a strong signal of recent regional endemic community assembly, as exemplified by the '*Bentosites'* trio of *etheridgei*, *coxi* and *yulei* (see Chapter 2). To the north, immediately the other side of the Burdekin Gap, a similar ecogeographical and biodiversity pattern is seen: much fine scale endemism but also young community diversification. This is highlighted in the species-0.1 lineage versus the 0.4-0.1 lineage DRL results: while the signal is much less in the latter, it is still present (Figure 8). This is one hot laboratory, the question is why?

In the SEQ-NNSW region from the Conondale Ranges to the Border Ranges and south to Barrington Tops (just north of the Hunter Valley) there is a broad band of high species and phylogenetic diversity with low endemism. However, there is a massive turnover of deep diversity across the Fassifern and Brisbane River Valleys underlying the pattern of high diversity and low endemism. There is a corresponding gradient in diversity and lineage turnover immediately to the west. This arc surrounding the Fassifern and Brisbane River Valleys from the Border Ranges via the Main Range and around to the D'Aguilar Range is as rich in historical biogeography as any in eastern Australia. Just considering SD and PD and endemic SD and endemic PD, this division would not be seen (the IBRA 6.1 groups SEQ and NENSW) – it is the turnover and especially the phylogenetic turnover that highlights this outstanding biogeographical feature; not a gradient marking attenuation of diversity but the narrow overlap of deeply divergent biotas (the northern Torresian and the southern Bassian: Spencer 1896; Burbidge 1960) across all groups of Camaenidae.

The Sydney Basin (south of the Hunter Valley) shows low phylogenetic diversity but more endemic PD, and high levels of species endemism and relative phylogenetic turnover, suggesting recent endemic diversification. Again, sidestepping the potential for taxonomic inflation (Isaac et al. 2004), the lineage only DRL continues to indicate regional endemic diversification (Figure 8). There is an offset of species and phylogenetic turnover with little absolute PD but substantial relative decline and turnover across the Hunter Valley, north of the rise in species turnover across the Sydney Basin. This is highlighted by the SD-PD βt turnover difference. The results indicate that the Sydney Basin has been or still is undergoing regional diversification and it is worth noting that the recently discovered *Saurochoncha caperteeana* (Zhang and Shea 2008) appears to be a derived endemic associated with *Pommerhelix* within clade 3(M. Shea pers. com.).

Then there are the gaps in between the centres of diversity:

The Laura Gap appears to act mainly as a pathway or filter of northern diversity but a barrier to southern elements (c.f. Kikkawa et al. 1981).

The Burdekin Gap, the foremost biogeographic barrier in Australia (Burbidge 1960 and others) shows deep divergence across mesic mesothermal lineages, and represents the southern limit of several

northern megathermal (and possibly Melanesian) elements. On either side and within the gap there is a profusion of recent diversity of various distinction.

The St Lawrence Gap shows only a partial separation between MEQ and south, with divergence in only some lineages (e.g. clades 1, 4).

The Hunter Valley – a noted biogeographical boundary (Spencer 1896 and others) - is the southern limit of many lineages, with only clades 3 and 9 further south, thus acts mainly as a filter, especially of phylogenetic diversity: PD declines before SD towards the southern range limit.

Third, gradients of diversity:

Compared to the big centres of diversity, south-east to central Qld appears more of a continuum with 'average' PD and low endemic PD, except of a pocket of endemicity and diversification (e.g. relative PD, DRL) suggesting a minor centre of endemic diversification in the Kroombit-Byfield region (see Chapter 4).

The tail of diversity south of the Sydney Basin shows a decline in numbers of species and lineages but less so for phylogenetic diversity. A large part of the region has SD>2 rending PD, relative PD and DRL uninformative. Endemic SD and PD remain relevant, and unique or stem PD might be useful but just looking at the phylogeny (Chapter 2) is probably the best. The Hunter Valley represents a substantial filtering of diversity but the extreme southern diversity is comprised of two endemic deep lineages both near as old as the entire hadroids, a distinct '*Austrochloritis'* lineage within clade 9 and *Chloritobadistes*, and is therefore not a subset of northern diversity but an endemic fauna of low species diversity but substantial endemic (at the relevant scale) phylogentic diversity – relicts. Some of this may be seen in the LOESS residual PD and relative endemic PD. Despite the low diversity there is signal of relatively young species (where the diversification indices are still informative). One interpretation of a combination of endemic lineages, low diversity but young species is that this southern limit of diversity is not governed by dispersal limitations (time since dispersal: Wiens et al. 2006; 2009) or low speciation rate but by long term environmental factors via extinction. In particular, the minimum temperature limit is mirrored in the decline in diversity across the higher parts of the New England Tableland further north (see also Figure 13 below).

Adjacent to the WT, west and south-west to the Einasleigh there is moderate decline in species diversity and PD, more decline in diversity of deep lineages but a massive turnover. The steep environmental gradient (see ΔE) from the fringe of the WT turning into the Einasleigh is highlighted by the βt difference. Given that the PD is contingent upon the SD in the method, the relative difference between SD and PD βt depends upon how closely related the species are. Therefore, proceeding west, first there is a turnover of major lineages from the deep forest groups such as (clades 4, 5.5, 8, 9) to a different set of lineages (clades 2, 5, 7, 10); a modest decline in species diversity and phylogenetic diversity but a large turnover. Next there is a turnover in species but only among taxa within clades hence relatively less PD turnover. The major turnover among lineages is associated with the ΔE, while the βt difference, the turnover within lineages, is adjacent to this.

The north-west slopes of the Great Dividing Range in NENSW show low diversity, low endemicty, low turnover but a diversification signal (in relative PD, endemic diversity ratio and DRL). One interpretation of this is that it has acquired diversity (recently) dispersed from the more diverse areas to the east. Towards the west of the SEQ-Border Ranges-NENSW region there is substantial turnover but

only a slight signal in turnover difference following the environmental gradient suggestive of parapatric diversification, most of the turnover is due to decline in total diversity. The western NSW areas are poorly sampled and the taxonomy still provisional (see Chapter 2; Clark 2005; 2009) therefore for the time being the patterns can not be much further explored.

Inland Queensland is somewhat obscured by patchy sampling of grids but has moderate species and phylogenetic diversities. There is a mix of widespread species (*mattea*, *mucosa*, *pachystylum* in clades 1, 10, 11), plus a series of clade 3 taxa (in '*Adclarkia'* and '*Pallidelix'*) contributing to turnover and diversification signals. Thus the region appears something of a continuum overlaying relicts associated with the high ranges; for example, BL7, clade 5 '*Austrochlorits'* in the Carnarvon Ranges, a minor 'mesothermal island' (Nix 1982). Again, more detailed sampling is desirable.

Distribution of βt difference

Low SD and patchy grid sampling affects interpretation of some of the western and far southern patterns but elsewhere the numbers are sufficient for the pattern to be robust. The difference in βt highlights two areas where SD turnover is relatively greater than PD turnover: 1) a band from the western fringe of the WT across the Burdekin Gap and into northern MEQ; 2) the Hunter Valley and the Sydney Basin. Across all indices both these areas show signal of diversification: low relative PD, high endemic diversity and DRL ratios. Biogeographically one is tropical the other temperate, with completely different fauna only sharing clade 9, and then utterly different sub-lineages (one a papuinid the other a chloritid). They are both at the edges of major biogeographic boundaries but that other great biogeographic demarkation, the Border Ranges area, does not show much signal in the diversification indices or in the βt difference. However the Border Ranges is a region of turnover due to an overlap in two rich but different biotas; the feature common to the two areas highlighted in the βt difference maps is that they straddle gradients at the edges of centres of phylogenetic diversity.

Diversification driven by an environmental gradient results in relatively closely related species either side (e.g. Moritz et al. 2000), hence relatively more difference in SD than PD but only for change in composition of diversity, not merely a change in the amount of diversity (Graham and Fine 2008). The standardized turnover metric allows us to make this comparison (Figure 2F & Appendix 5F). Here we can see that the observed difference between SDβt and PDβt has a significant correlation to environmental change (ΔE), in particular for change in composition of diversity but not for change in amount of diversity. Some of this can be seen in Figure 7 where the index traces along the Great Dividing Range in NNSW-SEQ, along the western edge of the Wet Tropics, and across two predominant biogeographical barriers, the Burdekin Gap and the Hunter Valley. Hence this result can be taken as a global signal of the contribution of adaptive diversification driven by environmental gradients.

Multivariate models of diversity

So much for maps and narrative description; in this section the data is subjected to a multivariate analysis of species diversity assessing the relative contribution of environment, environmental heterogeneity, and diversification. Broad types of questions are: how much of species diversity is governed by climate, by environmental spatial heterogeneity, by diversification, or by historical change (temporal heterogeneity)? While some of the questions may be simple to pose, over and above the

narrative description afforded by the maps, quantitative statistical analysis is a complex matter and here I merely broach the topic.

These analyses use the more-or-less contiguous array of grid cells shown in the maps but excluding grids with SD<2 and 20 grids with less than half land area, giving a total of 231 grid cells (see Appendix 5E). The grids represent individual data points in the multivariate analyses. This brings in the aspect of non-independence between data elements (grid cells). As mentioned in Chapter 1 to some extent using all grids in a contiguous complete coverage ameliorates some potential biases but in addition here I include another approach to help correct for shared diversity among grid cells - the occurrence of the same species (and phylogenetic lineages) in multiple grid cells. While the data points in biodiversity analyses are the grid cells, the proposition here is that the fundamental units contributing to biodiversity are the species, or in the case where we have a phylogeny, the lineages: these need to be accounted in analyses equally. Consider two adjacent cells with the same species: these are replicates of the same evolutionary history (process) not independent but two versions of same biodiversity in somewhat different environments. Compare this with two adjacent cells with near identical environments but different species: different evolutionary history in similar environments and therefore better thought of as more independent replicates of evolutionary process. Considering phylogeny further discriminates data interdependence: they could be close sister species or entirely unrelated. Rather than use geographic distance as a proxy we can directly use the evolutionary units as a measure of grid cell relatedness. Therefore the scheme proposed here is to weigh grid cells by proportion of endemic phylogenetic diversity (using the inverse function: p-invPD). This allows for a more even contribution among the individual units of diversity, thereby ameliorating auto-correlation among grid cells. For example, two cells may have the same species and therefore represent replicates of the same evolutionary diversity – weighing the cells by the inverse of the species range (in this case ½) allows that species to carry a weight in the analysis of one. Similarly, the two cells may have different species but these could be either very distantly related or actually very close sister species – if we wish to investigate patterns of PD then weighing branches by the inverse of the numbers of cell in which they occur results in more equal contributions.

In this analysis SD = tips in the tree, which then determines the PD. The basic framework is standard least squares with selection among multiple effects using linear and quadratic terms based on AIC. All analyses include area although there is negligible correlation between SD and area after excluding the 20 grids with the least area. All analyses use the p-invPD weighting but the results are much the same in sign, strength and significance using equal weighting (Appendix 5J). Each grid cell has environmental mean and standard deviation values based on the 1km scale ANUCLIM surfaces. The standard deviation is used as a measure of heterogeneity within a grid cell. The results of these analyses are shown in Table 1 and Figure 11.

Many climate variables show a strong correlation to SD, in particular the plant growth indices developed by Henry Nix (see Chapter 1). A small subset can be selected to represent climate and heterogeneity: mesothermal PGI, temperature of the wettest quarter, and standard deviation in megathermal PGI (denoted meso, TWeQ, mega_sd). Mesothermal PGI embodies much of the world of these snails, TWeQ accounts for some measure of rainfall seasonality, mega_sd largely reflects environmental variation due to topography and east-west gradients. In a simple linear term only model of SD these three can account for more than half the pattern of SD ($R^2 = 0.54$; Table 1A). Other combinations give similar results (not shown). Screening indicates that quadratic but not interaction terms are of some value but do not change the overall picture ($R^2 = 0.568$; Table 1A; Figure 11A): SD

increases as it becomes more mesothermal – at least to some limit, and increases with environmental variation – beta diversity, and in summer rainfall regions. That species diversity is heavily influenced by climate is to be expected, the real point is how is it related to underlying patterns in phylogenetic diversity?

As some of these measures are intrinsically correlated with SD it is desirable to use the null corrected versions. Three indices that capture some of the variation in phylogenetic diversity and in diversification rate are relative PD, endemic diversity ratio and corrected DRL ratio (Figures 5C, 6C & 6F). Relative PD and DRL are expressed as a proportion of the null expectation, in order to reduce correlation with SD via variance. Thus the three indices have little or no intrinsic correlation to SD (Figure 3). Relative PD has been used as a measure of phylogenetic clustering, and interpreted in terms of ecological constraints on niche evolution and community assembly (e.g. Webb et al. 2002; Graham et al. 2009). There is no measure of speciation rate *per se* but endemic diversity and DRL ratios capture some element of recent local diversification rate. Relative PD is perhaps better thought of as relative age of species and hence a mix of retention of phylogenetic diversity and diversification. This is more along the lines of Davies et al. (2007) where it was used as an index of relative speciation and extinction rates. As the three indices are reasonably correlated they involve some level of redundancy.

By themselves these measures show weak (but nominally significant) negative correlations to SD: both measures of diversification rate (endemic & DRL ratios) decline with increasing SD, while relative PD increases (Figure 11B). Considered together and with quadratic terms (interaction terms were not important) they provide only a weak fit to SD (R^2=0.11) with negative trends (Table 1B; Figure 11C). It is obvious that at the large scale climate has an overarching control on SD and to understand the more local secondary effect of phylogenetic factors it is desirable to analyse this in the context of climate. Therefore the combined model uses both climate and PD indices simultaneously (Table 1C). In the combined model (either linear or linear plus quadratic) including the phylogenetic diversity terms makes little contribution to fit and even allowing for the large scale effect of climate the diversification trends are still negative. Relative PD effectively shows no relationship with SD, DRL ratio shows a weak negative trend and endemic diversity ratio shows a negative or modal pattern. Removing environmental heterogeneity (mega_sd) does not change the sign but relative PD shows a more modal function (not shown). Either alone or together, with and without controlling for climate, the relative PD and diversification indices are not correlated with SD; if anything, high SD is associated with low diversification indices and high relative PD. This can be seen qualitatively in the maps.

What are the relative PD and diversification indices correlated with? Preliminary investigations show that a suite of climate parameters can yield moderate fits (R^2 = 0.29-0.50). Relative PD shows strong correlations to temperature, rainfall seasonality, and to latitude (R^2 = 0.35; Figure 12). Using stepwise selection of linear and quadratic terms for eight variables encompassing temperature, rainfall, heterogeneity and latitude, there are complex interactions but latitude shows a strong signal (Table 2; Figure 10). Such an analysis requires further consideration but broadly these results resemble those of Davies et al. (2007). Some of the complexity may be appreciated by considering the spatial distribution of patterns in the diversification indices (including relative PD) and the turnover difference. Diversification is not particularly associated with the major centres of SD, and not with regions of very low diversity but with regions of intermediate diversity adjacent to major centres of endemic diversity (regarded as refugia) and lie in or around major biogeographical boundaries: Burdekin Gap, Hunter Valley, west of the WT, and the western slopes of NENSW. The difference in SD and PD βt also

captures some of this: large amounts of signal associated with mid-level SD and eco-geographic gradients.

Latitude is better thought of as a complex proxy for more underlying factors: not only environment, as in energy and seasonality but also in historical fluctuation in environment; as well as contingent effects of geographical distribution of environments and sources of diversity (Gaston 2000; Cardillo et al. 2005; Jansson & Davies 2008). That latitude knocks out temperature in stepwise modelling suggests something more than just climate is involved, aspects which may not be readily represented. The core high diversity regions, the refuges, by definition represent somewhat stable domains that can accumulate and retain diversity. The adjacent boundaries and gaps are by contrast diversity troughs or sinks (c.f. Goldberg 2005). Two interpretations of this are that 1) species diversity is governed not by speciation but by extinction; 2) the signal of ecosystem change with diversification now concentrated in expanding environments.

One approach to investigating this further might be to consider historical change - temporal heterogeneity – in environment (Graham et al. 2006; Jansson & Davies 2008; Araujo et al. 2008). Where current diversification is associated with historically unstable (fluctuating) environments, this suggests that the diversity is ephemeral and extinction is the overarching governing factor; where it is associated with expanding (changing) environments then it more reflects ecosystem turnover. At the moment there is no index of historical change but in principle it might be possible to build paleoclimate proxies, akin to the LGM paleomodels used in Chapters 3 & 4, based on paleo-environmental hypotheses inferred from the fossil record (e.g. McGowran et al. 2004; Kershaw et al. 2005).

Before moving on, this section would not be complete without a map, or two. Figure 13 shows a spatial projection of a climate model similar to the one used above but because there are no PGI layers for the whole continent and mega_sd is grid scale specific, four other climate variables are used giving a similarly high fit to SD (R^2 =0.53). Otherwise it is the same general procedure using the 231 50 km grid cells as data points in a GLM with linear and quadratic terms to predict SD. This model has then been used to predict SD for $^1/_3$ degree grid cells across Australia. The companion map shows the sum of logistic bioclimatic distribution models for seven big clades (shown in Figure 10), using the same climate variables and linear and quadratic terms. Here the presence data is the observed $^1/_3$ grid cells, weighted by the species diversity, while the absence data are all the other grids across Australia, weighted to give a combined weight equal to that of the presence data. This is essentially the same type of distribution model structure that is now widely used (e.g. Elith et al. 2006; Moussalli et al. 2009) except that it is for large clades and presence is weighted by SD. The resulting maps highlight the role of climate in governing the snail distribution and diversity, in particular: the constraint to east of the Great Divide but with regions of diversity in inland Qld, attenuation or absence in Victoria, and the near absence from the rest of the continent except some areas where they are or may be – Red Centre, Kimberley-Arnhem Land, western NSW-Flinders Ranges. The attenuation in diversity across the southern NSW and New England Tablelands highlights the role of temperature in controlling the southern limit of distribution (see Chapter 1 & Appendix 1E).

Temporal framework

To the above descriptions of east coast diversity we can attach a relative temporal scale to a range of key ecogeographical divergences in the phylogeny. While this is only a tentative absolute time scale, the relative time scale (c.f. Loader et al. 2007) among discrete divergences across biogeographic

boundaries and/or environmental biomes is more robust, and allows a consilient and coherent description.

To provide a relative temporal framework the two gene 147 taxon data of Chapter 2 has been subjected to relaxed-clock analysis. This dataset provides perhaps the best mix of phylogenetic robustness while containing sufficient taxonomic coverage to provide relative dating from a range of key divergences. These comprise examples spanning known biogeography at several different geographic scales (from within region phylogeography to among region biogeography), shifts between coastal mesic forest and more xeric sclerophyll inland environments, and biogeographically and phylogenetically large scale diversifications and relict disjunctions.

For relaxed-clock dating two methods were used: BEAST (Drummond et al. 2006) analysis of the 147 taxon dataset; and PLRS (Sanderson 2002) analysis of the Chapter 2 Bayesian trees. Both analyses use the same GTR-G-inv sequence evolution model for each gene (=3 partitions). Both use the same calibration scheme: a normal distribution prior on the substitution rate of mean = 0.017 and standard deviation =0.002, combined with uniform constraint of 10-20 mya for the *Sphaerospira* lineage Burdekin Gap split. This mtDNA rate distribution allows the prior to span most estimates of mtDNA substitution rates (2-4 % pairwise per mya: Quek et al. 2004; Pulquerio 2007; Nabhotz et al. 2009; Ribera et al. 2010; and see also Chapter 3). The BEAST v1.4.8 (Drummond and Rambaut 2007) analysis used the uncorrelated lognormal relaxed-clock model with Yule node height (speciation) prior, 10 million steps sampling 1/1000, burnin 10%, with ESS for all parameters >300. The PLRS analysis in R8s v1.7.1 (Sanderson 2003) used 400 samples taken from the MrBayes analyses (Chapter 2, Figure 4). These were then rate smoothed by Penalized Likelihood with the TN algorithm, ADD penalty function and optimal smoothing factor = 50. The analyses were calibrated with the uniform Burdekin Gap constraint and a root age drawn from a normal distribution such that the set of rate-smoothed trees returned a mean rate = 0.017 with stdev =0.002. The BEAST v1.4.8 tree is essentially the same as the results of the MrBayes and RAxML analyses in Chapter 2 (see Appendix 5M).

Figure 14 shows the results for both methods for the set of divergences considered, arranged by age. Some of the deep hadroid splits may be redundant (due to topological uncertainty) but are intended to bracket the age of this region of the phylogeny. Owing to the tentative nature of the calibrations the results are shown as quartile and 95%CI plots but it is clear that broadly both methods (quite different branch length, topology and relaxed-clock assumptions) return much the same divergences, and the same relative pattern. The hadroid splits are also indicated on Figure 16.

At the lower end there are the two phylogeographies of *G. bellendenkerensis* in the north WT (see Chapter 3) and the peripatric speciation of SQ1 from the phylogeography of the *Sphaerospira fraseri* complex in SEQ (see Chapter 4). These both biogeographically and temporally represent examples of Plio-Pleistocene phylogeography. Next is a series of disjunctions and ecogeographic divergences: the disjunct WT northern outlier of the clade 1 complex dominated by MEQ groups; the clade 4 (*Sphaerospira* lineage) divergence across the St Lawrence Gap (Chapter 4); the leapfrog distribution of WT and SEQ-NENSW papunids; WT upland RF endemic *'Jacksonena' delicata* phylogeography; the extended phylogeography of the SEQ-NENSW *Thersites* species complex; and two notable ecological diversifications: the rapid radiation of clade 2 spanning Cape York and Far North Qld including WT mesothermal endemic *J. rudis* and megathermal xeric adapted *Spurlingia* spp across the Einasleigh, and the divergence of inland sclerophyll adapted clade 1 *mattea* from coastal forest assemblages across central and SE Qld.

Slightly older than this are: the clade 4 *Sphaerospira* lineage Burdekin gap benchmark; two large scale regional diversifications (clade 1 with basal split to the xeric inland adapted *zebina* and the radiation of South Australian members of clade 3); and the two highly restricted Cape York Iron range and WT Thornton Peak endemic relict taxa, belonging to the *Sphaerospira* lineage.

The outstanding relicts within the hadroids, the inland Qld BL7, the basal chloritid types (the SEQ *Offachloritis*-SQ4 lineage and Victorian *Chloritobadistes)* and the far inland desert *Aslintesta*, represent a signal of a global ancestral distribution and radiation of the hadroids. At a similar timescale to this very large-scale radiation and attenuation, is the extant diversity of the clade 5.5 '*Austrochloritis'* endemic to the WT, and a group of chloritids spread across Melanesia (clade 7: PNG to Solomon Islands; see Chapter 2). Although the dataset only contains seven representatives in (two genera) of this group, overall sampling is enough to highlight that it is not present in eastern Australia and that it is of considerable (relative) age considering ideas on the age of Melanesian radiations (Mayr and Diamond 2001; Meredith et al 2009).

Beyond this fall the deep disjunctions in clade 8, the large scale diversifications in clade 9 including the far southern endemic chloritid lineage and chloritid - papuind ecological differentiation, and a matching chloritid - papuind ecological divergence in clade 10. The deep divergences in clade 8 (and 9) likely represent more relicts (attenuation) of ancestral diversity than vicariant (endemic) divergence.

By biogeography and timescale, below the level of the *Sphaerospira* lineage Burdekin Gap reference point (ca. 13 mya) patterns represent largely *in situ* endemic regional vicariance and radiation, with some or all of the divergences falling into the mid to late Miocene period (ca. 5-10 mya) of substantial change in Australian vegetation - the 'Hill Gap' – inferred from the geological and fossil records (Bowler 1982; Truswell 1993; Martin 1994; McGowran et al. 2004; Greenwood and Christophel 2005; Byrne et al. 2008). Above this level, divergences more and more represent elements of attenuation and re-distribution of ancestral diversity – 'old wine in new bottles'.

The intention here is not to claim specific dates or that there are eras of simultaneous divergences, or to arbitrarily select nodes to give such appearances but that out of all of the extant diversity there are at least a few nodes that have clear biogeographical and environmental correlates, and if we presume that the relative node ages are reasonably accurate then the relative ages of these together form a coherent series that chart some of the major trends in the Tertiary historical biogeography of Australia: the absolute dating should be consistent with this. For example halving the timescale may be still possible, doubling it, not.

Diversification in the hadroids

The hadroids are a suitable focal group for detailed diversification analysis as they are well sampled both distributionally and genetically, appear to be a monophyletic endemic Australian group, and represent a large part of the east coast camaenid diversity (160+ species, near half the total). They largely comprise four big clades: Clades 1, 2 and 3 have a series of adjacent distributions with little overlap spanning the entire east coast region, which across the northern part, broadly overlaps the sister lineage, clade 4 (Figure 10). In particular, clade 1+2 is regionally co-distributed with clade 4 across eastern Queensland but in a different environmental milieu. Clade 4, the *Sphaerospira* lineage plus sister lineage (*Offachloritis et al.*), has a strong association with the mesic forests of the mesothermal archipelago (see Chapter 4) while clade 1+2 contains a variety of ecotypes including some

mesothermal endemics but is dominated by megathermal forest and vine scrub taxa with multiple instances of xeric sclerophyll adapted taxa. Especially clade 2 exhibits a dramatic radiation across environmental gradients in FNQ and Cape York. Both Chapters 2 and 4 provide some detail on this but in addition Figure 15 shows a simple bioclimatic distribution of the main hadroid clades for the mesothermal and megathermal plant growth indices (as described in Chapter 1). Selected individual species averages are overlain. To assist the reader a similar colour coding scheme for these hadroid clades will be used throughout. Note the wide range in clade 2 (e.g. *barneyi* vs *rudis; dunkiensis* vs *gemma*), especially given the relatively young age compared to clade 4 (e.g. *bellendenkerensis* vs SQ1) which is at least twice as old.

Given the near completeness of the sampling, co-distribution and environmental correlates, the hadroids and clades within the hadroids are a suitable group for lineage accumulation (or lineage – through-time; LTT) diversification analyses (Harvey et al. 1994; Rabosky 2006a; see Chapter 1). The South Australian radiation has only been partially sampled (11/31; see Chapter 2) but sufficient to indicate a reasonably discrete phylogenetic origin within clade 3, and should have only minor effect on the analyses that focus on east coast diversity. The remainder of the hadroids should be well enough sampled for LTT analyses (>80% of known diversity; Pybus and Harvey 2000; Ricklefs 2007b).

This uses the hadroid part of the all-gene 327 taxon dataset of Chapter 2 and five outgroups to provide information on the root position. Methods follow those of the 147 taxon dating analyses: ultrametric chronograms via BEAST or PLRS of MrBayes trees, using the same sequence evolution and relaxed-clock models, and calibration scheme. The difference here is the all-lineage analysis required of lineage accumulation methods. These ultrametric trees were then subjected to diversification rate analysis using the R package LASER v2.3 (Rabosky 2006b) and a lineage accumulation deviation analysis that is somewhat novel. Figure 16 shows a simplified version of the BEAST posterior consensus with the main clades collapsed to represent relative size and selected sub-lineages super-imposed. Once again the absolute timescale is provisional – the key assumption is that the relative timescale is reasonably accurate. Where comparable, these results are consistent with the previous dating using the 147 taxon dataset. All analyses were done with both BEAST and PLRS trees, with a mix presented for brevity. The BEAST results show qualitatively similar but less distinct LTT patterns as the PLRS tree, with only non-significant deviation from constant diversification rates. This may imply constant rates but also it reflects the BEAST relaxed-clock method (unlike PLRS) being in part predicated on a Yule node height (speciation) prior, which will tend to smooth any LTT towards log-linear wherever the likelihood signal from the molecular data is relatively weak. The hadroid splits mentioned in the previous section are indicated again, along with era of the ‘Hill Gap’.

The hadroid PLRS lineage-through-time plot is shown in Figure 17: the overall impression is a broadly log-linear or slightly concave function consistent with pure birth or birth-death equilibrium (with minor death). Two other features are worth noting: 1) a decline in the most recent time interval towards the present (ca. 1-2 mya); 2) some deviation from linear at deeper level, assuming the timeframe around the mid-late Miocene ca. 5-10 mya. As the former (most recent time-interval) spans parts of the tree where the link between gene-tree and taxonomic discrimination and interpretation becomes uncertain, this part of the LTT is best excluded from analysis (Avise and Walker 1998; Rabosky and Lovette 2008; Ricklefs 2007b; Purvis et al. 2009). It is also necessary to exclude it in order to focus attention on the deeper possible non-linearity, because the sensitivity of the method increases with the number of lineages resulting in a tendency in fitting to tip patterns (Rabosky 2006a,b). Therefore LASER analyses

used the Truncate function to exclude the most recent 1-2 mya set of nodes, roughly demarked in grey on the figures.

As apparently slight as this deviation might appear, and while pure-birth or birth-death (with 40% extinction) are good fits, diversification analysis indicates some support for rejecting rate-constant equilibrium models (Table 3A). As current methodology only allows fitting to a limited set of discrete non-equilibrium models (such as a two-rate Yule model), none of which may best reflect the deviation, another approach was taken whereby the deviation between the observed data and the best fit model is compared to the deviation expected among stochastic simulations of the best fit model. Best fit models were inferred in LASER, and matching phylogenies simulated in Phylo-Gen 1.1 (Rambaut 2002). Log-linear regressions were then fitted to each of these and residuals taken to represent the deviation from a pure birth model. As there is no dedicated program available, this was done in the general statistics platform JMP v1.3.5, fitting only a log linear function, which equates to a pure-birth model. Fortunately, this is appropriate given the general LTT trend (as a *b/d* model with 0.4 extinction is similar to linear; see Appendix 5I). Due to computational limitations only 200 replicates were fitted. The first (oldest) 10 nodes were excluded from this fit, as this part of any cladogenesis simulation is prone to huge variance (Harvey et al. 1994; Ricklefs 2007b). To accommodate phylogenetic uncertainty a sample of Bayesian trees have been used to represent the observed result. The bottom panel in Figure 17 shows the results of this deviation analysis: the observed deviation approaches or extends beyond the expected stochastic variation around 5-7 mya on the figure. Several clades contribute to this signal (Figure 16 & below). Near the present and towards the base deviation also appears high but these are separate issues and better disregarded as discussed previously.

Therefore there just may be something in this deviation from equilibrium worth investigating further. An appropriate way to do this is to look for congruence among independent lineages by deconstructing this into the signal from individual clades comprising the hadroids. Of particular relevance here are the co-distributed but ecologically dissimilar (occupying different bioclimatic domains) lineages comprising clade 4 and sibling clades 1 & 2. Clade 3 has a largely separate distribution and includes a South Australian radiation which is only sampled at the generic level (11 of 31 species). Clade 4 contains the *Sphaerospira* plus sister lineages and therefore is older than that analysed by similar methods in Chapter 4 but only includes species and main phylogeographic units = 30 taxa (OTU). Clade 1+2 comprises 74 taxa almost all of which are named species or informal codes. This LTT breakdown is shown in Figure 18 plotting clade 4 and clade1+2 on the same scale. These lineages show quite different diversification rates and patterns with a contrasting trend coinciding with the deviation shown in the overall hadroid LTT (and to the general time frame of the mid-late Miocene), where the clade 4 slope decreases and the clade 1+2 increases. This is largely due to the radiation of clade 2 but qualitatively there appears to be a common independent trend in clades 1 and 2, and perhaps even 3 (caveats notwithstanding). Again, quantitative evaluation (Table 3B,C) excludes the tip component and while clade 4 (minus tip) cannot be distinguished from equilibrium (c.f. Chapter 4), for the PLRS trees at least, discrete models are the better fit, with opposite rate trends but similar shift times. It is worth noting that two strong contrasting or complimentary patterns could sum to a less strong pattern overall, irrespective of increases to the sensitivity of the method due to increased numbers of lineages.

In summary, overall we have a roughly log-linear lineage accumulation (unlike the pattern of declining diversification rates seen in many groups; e.g. Rabosky 2009a,b) but interrogating this further suggests some congruent region-wide signal, differentially affecting lineages occupying different bioclimatic domains with a relative decline in the mesic mesothermal clade 4 compared to the megathermal and

more xeric tolerant clades 1 and 2. This whole hadroid lineage wide pattern, seen (at least qualitatively) in separate sub-lineages, corresponds to several of the discrete ecogeographical divergences discussed above, and if we accept the tentative timescale, to the mid-late Miocene 'Hill Gap' proposed to be an era of considerable flux in the Australian biota. While perhaps no one piece of evidence is without doubt, together they do form a coherent scenario, where extrinsic factors produce temporally congruent but ecologically specific signals (Dolman and Hugall 2008; McPeek 2008).

Ecogeographical divergence

Figures 14 & 16 indicate some specific instances of divergence between more coastal mesic and inland xeric taxa. A possible approach to presenting this more analytically is by charting inferred changes in ancestral states along the phylogeny. For this, the camaenid taxa were classified into two categories, mesic and xeric, using various bioclimatic parameters. Four classifications were made using different combinations of bioclimatic indices resulting in slightly different assignments and numbers in each category (Figure 19D; see also Appendix 5G). Ancestral states were then estimated using both the symmetric and asymmetric Mk1 and Mk2 models in Mesquite (Mesquite v2.6; Maddison & Maddison 2009). Figure 19 shows two plots of the relative amount of state difference between daughter lineages through the hadroid phylogeny: A) state change per node by divergence bins (geometrically scaled to give more even numbers of nodes per bin); B) ratio of cumulative state change to cumulative nodes. Both charts show minimum and maximum values across the four classifications, by two models (=eight analyses), and 19A comprises three sets of divergence bins.

Ancestral state estimations are better thought of as null models showing an overall expectation according to a stochastic process rather than an estimation of actual ancestral state. In particular the standard stochastic models effectively assume a uniform landscape in space and time, and are therefore inappropriate to many historical processes. Most of the information may be in the difference between observed and null, and it may be better to think of ancestral state estimation in an ABC framework (Beaumont & Rannala 2004) where models representing different processes are simulated and the results compared to the observed. Notwithstanding these criticisms, measuring trends in amount of change may be more robust to assumptions, and the results highlight the previous narrative pattern of divergence between coastal and inland taxa. This is consistent with the pattern of relative clustering of state change (increase and then decrease) coinciding with the diversification deviations in Figures 17-19, which assuming the timescale, is around 5-10 mya.

For linage accumulation (LTT) patterns, even if sub-clades share a deviation, random phylogenetic noise (over and above specific effects of Bayesian priors) might be expected to result in the sum of these appearing relatively smoother, dispersing the signal. On the other hand if there are large scale limits to diversity then gains in one type will tend to be offset by losses in another and therefore in an analysis of the entire biota, summing across differential diversification among types will reflect this zero-sum quality (e.g. MacArthur & Wilson 1969; Hubbell 2001), resulting in apparently less overall deviation (Chapter 1 Figure 1). The individual LTT deviations are not particularly abrupt, large or extreme. While this might be because there is nothing there, or because of weak power inherent in the data and methods, it is also expected of a broad ecosystem trend, and the correlation to ecogeographical divergence is consistent with this interpretation. While occasionally nature will throw up dramatic events (Krakatoa, the KT boundary etc) we should not expect this or demand it or look only for such patterns. What I suggest here is that there is a signal of an era of change, of the turnover of biodiversity,

not a dramatic event but a process drawn out over evolutionary timescales – longer than speciation but shorter than background extinction, take to write over the past.

This section is an example of a by-lineage (or by-clade) analysis that depends on (near) complete sampling of monophyletic assemblages. Here we could combine the need for monophyly with that of ecogeographical coherence because nature provided some reasonably distinct examples. But often this is not the case and what we really wish to study is by region or by environment – by class not by clade. The following approach shares some methods used in community ecology of phylogenetic clustering and turnover within and among assemblages to investigate evolutionary trends.

By region, by bioclimatic category assemblage analyses

The dynamics of mesic-xeric diversification may be investigated by assemblage methods that do not require *a priori* division into clades. Broadly there are two ways to this, one via phylogenetic diversity methods used in community ecology – phylogenetic clustering or relative PD and phylogenetic turnover among assemblages (e.g. Goldberg et al. 2005; Graham et al. 2009). Assemblages can be local communities or spatial units – grid cell maps (e.g. Figure 5C), or larger regions or even biome types (e.g. Crisp et al. 2009). A second approach is via stochastic models of speciation, transition and extinction across categories within a phylogeny (e.g. Maddison et al. 2007; Fitzjohn et al. 2009; Rabosky & Glor 2010). Biodiversity can be divided up into categories that can be polyphyletic assemblages such as spatial units like grid cells or regions, and/or into ecological categories such as mesic and xeric.

Here I explore the relative phylogenetic structuring and diversification patterns between mesic- and xeric-like categories within the Far North Queensland region (FNQ). In the first instance this requires using a suite of bioclimatic parameters to classify, via the distribution data, the species into categories. Figure 20 shows an example of this for the FNQ region: using a few parameters including the PG indices, the 80 species of the FNQ region can be classified into three categories that map rather well to upland (mesothermal) rainforest, lowland (megathermal) rainforest and the rest – more xeric sclerophyllous woodland and vine thickets. In these approaches species are assigned to single category and *ipso facto* species turnover among categories = 1. Appendix 5N shows the phylogenetic distribution of the 80 FNQ species in the total east coast camaenid phylogeny pool.

The world rarely comes in such discrete categories but hypotheses do: while binning species into mesic or xeric is a crude simplification it is a necessary evil; it is not meant to imply that the world is thus but that if the categorization captures some reality and the hypothesis is sufficiently relevant, then the results should show through (c.f. Mayr 1976). By these criteria the FNQ region should be suitable as it is biogeographically distinct, rich in diversity patterned across strong environmental gradients (Chapter 3). Using this approach we can ask whether patterns of phylogenetic clustering within and turnover between the categories meso, mega and xeric show any consistency with the hypothesized direction of evolutionary trends - from old ancestral mesothermal mesic ecosystems to derived more xeric ones – proposed for Austral biodiversity and suggested previously in this chapter.

Results and additional details of methods are presented in Tables 4 & 5. The total camaenid diversity can be divided into hadroids and non-hadroids to provide phylogenetically independent comparisons. Using just the hadroids may represent the best option as they are an endemic group free from possible confounding effects of immigration of extra-limital megathermal elements from north (i.e. Melanesian

links such as clade 7, 10, *Meliobba* etc). Two statistics are used: 1) relative PD or phylogenetic clustering; 2) phylogenetic turnover PDβt. Turnover can also be compared to null expectations, therefore both can be described as a frequency statistic. Randomizations used both total and regional pools, with and without distribution range weighting. As the results are qualitatively similar and regional pool seems more appropriate (Hardy 2008; Vamosi et al. 2009), the equal weighting regional pool results are used in discussion.

Summarizing the results for relative PD (Table 4), in all cases it is less than the null, and while perhaps we should not expect separate groups to show the same patterns, the xeric category is consistently the most clustered; quantitatively by relative PD and qualitatively by significance ($p<0.01$). For turnover (Table 5), in all but one case PDβt is greater than null, and turnover between meso and xeric categories tends to be the greatest. Considering just hadroids the results are quite consistent: the order of relative PD (of phylogenetic clustering) from least to most is meso-mega-xeric; the level of turnover places the order of similarity as meso-mega-xeric. The xeric category is the only one significantly clustered while the mega-xeric turnover is the only one less than the null expectation. Results using total pool are qualitatively same (Appendix 5K). One interpretation is that the xeric domain has undergone recent endemic diversification and some of this is being accumulated in the more megathermal parts of the WT. Crudely, the maps of turnover (Figure 7) suggest some endemic diversification across southern and western fringes of the WT but do not indicate direction. More speculatively, can the pattern be interpreted as saying that the meso category is the oldest, ancestral one, the xeric category the youngest most derived one, derived mainly via the mega category? If so it is consistent with the prediction of the hypothesis of historical ecosystem trends.

For some of these types of analyses geographic distance can represent a confounding factor affecting turnover where comparisons spanning more distance will tend to have higher turnover. This means that assemblages with smaller, closer distributions may *ipso facto* tend to have lower turnover, due to the pervasive spatial effect (e.g. Figure 2). For example we might expect turnover between FNQ meso and mega to be relatively less than compared to xeric because they occupy smaller closer ranges, and comparisons within a region might be lesser than between distant regions (e.g. between mesic and xeric categories within FNQ versus between FNQ and MEQ mesic categories). However, in this instance such biases cannot explain the observed result.

Distribution overlap among major clades

A feature of many of these clade distributions is large range, substantial sympatric/local diversity within lineages but little overlap between. A number of the range boundaries correspond to biogeographical features but others show less clear associations. Of the balance of forces governing such large scale distributions, how much is historical biogeography, how much is abiotic environment, and how much is biotic competition?

A particular example that is amenable to investigation is the sibling trio of hadroid clades 1, 2 and 3 (Figure 21, Tables 6 & 7). Each of these is phylogenetically well defined and together form a series of adjacent ranges, in total extending from South Australia to south-eastern NSW all the way up to Torres Strait. There is some overlap but it is the lack of overlap that is striking (Figure 21B; Table 6). While each clade spans a large area, and with local diversity of 2-3, the proportion of area (as measured by $^1/_3$ degree grid cells) where they overlap is small: for clades 1 and 2 only 3.7% of the joint area; for clades

1 and 3, 5.7%. If these distributions were randomly assigned we would expect the overlap between clades 1 and 2 to be of the order of 9%, for clades 1 and 3 19%, 2-4 times the observed.

Do clades abut at diversity troughs, at environment gradients, or at biogeographic barriers? Competition presents us with something of a conundrum because within the range of each clade endemic sympatric diversity is widespread (30-40% of range has SD=2, 10% = 3 up to grid cell max SD=5-6). Therefore this involves a combination of within clade ecological divergence allowing co-distribution, and between clade ecological competition constraining range boundaries (c.f. Schemske 2009). If distribution is governed by environment (i.e. distribution is in some way at equilibrium) the areas at the periphery might well be diversity troughs.

The boundary between clades 1 and 2, and 1 and 3 are near major biogeographical features (the Burdekin Gap, SEQ-Border Ranges) but do not overlay them (as compared to some other clades: see Figure 10), being offset by 100km or more from the Burdekin Gap (with a clade 1 disjunct northern outlier *Montheithosites*) and to the south straddling the St Lawrence Gap towards the northern end of SEQ, with a long front of contact through inland Qld. The turnover between clade 1 and 2 south of the Atherton region does align more with sub-regional divisions in Wet Tropics (see Chapter 3, and also Moussalli et al. 2009). The western edge of clade 1 contacting clade 3 is mostly due to a single species *S. mattea*. This western distribution range is somewhat paralleled in clades 5 & 10 (Figure 10), and loosely follows the Great Dividing Range and northern extension of the mesothermal domain (Nix 1982).

The role of extrinsic environmental factors may be addressed with bioclimatic modelling: how much of the distribution and hence lack of overlap can be explained by bioclimatic factors? This is not to say that they do control it but how much of the range pattern can be modelled by a few simple factors? To investigate this, a simple BIOCLIM model is applied to the distribution data for clades 1-3 (Figure 21C), using the bounding values of a few parameters (AMT, AMP PDQ, TWQ; see Chapter 3 & 4 for further details). At this large spatial scale quite a lot of the distribution can be explained by these few parameters, especially the abutting range boundaries with only 6% overlap (Table 6). These are large clades that encompass substantial evolutionary history and environmental change. However the two main environmental gradients contributing to the fit, the large scale latitudinal range and the gradient across the Great Dividing Range (e.g. Nott 2005), have even longer histories and hence a continuous pervasive influence.

This poses the question of whether the areas of overlap are areas of low diversity (troughs or sinks) or high diversity (accumulation and competition). This is not easily assessed (Table 7). Overlap necessarily means minimum SD=2. Let SD_2 refer to grids with species diversities ≥2 and overlap grids as SD_o. The average SD_2 for clades 1, 2 and 3 are 2.50, 2.26, 2.38; with overall of 2.46. Near to the overlap between clades 1 and 2 SD_2=2.7, compared to SD_o=2.90; near the overlap between 1 and 3 SD_2= 2.42, while SD_o=2.57. The results suggest that diversity in the overlap is as high as elsewhere, not less, suggesting that the overlap areas are not troughs *per se* but overlap results in carrying capacity being split between the two lineages. Perhaps no one rule suits all cases and it is a balance of processes, effective over different scales that govern the boundaries: broadly constrained by geography, more proximally by environment and locally by carrying capacity.

Here I take the opportunity to discuss the rationale behind using distribution modelling in evolutionary biology and defend the use of the simple BIOCLIM model. The choice of model type – of complexity

of model – depends on the question it is used to address not on the need for sophistication *per se*. For the purposes of investigating why ranges are not similar (as in the above case), logically it is preferable use as simple a model as possible and one that does not include a procedure that automatically down-weighs regions not occupied, for example ones that use logistic regression on presence and absence data (underlying most methods: Guisan & Zimmermann 2000; Phillips et al. 2006; Moussalli et al. 2009). Alternatively if the question is to investigate if bioclimatic distributions are similar (for example vicars) it is more convincing if complex models with many parameters are used. The object of distribution modelling is not to provide a fit as close as possible to the observed distribution but to see how much of that distribution can be explain by a few simple bioclimatic factors – a null model - and it is in deviations away from this where much of the interest lays, or should lay (c.f. vanDerWaal et al. 2009; Lobo et al. 2008). There are many reasons why a range may not fit bioclimatic parameters, such as historical contingency, competition, extinction, vicariance; using sophisticated systems to create complex (abstruse) algorithms with numerous terms is not necessarily the rational choice. Nevertheless, for the question here of large scale distribution of clades 1,2 and 3, more complex GLM presence/absence logistic models with quadratic terms, of the type widely use (and considered necessary; Elith et al. 2006; Araujo and New 2007;) produce much the same results – good explanation of the range boundaries and limited overlap (not shown but see Figure 13).

Sympatric diversification

It is appropriate to end with an investigation of the emergence of sympatric diversity: the 'trajectory' from allopatric populations to sympatric species, where 'speciation can be said to be complete' (Mayr and Diamond 2001).

There are 321 species across eastern Australia but for any one area such as 50km grids there are only up to around 20, for local areas (within 5km) up to 12-15 and at any one point a maximum (α alpha diversity) of 10-12 (see SD/area graph Chapter 1 Figure 10). Across the entire phylogeny there are no (clear) examples of sympatric sister species. However, repeatedly across the phylogeny sympatric species groups appear at quite low divergence levels. Figure 22 shows a histogram of divergence level for sister tip lineages and sister sympatric lineages in the hadroids. While sister species (taxa=tips) can be as close as 0.003, the minimum divergence (lineage depth) for sympatric species is 0.04-0.05.

Local diversity comprises some members belonging to the same clade and members of much more distant lineages: some element that can be described as regional endemic diversification, within an assemblage of what is in effect an assortment of ancestral diversity (Gaston 2000; Ricklefs 2006; 2007a). It is the recent endemic component that is most relevant to investigating the emergence of sympatric diversity. Within the hadroids the youngest (least divergence depth) examples represent eight cases across the four clades, spread across four different biogeographic regions: FNQ, MEQ, SEQ and the Sydney Basin (Table 8). This is based on known widespread local co-occurrence, as well as analysis of the distribution dataset at a fine scale (within 2km radius). The *P. duralensis*-SN20 pair is somewhat questionable but is include here provisionally.

From the phylogeographical arrangement of populations and sibling species (Chapters 3 & 4) it is clear that speciation largely proceeds via allopatric and peripatric population structuring: towards the most recent part of the phylogeny there is a profound relationship between geographic and genetic distance. The details of this very much depends upon ecology and environment but Figure 23 shows a crude general trend for the hadroid, with clades 1 and 4 showing the strongest structuring (greatest

divergence per distance) as might be expected of their ecology and environment (more mesic, across regions of eastern Queensland; see Figure 15).

Figure 24 shows the trend in phylogenetic divergence and geographic range of the nodes within the clades spanned by each of these eight examples of sympatric diversity. From Figures 22-24 we can see there is a nested series of divergences and ranges, a time-lag, before the emergence of sympatry. Is this time-lag in emergence of sympatric diversity just an extension in time of the finer scale geographic trend or do other factors contribute? Exactly how to investigate this or how to interpret any results is not obvious (c.f. Lynch 1989; Barraclough et al. 1998; Mayr and Diamond 2001; Malay and Paulay 2010) but if there were factors delaying sympatry other than just time required for geographic dispersal then perhaps range per divergence should tend to fall to the right of the background trend, or the opposite if restrictions on range expansion were reduced. An example could be where dispersal requires adaptation: if this takes time after divergence (speciation) it will delay sympatry but if it drives speciation it may accelerate dispersal.

For most of the eight examples in Figure 24 the sympatric node appears to be an extension of the background trend, although most fall above the trendline, which if anything suggests no difference from an extension in space and time of the finer scale pattern. These most recent cases of sympatry tend to contain more PD and subtended nodes (Appendix 5L) indicating they belong to groups undergoing relatively higher levels of diversification, as is consistent with regional scale trends seen in Figures 5 & 6 (e.g. northern MEQ, Sydney Basin). These are regions of high endemic diversity with dense fine scale population/taxon structuring. Thus while the results are not enough to claim a sign of adaptive divergence they are not entirely inconsistent with it.

A Mayrean perspective

Site to site across regions, there are many different combinations of species comprising sympatric diversity, however, this complexity can be broken down into local representatives of a smaller set of lineages (c.f. genera, families). Thus of the entire diversity, all instances of sympatric diversity can be represented by a distinct unambiguous subset of nodes within the phylogeny of the entire diversity. For each node we can ask, does one daughter lineage contain any species that is in sympatry with any species in the other daughter lineage? If so, that node is assigned to the sympatric category, if not, it is assigned to the allopatric category. For the hadroids, of the entire diversity as represented by the 164 species and hence 163 nodes in the phylogeny, 30 nodes subtend all instances of sympatric species: for each of these nodes the daughter lineages contain species that are in sympatry with one another – for the rest of the nodes the daughter lineages are not sympatric.

The following analysis is quite speculative but it is based on the fact that the sympatric diversity simply, exactly and unambiguously can be partitioned into those phylogenetic nodes that delimit sympatric diversity and the rest that are allopatric. Diversity is complex: hundreds of species in thousands of communities but this can be cut down, into a simple precise graph (notwithstanding some phylogenetic uncertainty). Therefore I present it for interest, what it means is the question. The categories can be compared by cumulative distribution of node depths. While this is analogous to lineage accumulation (aka lineage-through-time) approaches, it is better to avoid thinking of it in terms of diversification rate theory. Figure 25 shows a cumulative plot of node by divergence for the hadroids, broken down into the sympatric and the allopatric components. This is just for the PLRS Bayesian consensus tree and does not attempt to account for phylogenetic uncertainty.

There are quite distinct differences between the two categories, and within them. The shift in accumulation slope between sympatric and allopatric occurs around 0.14 divergence corresponding with the deviation observed in the total hadroid lineage accumulation plot (bottom plot and Figures 15-17). This marks a 4-5 fold difference in slope between steep allopatric and shallow sympatric accumulations. What does this mean? Broadly speaking we can imagine a 'meta-population' process with units of varying distinctiveness and age where some components die out, others coalesce (merge), and others goes on to become long lived lineages, finally even sympatric diversity. The relative proportion or rate at which these sympatric elements emerge is measured by this lineage accumulation pattern. Beyond the population coalescent and quasi-stochastic realm of species birth and death, older lineages increasingly show different dynamics (behaving increasingly resistant to extinction; c.f. Hubbell 2001) where sympatric diversity is governed by ecological factors sorting 'old wine into new bottles' – species are older than the communities they comprise (Joe Connell pers. com.). This aspect has little to do with the emergence of endemic sympatric diversity, and accumulation plots begin to look similar and eventually merge (allopatric → 0, sympatric → total). This would be an extension of the process hypothesized for the *Sphaerospira* lineage in Chapter 4. Just how varied in shape and relative intensity such plots can take may be worth investigating, empirically and also theoretically, exploring different speciation processes and phylogenetic and biogeographic scales. For example, adaptive radiations might be expected to show relatively quicker shift to the sympatric category that vicariance. Biogeographical sub-structuring may limit accumulation of sympatry, set a minimum base to the allopatric component. What could influence the shape (e.g. sharp transition or smoother gradation) of the plots? What effect might extinction have in re-setting the accumulation patterns?

The significance of the coincidence of the observed shift with that of the differential lineage diversification, ecogeographical divergence, and hypothesized era of change in Australian ecosystems, remains to be explored. However, consider a pre-existing ecosystem with established patterns of community diversity that then suffered an era of differential extinction and diversification, and re-ordering of diversity. Perhaps the widespread ancestral sympatric component is less prone to extinction – due to large range and wide diversity – and hence a relatively larger portion survives compared to the younger local endemic sympatric diversity, then subsequently the local emergence process begins again: hence the appearance of a switch in accumulation curves. It may be of value to compare these analyses with various theoretical expectations of turnover and mass extinction.

A few final comments

1) Consistency among inferences: No one diversity index captures all of the complexity but comparison of several indices can help discriminate patterns. There are some inconsistencies or contradictory inferences among some of the multitude of analyses but given all the complexities and approximations it would be troubling if there were not. Some of the diversification indices show biases at low levels of SD. This is likely always to be a problem and it may be necessary to further exclude low diversity areas. Many of the patterns appear at regional scales or smaller, highlighting the limitations of interpreting the diversity in terms of overarching trends but also that many of the processes occur at local scales and are hence obscured by the fixed 50 km scale used. The analyses of Chapter 4 emphasizes the continuity of the phylogeographical speciation process in clade 4 while Chapter 5 emphasizes the concept of differential diversification and ecosystem turnover but perhaps this is largely because one studies a single group, the other the diversity of groups. At the end of the day, so to speak, does the macroecological gambit pay well, or is it better to break the diversity down into component parts, analyse each in detail and then attempt to reassemble the complexity from the

diversity of patterns and inferred processes? At the very least spatial mapping of the diversity and diversity patterns is essential in appreciating the tangled bank from the bottom line.

2) Determinants of diversity: In summary, in the case of the camaenid land snails of eastern Australia, species diversity has nothing to do with speciation rate and everything to do with environment and environmental change. Broadly, the results resemble some of the findings of Davies the al. (2007) and Weir and Schluter (2008), in the complex associations of diversification and diversity, and of environmental correlates of relative PD; that species diversity is not so much associated with speciation rate but relative extinction (as compared to Cardillo et al. 2005), or more generally 'carrying capacity' (e.g. McPeek 2008; Rabosky 2009a,b). As most lineages have been present prior to and throughout the mid-Miocene onwards, species diversity has nothing to do with time of immigration as proposed by Wiens et al. (2006; 2009). The regions of high species diversity have low signals of diversification while the regions of medium to low diversity have high signals of diversification (despite some methodological biases). Diversity is governed by environment via extinction: in the rich regions extinction is low and so replacement rate is low; in the poorer regions the appearance of high rate of diversification is merely the tips waiting to be pruned. Most likely extinction is largely intermittent reflecting historical fluctuations but across evolutionary scales pervasive. The above summary is specific to these snails, different organisms will have different patterns but I suggest that casting questions of diversity in terms of speciation or extinction rates are in effect elaborate ways of restating the problem: it is unlikely ever to be speciation rate except in special cases, while extinction is largely a agent via which more fundamental forces act. Diversification rate does not *explain* diversity it just *describes* it.

3) Diversification versus turnover: One way to read the tree is to see it as increasing diversity through time; at a glance it appears log-linear and diversification models infer a limited amount of extinction. There were fewer species in the past and more today, for example in colonization of new regions or adaptation to expanding environments. However if total diversity is strongly limited by environmental factors, room for the 'new winners' need to be at the expense of the 'old losers'– the zero-sum game element (e.g. MacArthur 1969; Hubbell 2001). This is at the heart of proposed history of environmental change and ecosystem turnover in Australian biodiversity. The difficulty is that while the winners are self evident – are the extant lineages in the diversity, the losers can only be inferred; and in many cases may belong to entirely unrelated groups, i.e. not represented in the same phylogeny. Thus the tree, and the lineage accumulation, should not necessarily be seen as a growth of diversity so that there was little in the past and more now. Many different processes can yield the same LTT pattern, and the method alone can distinguish few (see Chapter 1). To disentangle the possibilities requires other information. As in the discussion of pattern and process in Chapter 4, consider that the phylogeny represents extant lineages only of a system that has undergone some turnover. For example the *b/d* model extinction fraction of 40%, which only produces minor curvature in LTT, is more than adequate to encompass substantial turnover. Furthermore, while LTT in principle can infer random extinction in an equilibrium *b/d* system, the more lineage-specific the extinction, the more invisible it is to LTT methods *per se*. Then there is the strong relationship between geographic scale and phylogenetic distance among the more recent allopatric dominated part of the phylogeny, contributing to the log-linear appearance near the tips.

The Hubbell (2001) neutral theory posited a fractal or scale invariant self-similar quality to the underlying phylogeny however no such pattern is seen in species level phylogenies, and is only found in more coalescent processes. The difference is in the different extinction processes that occur at

different scales. An example of this is the trans-species phylogeny in Chapter 4 where only the all-haplotype tree resembles the Hubbell expectation (c.f. Womble 1951; Mallet 1995; Monaghan et al. 2009). A system exhibiting equilibrium constraints on diversity may be expected to show something of a density dependent convex curve LTT. A system exhibiting an equilibrium *b/d* process is expected – given enough time – to show something of concave curve. The net summary shape of the lineage accumulation depends on the combination of extinction processes, and these vary across the diversity scale. At the lowest divergence and smallest local geographical scale the process resembles a coalescent ($b=d$), at a larger phylogenetic and regional geographic scale it is more like a meta-population stochastic extinction process ($b>d$), at a global scale the sympatric diversity assembled among deep lineages is the remaining net birth only ($d=0$) component. Large geographic range and ecological divergence all but ensures lineage immortality. Thus the broadly linear LTT can reconciled with a system carrying similar levels of diversity through time but high turnover because: 1) it is compatible with at least 40% turnover as is; 2) there can be more lineage specific extinction hidden in this, as is implied in the comparative analyses; 3) the nested phylogeographic nature of a lot of the more recent allopatric lineages and inherent death-lag in *b/d* processes (c.f. Chapter 4); 4) some actual recent increase due to change suiting seasonal megathermal groups, from which the camaenids appeared to have originally stemmed (Solem 1979a; 1997).

Because the analysis spans the entire biota it summarizes both the radiating groups and (at least some of) the groups that they replace. Considering that many lineages have been widespread prior to and during the period of change in question (mid-Miocene onwards), and the relative depth of the deeper parts of the phylogeny, twice as deep as the Burdekin Gap reference point (ca 13 mya), there is more than enough scope for the phylogeny to span substantial turnover. Turnover should be distinguished from mass extinction. While mass extinction need not be instantaneous (just a relatively short with respect to the scale of the LTT diversification trend), turnover implies a process of change over an evolutional time scale. Mass extinction implies (and for it to be detectable, needs) a very large proportion of extinction spread across most lineages and sufficient time afterwards for subsequent diversification, resulting in a common pattern across lineages. Turnover is different in that it is not necessarily the same set of lineages suffering extinction and exhibiting radiation: it is not mass extinction across the board but lineage specific differential diversification. Trees should be read from the tips down not the bottom up, through the lens of evolution not the mirror of classification.

4) Accumulation of lineages versus accumulation of diversity: Considering diversification *per se* there is substantial difference among lineages. For example clades 1-2 have a higher diversification rate than clade 4, being half the age with more than twice the diversity. However lineage accumulation plots indicate that clade 4 can appear as fast as clade 1+2, at a point where most ($^2/_3$) lineages in clade 4 are phylogeographical. Beyond this (at lower divergence levels) clades 1-2 have much higher diversification rate but this is due to an allopatric array of taxa at ever smaller geographic scales– akin to phylogeography but with names. What do we mean by diversity? Sympatric diversity is a key aspect because: 1) it is a quintessential meaning of diversity; 2) it is freer of taxonomic vagueness (Mayr 1976; Isaac et al. 2004; Hey 2009), a more objective reality. Considering this sympatric component only, curiously there appears to be much less difference in diversity among lineages. For the sake of argument, many of the tips are arbitrary, ephemeral, perhaps better thought of as sub-species, and only at the level of sympatry does the real story emerge.

Chapter 5 Tables

Table 1. Multivariate general linear modeling of Species Diversity

A: Climate model	Source	Estimate	F Ratio	Prob>F
linear terms only	TWeQ	0.04263	53.1767	*<.0001*
RSquare Adj = 0.5461	meso	0.08892	29.1067	*<.0001*
	mega3_sd	0.86184	63.9431	*<.0001*
	area	0.00007	0.0327	*0.8567*
linear and quadratic terms	TWeQ	0.04323	56.0284	*<.0001*
RSquare Adj = 0.568	meso	0.35667	18.5462	*<.0001*
(no interaction terms	meso*meso	-0.00278	10.2604	*0.0016*
were significant)	mega3_sd	-0.07060	0.0414	*0.8389*
	mega3_sd*mega3_sd	0.11184	6.5156	*0.0114*
	area	0.00029	0.6324	*0.4273*
B: Phylogenetic Diversity model	Source	Estimate	F Ratio	Prob>F
linear terms only	ln(sigsd/sigpd)	-1.02943	13.0331	0.0004
RSquare Adj = 0.0998	corrPD	-1.24617	0.4411	0.5073
	corr_sd10drlr	-2.13734	2.8421	0.0932
	area	-0.00115	6.3275	0.0126
linear and quadratic terms	ln(sigsd/sigpd)	0.26446	0.1277	0.7211
RSquare Adj = 0.109	ln(sigsd/sigpd)*ln(sigsd/sigpd)	-0.20682	3.2042	0.0748
(no interaction terms	corrPD	-17.82082	1.8239	0.1782
were significant)	corrPD*corrPD	8.53027	1.7345	0.1892
	corr_sd10drlr	-2.93593	4.3303	0.0386
	area	-0.00117	6.635	0.0106
C: Combined model	Source	Estimate	F Ratio	Prob>F
linear terms only	ln(sigsd/sigpd)	-0.00854	0.0015	0.9687
RSquare Adj = 0.540	corrPD	0.41327	0.0787	0.7794
	corr_sd10drlr	0.19174	0.0426	0.8368
	area	0.00010	0.0639	0.8006
	TWeQ	0.04195	42.5292	<.0001
	meso	0.08976	27.0599	<.0001
	mega3_sd	0.85747	59.3837	<.0001
linear and quadratic terms	ln(sigsd/sigpd)	1.20526	5.7361	0.0175
RSquare Adj = 0.575	ln(sigsd/sigpd)*ln(sigsd/sigpd)	-0.20610	6.6862	0.0104
	corrPD	0.71028	0.2422	0.6231
	corr_sd10drlr	-0.34281	0.1422	0.7064
	area	0.00022	0.3352	0.5632
	c8_twq	0.04362	48.9371	<.0001
	meso	0.34314	15.9305	<.0001
	meso*meso	-0.00263	8.571	0.0038
	mega3_sd	-0.10859	0.0979	0.7546
	mega3_sd*mega3_sd	0.11715	7.128	0.0082

Total of 231 grid cells: SD>1 with 20 excluded due to low area.
Weighted by p-invPD: Observations (or Sum Wgts) = 10.028.

Climate variables:
1) climate (meso) = mesothermal PGI
2) seasonality (TWeQ) = mean temperature of the wettest quarter
3) grid heterogeneity (mega_sd) = standard deviation of the megathermal PGI within the grid
Three diversification indices:
1) relative PD (rPD_av)
2) endemic diversity ratio [ln(sigsd/sigpd)]
2) corrected DRL ratio (delta_sd10drlr)

Table 2. Environmental model for relative PD.

	Source	Estimate	F Ratio	Prob>F
Stepwise selection	AMP	0.00031	6.2666	0.013
linear and quadratic terms	AMP*AMP	3.38E-08	2.5761	0.1099
RSquare Adj = 0.442	TWeQ	-0.01065	4.5398	0.0342
	TWeQ*TWeQ	0.00002	2.7507	0.0986
	PWaQ	-0.00115	27.9556	<.0001
	DEM_Std	0.00038	3.8876	0.0499
	lat	-0.02423	1.8972	0.1698
	lat*lat	-0.00128	9.9923	0.0018
	area	0.00016	1.8371	0.1767
	area*area	-7.13E-08	3.2715	0.0719

Eight variables used: Annual mean temerature AMT; Annual mean precipitation AMP; Temperature of the wettest quarter TWeQ; Precipitation of the warmest quarter PWaQ; Standard deviation of elevation DEM_sd; Standard deviation of mesotherm PGI meso_sd; latitude; area.

Table 3. LASER diversification rate analyses of hadroid lineages (LTT).

A: Hadroids n=164/120	model	dAIC	r1	a	st	*p*
PLRS	pure birth	4.54	0.19			*0.07*
	b/d	4.25	0.15	0.39		*0.08*
	DDX	5.12	0.12			*0.05*
	yule2rate	5.25	0.16		5.58	*0.05*
	yule3rate	0.00	0.16		8.04	*0.50*
BEAST						
	pure brth	8.09	0.19			*0.01*
	b/d	7.60	0.14	0.40		*0.02*
	DDX	8.61	0.11			*0.01*
	yule2rate	8.35	0.16		5.09	*0.01*
	yule3rate	0.00	0.16		4.81	*0.50*

B: Clade4a n=30/28	model	dAIC	r1	r2	a	st	st2	r3	*p*
PLRS	pure birth	0.19	0.168						*0.277*
	b/d	0.096	0.087		0.666				*0.29*
	DDX	1.21	0.084						*0.171*
	yule2rate	0	0.107	0.242		5.71			*0.306*
	yule3rate	0.486	0.114	5.001		5.42	5.40	0.22	*0.24*
BEAST	pure birth	0.138	0.141						*0.285*
	b/d	0	0.074		0.661				*0.306*
	DDX	0.991	0.07						*0.191*
	yule2rate	0.467	0.094	0.196		6.55			*0.242*
	yule3rate	1.324	0.132	0.971		3.14	3.01	0.11	*0.164*

C: Clade1+2 n=74/60	model	dAIC	r1	r2	a	st	st2	r3	*p*
PLRS	pure birth	1.182	0.247						*0.231*
	b/d	3.175	0.239		0.053				*0.089*
	DDX	3.164	0.268						*0.09*
	yule2rate	3.698	0.362	0.221		7.76			*0.07*
	yule3rate	0	0.322	43.31		7.76	7.76	0.22	*0.409*
BEAST									
	pure birth	0	0.226						*0.409*
	b/d	0.455	0.151		0.513				*0.326*
	DDX	1.121	0.134						*0.237*
	yule2rate	1.63	0.203	0.365		2.67			*0.186*
	yule3rate	0.042	0.207	20.25		2.63	2.63	0.31	*0.397*

MrBayes posterior consensus PLRS; BEAST posterior consensus; PLRS includes *Vidumelon*, BEAST does not; truncateTree omit.time=2; n= total lineages/lineages after truncation; *p* = dAIC test statistic using n=truncated lineages, 5000 simulations; DDX = density dependent exponential function; r = rate, st = shift time; LASER v2.3.

Table 4. Relative PD of bioclimatic categories of FNQ camaenid species

	all			hadroids			rest		
category	n	relPD	*p*	n	relPD	*p*	n	relPD	*p*
meso	30	0.87	*0.06*	13	0.98	*0.35*	17	0.76	*<0.01*
mega	16	0.9	*0.24*	9	0.83	*0.09*	7	1.00	*0.46*
xeric	34	0.71	*<0.01*	20	0.75	*<0.01*	14	0.72	*<0.01*

n = number of taxa in category
relPD = relative PD = observedPD/null PD
p = proportion of null randomizations with PD exceeding observed PD
null determined using regional pool of the 80 FNQ taxa using 300 replicates, equal probability of inclusion. See Figure 20 for bioclimatic and spatial distribution maps.

Table 5. Phylogenetic turnover among biclimatic categories of FNQ camaenid species

	meso			mega		
	βt	rel βt	*p*	βt	rel βt	*p*
all						
mega	0.72	1.22	*0.98*			
xeric	0.82	1.48	*>0.99*	0.75	1.27	*>0.99*
hadroids						
mega	0.62	1.43	*0.99*			
xeric	0.67	1.56	*>0.99*	0.42	0.94	*0.31*
rest						
mega	0.79	1.18	*0.97*			
xeric	0.91	1.5	*>0.99*	0.92	1.36	*>0.99*

As for relative PD but rel βt is observed βt/null.
p = proportion of null randomizations with βt exceeding observed.

Table 6. Observed and modelled ranges for hadroid clades 1, 2 and 3

	observed	Bioclim
clade 1	179	191
clade 2	99	576
clade 3	422	1456
union of 1 and 2	268	689
union of 1 and 3	571	1608
overlap of 1 and 2	10	78
overlap of 1 and 3	30	39

Numbers of $^1/_3$ degree grid cells all records
Modelled in 1000 km^2, 95% bioclim bounds
Bioclim model parameters AMT, AMP, PDQ, TWeQ

Table 7. Species diversities for hadroid clades 1, 2 and 3

	SD		SD_2		SD_o
	overall	adjacent	overall	adjacent	
clade 1	1.55		2.60		
clade 2	1.52		2.26		
clade 3	1.49		2.38		
combined	1.60		2.46		
1 and 2 overlap		2.05		2.77	2.90
1 and 3 overlap		1.72		2.42	2.57

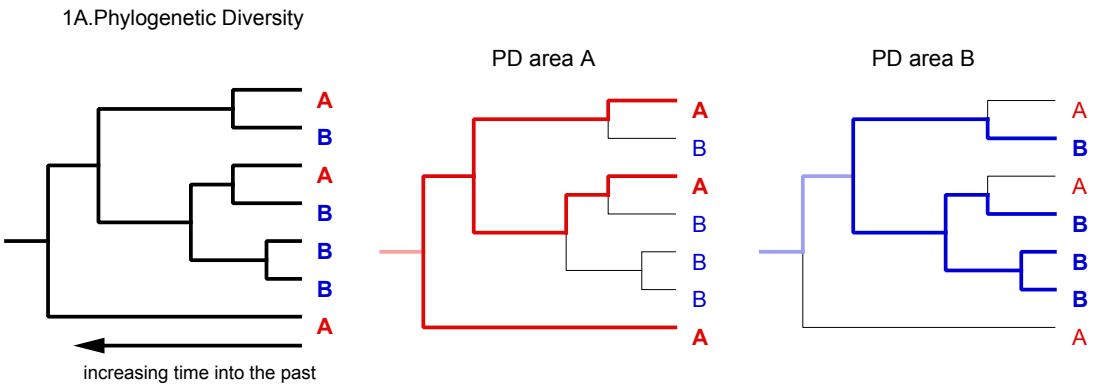
Using $^1/_3$ degree grid cells all records
SD_2 = SD of grids with two or more species
SD_o = SD of grids in the vicinity of the range overlap
37 cells adjacent to 1 and 2
123 cells adjacent to 1 and 3
See Figure 21 distribution maps

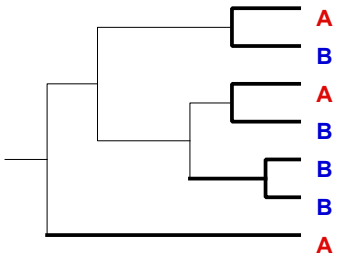
Table 8. Eight most closely related hadroid sympatric species

taxon1	taxon2	clade	region	av_div	range	nodes	PD	overlap
P. duralensis	SN20	3	Sydney Basin	0.044	539.4	8	0.231	0.41
S. rawnesleyi	BL33	1	FNQ	0.063	250.0	8	0.332	0.47
S. etheridgei	*S. coxi*	1	MEQ	0.072	282.7	9	0.318	0.29
M. marshalli	*M. middenese*	3	Sydney Basin	0.084	272.5	11	0.565	0.50
S. gemma	*S. praehadra*	2	FNQ	0.096	354.6	10	0.578	0.31
M. corneovirens	*P. duralensis*	3	Sydney Basin	0.125	680.0	21	1.082	0.56
S. blomfieldi	SQ1	4	SEQ	0.141	804.8	13	1.144	0.19

Maximum linear range in km of all individuals in clade defined by MRCA of the sympatric species
Divergence depth in PLRS Bayesian consensus tree
Number of nodes in clade
PD is sum of branch lengths in clade
Overlap is (sum of species area - net clade area)/net clade area
Distribution area based on 50 km grid cells occupied

Figure 1. Schematic definitions of Phylogenetic Diversity (PD) indices





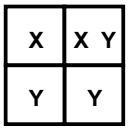
Total turnover
DPD = $b+c$

Relative turnover
bt = $(b+c)/(2a+b+c)$

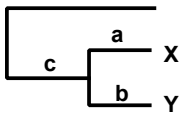
where
b = branches unique to A
c =branches unique to B
a =branches common to both

1C. Endemic Phylogenetic Diversity

Distribution of species X and Y

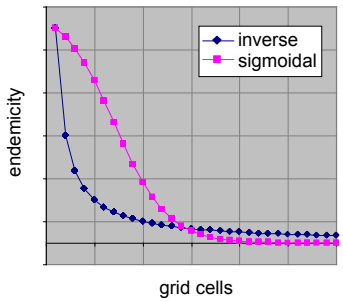


Phylogeny containing species X and Y



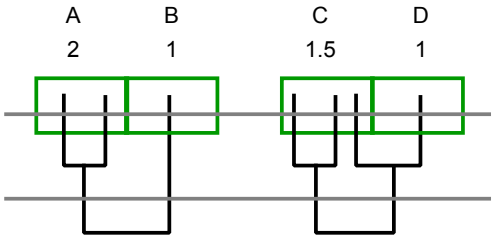
For inverse function endemic quotient:

species X = 1/2; Y = 1/3
branches a = 1/2; b = 1/3; c=1/4

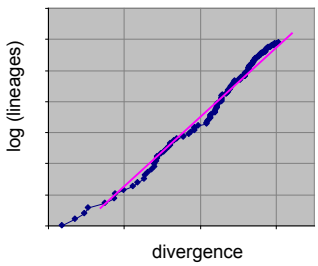


1D. Spatial Lineage by Divergence

Four senarios: sibling lineages occupy same grids (sympatric) or different grids (allopatric)



Lineage-though-time plot of entire phylogeny



Gross doubling time
(using slope = pure birth rate):

all: ~ 0.1 divergence
hadroids: ~ 0.06

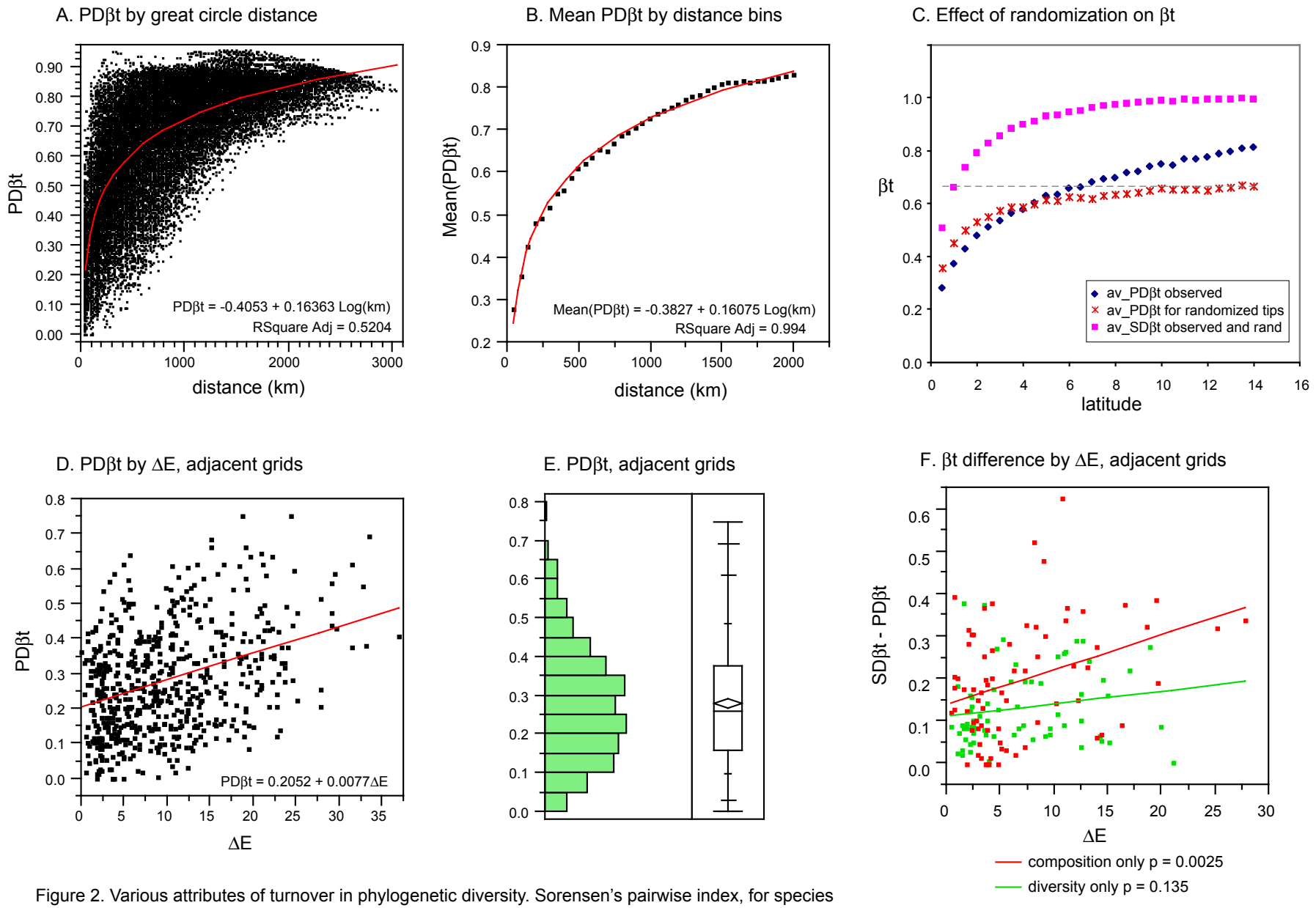


Figure 2. Various attributes of turnover in phylogenetic diversity. Sorensen's pairwise index, for species diversity (SDβt) and for phylogenetic diversity (PDβt). Environmental difference (ΔE) is Euclidean distance among the three Nix PGI; randomized tips means species names have been randomized across the phylogeny; βt difference is SDβt minus PDβt; composition only is for grids with equal SD (n=74), diversity only is where one grid is a subset of the other (n=77); see also Appendix 5N.

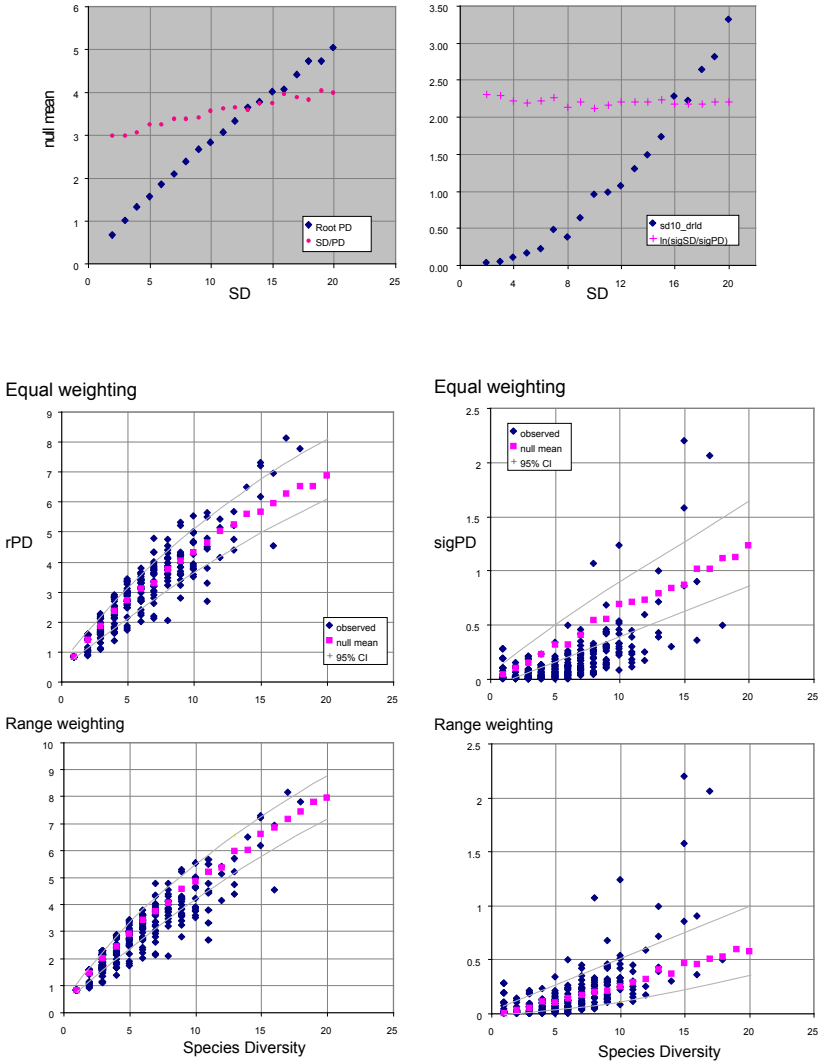


Figure 3. Examples of intrinsic correlations between various PD indices and SD based on randomizations. Top: Root PD, SD/PD ratio, endemic diversification ratio [ln(sigSD/sigPD)] and local lineage accumulation ratio (species to 0.1 divergence DRL difference). Randomization using equal probability of inclusion. Bottom: Correlation between PD and SD for observed and randomized taxa sets. Left: Root PD, right: Sigmoidal function endemic PD; with randomizations done with equal probability for all taxa (top) and probability of inclusion weighted by range (bottom). Null expectations are based on 300 random samplings for each SD level 2-20 from the pool of east coast species in the 327 taxon tree. For each random draw a species is selected from the pool (without replacement) with either equal probability, or probability weighted in proportion to the observed range (in number of $^1/_2$ degree grid cells).

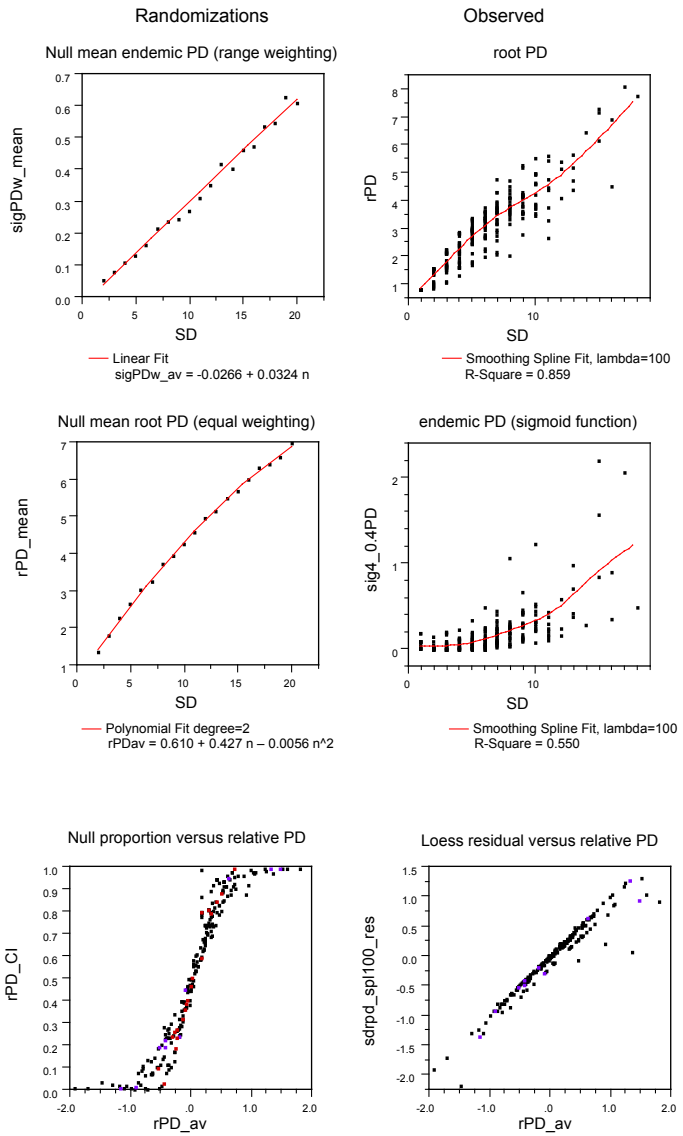
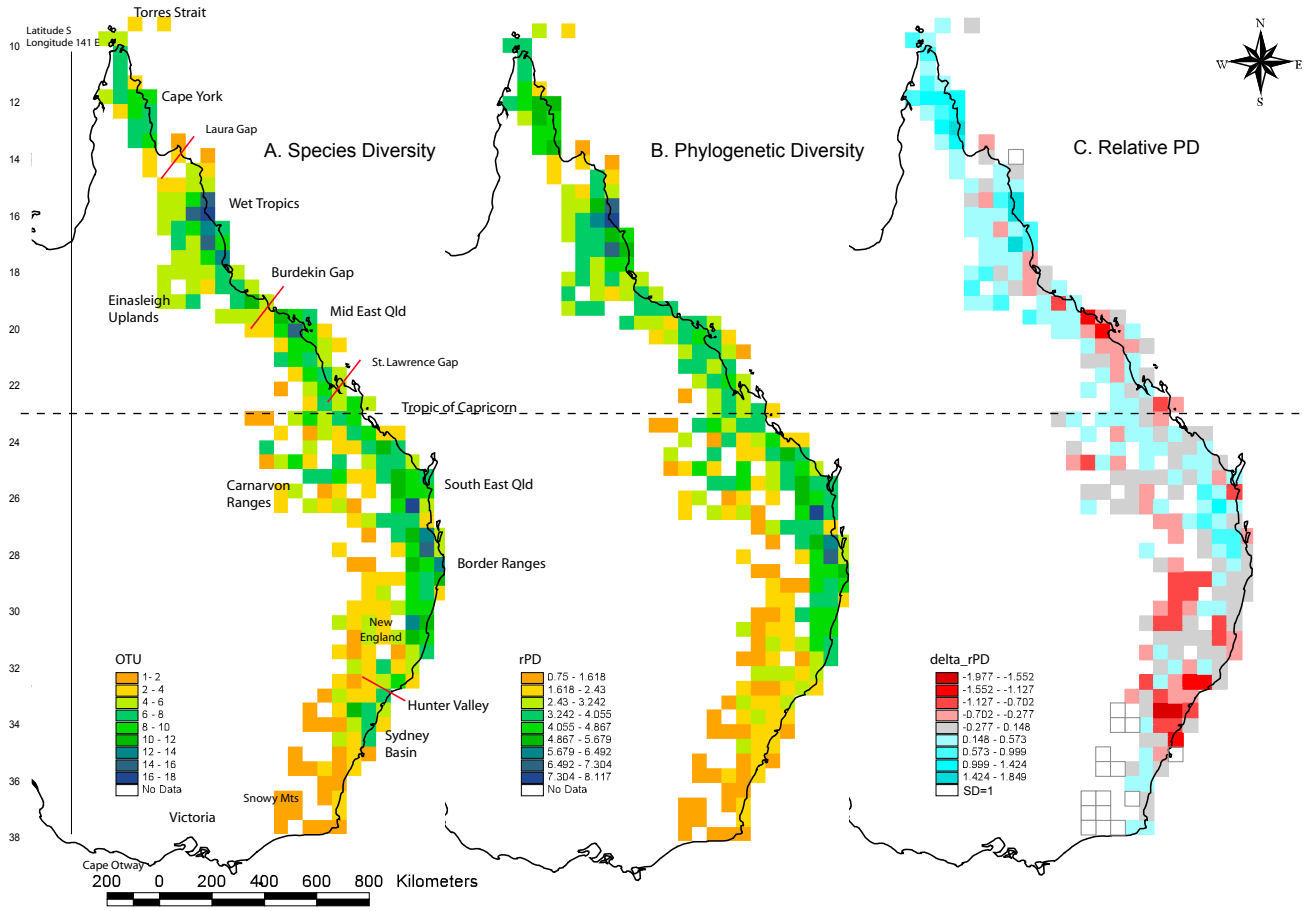


Figure 4. Various characteristics of PD indices. Top: Regression analyses for determining relative PD indices. Left shows intrinsic relationship with SD via randomizations, right shows loess regressions of observed data. Bottom: Relationships of various relative measures. Left shows relative PD versus proportion of times the null PD is less than the observed PD; right shows relative PD from randomizations versus the residuals from a LOESS regression of observed PD on SD. Relative PD is the difference between observed PD and mean PD from randomizations of the same number of species.



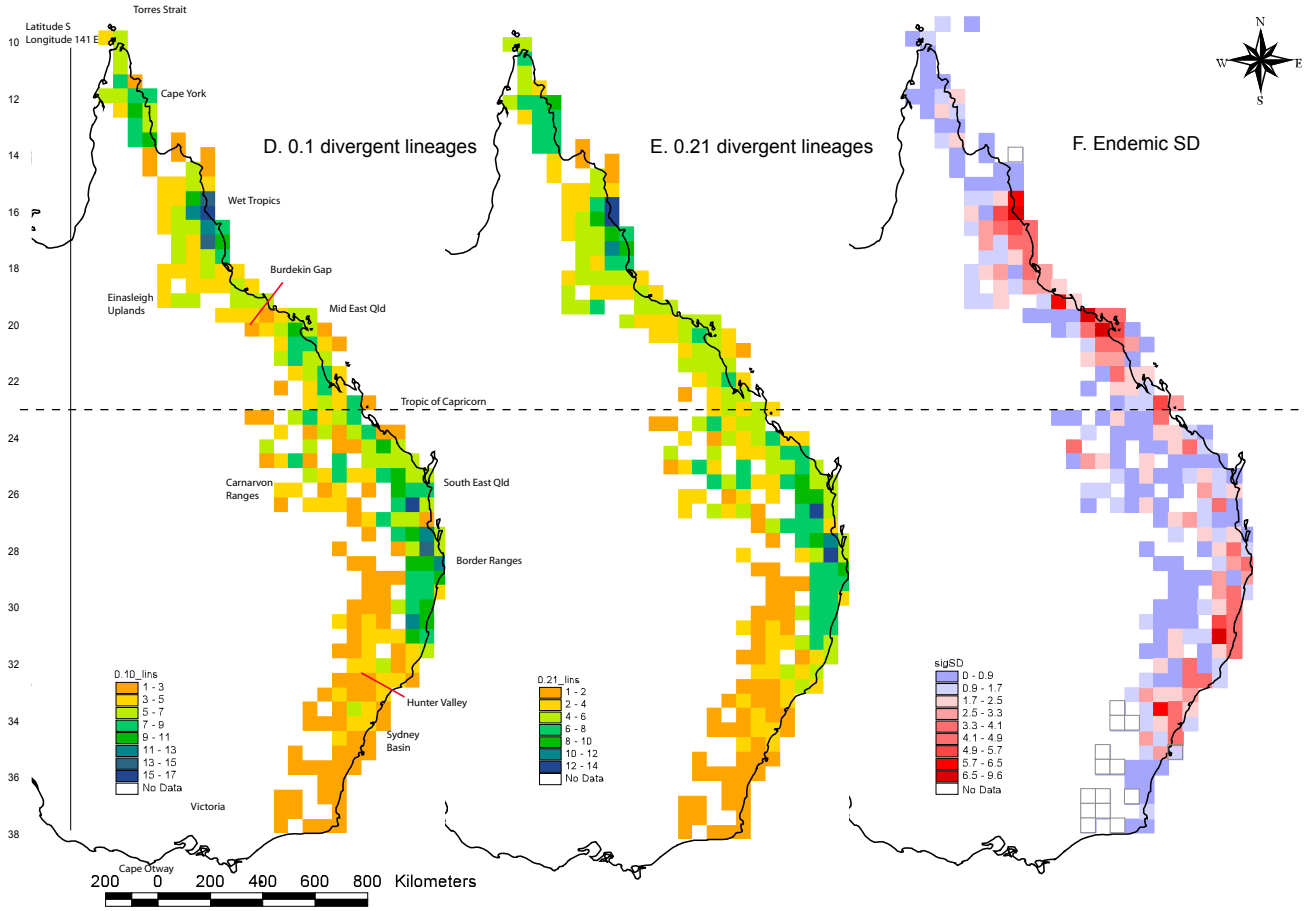


Figure 5. Various patterns of Species and Phylgenetic Diversities. **Top:** Distribution of SD and root PD at the 1/2 degree grid scale (for the set of included cells; see Chapter 1); right, relative PD is the difference between observed PD and expected (root) PD for that number of species (SD) randomly drawn from the full set in the phylogeny. Negative values indicate regions with relatively low PD. Details of the randomization statistics are in Figures 3 & 4. For these analyses SD is defined by the number of tips in the phylogeny assigned to a grid cell (SD=OTU). **Bottom:** Number of lineages that are 0.10 and 0.21 divergent; right, endemic species diversity using the sigmoidal function (a=0.4,b=1; see Chapter 1) There are about twice as many species as 0.1 lineages and about twice as many 0.1 as 0.21 lineages. These levels are used in the Divergence Rank Lineages (DRL) indices (see Figures 6 & 8).

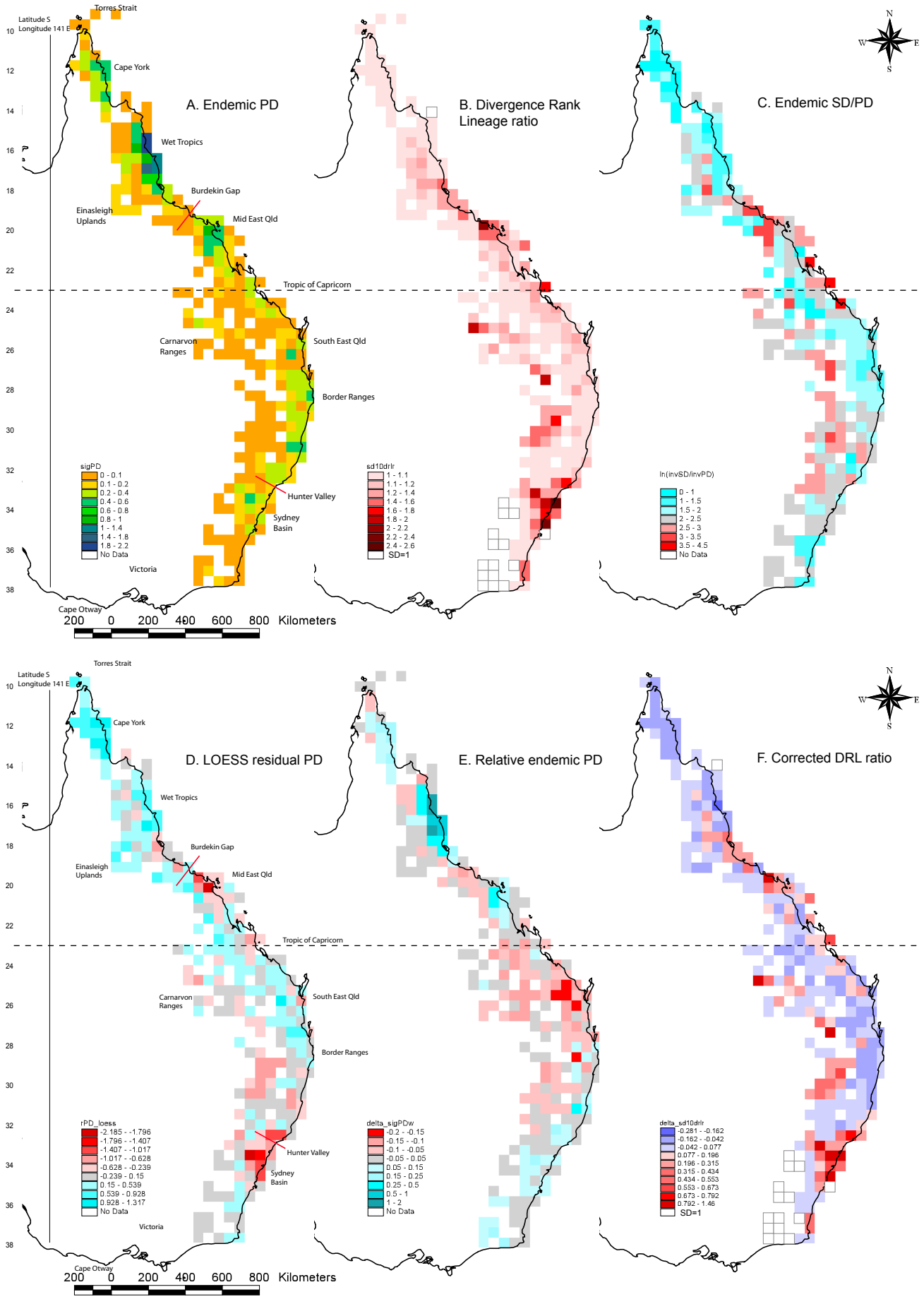


Figure 6. Various measures of diversification. **Top:** Endemic PD (using the sigmoid function); Divergence Rank Lineage (DLR) ratio is the ratio of the number of lineages between two different divergence levels, tips/0.10 (see also Figure 8); Endemic SD/PD is the natural log of the ratio of endemic SD to endemic PD (inverse function). **Bottom:** LOESS residual PD is the residuals from a LOESS (spline) regression (lambda=100) of (root) PD onto SD. Negative values indicate relatively low PD for the number of species. Note similarity to relative PD (Figure 5C). Relative endemic PD is the difference between observed endemic PD (sigmoidal function with a=0.4, b=1) and average endemic PD from randomly drawn sets of taxa (probability weighted by range). Corrected DRLratio is the difference between the observed and null Divergence Rank Lineage ratio (6B). Negative values indicate regions with relatively low values.

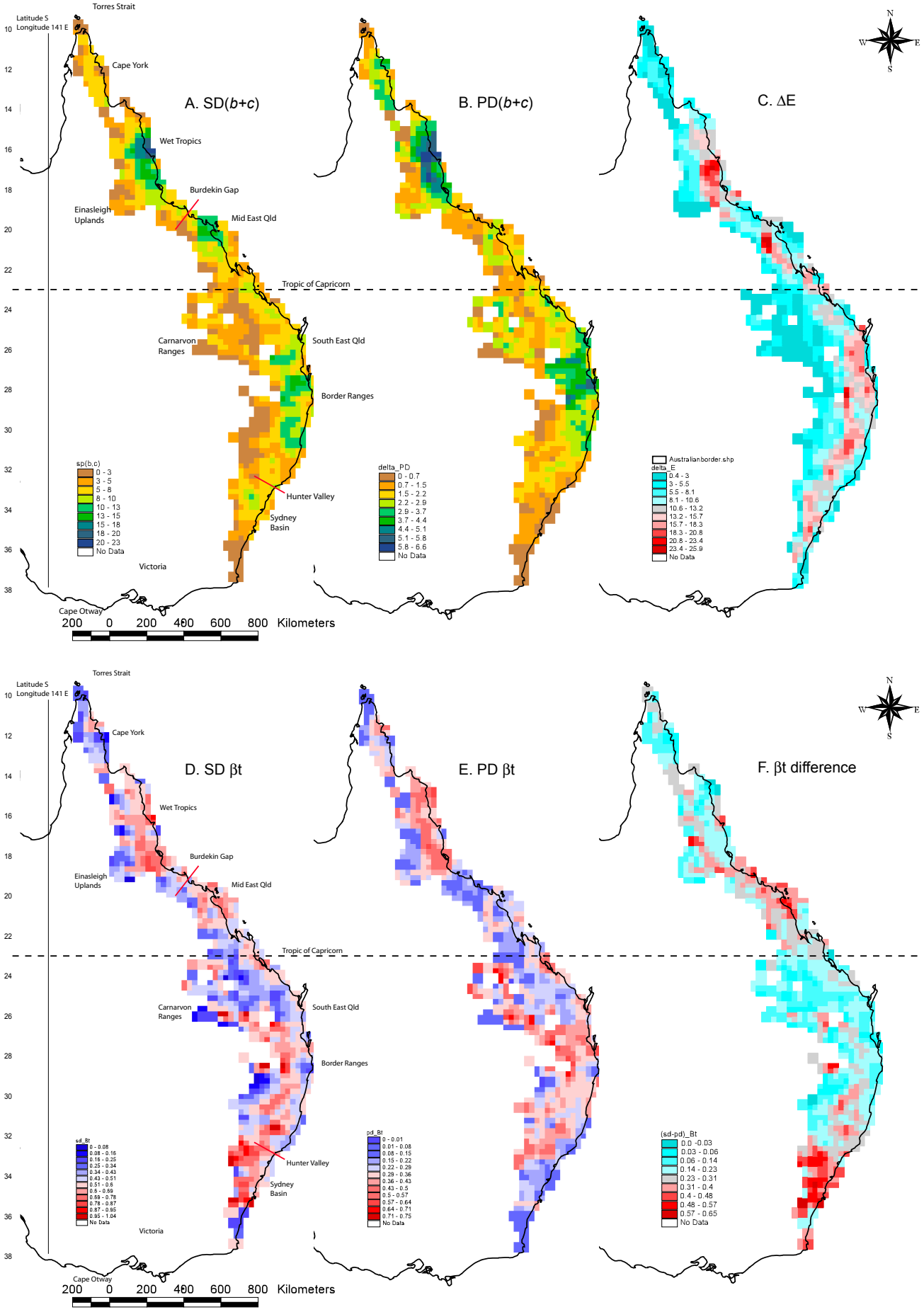


Figure 7. Turnover between adjacent grid cells. **Top**: Total turnover (*b*+*c*) in species diversity (SD) and phylogenetic diversity (PD), and environmental turnover (ΔE) as measured by the Euclidean distance among the three plant growth indices mesothermal, megathermal and microthermal (see Chapter 1 for details). **Bottom**: Relative turnover in species (SDβt) and phylogenetic diversities (PDβt), and the difference between the two. Turnover calculated between grid cells with SD>2 only.

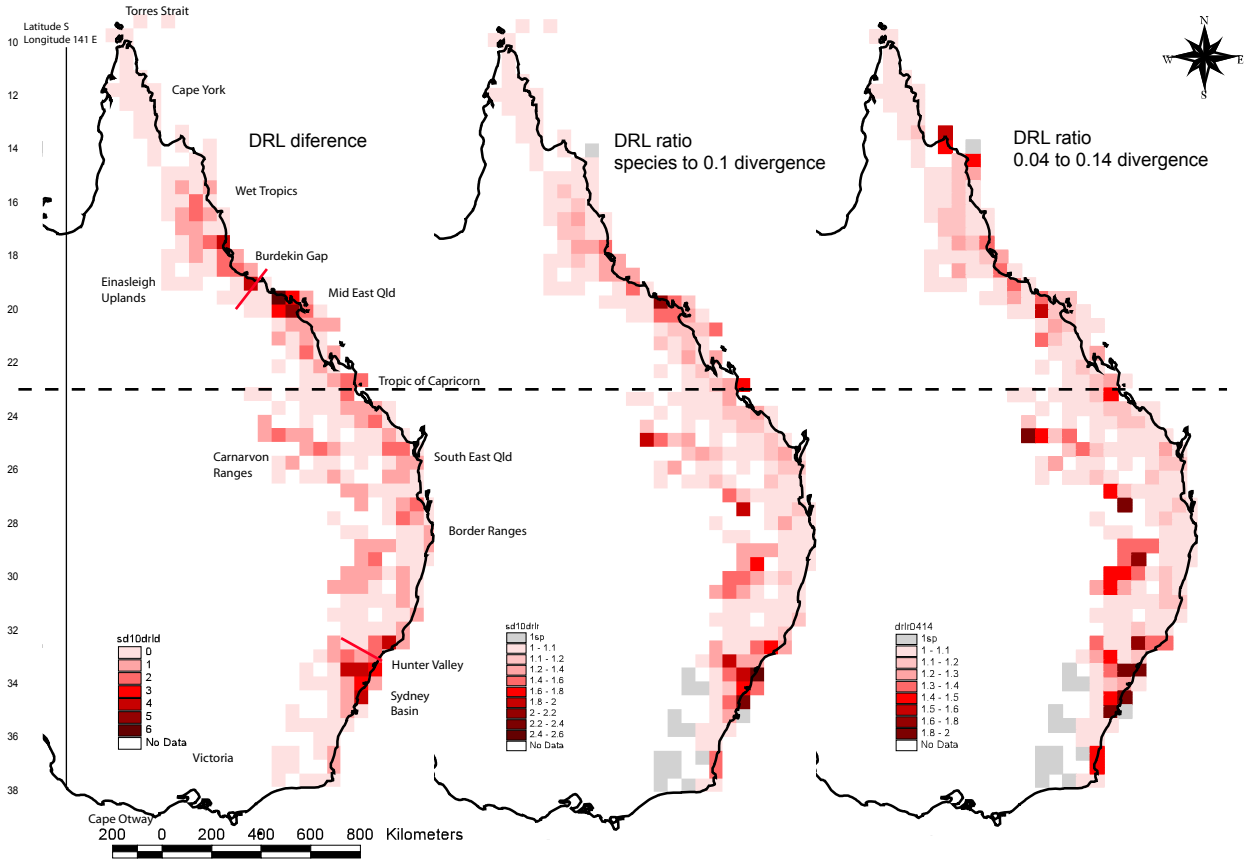


Figure 8. Left: DRL difference for species to 0.1 divergence lineages; Middle: DRL ratio for species to 0.1 divergence (see Figure 6); Right: DRL ratio for 0.04 to 0.14 divergence.

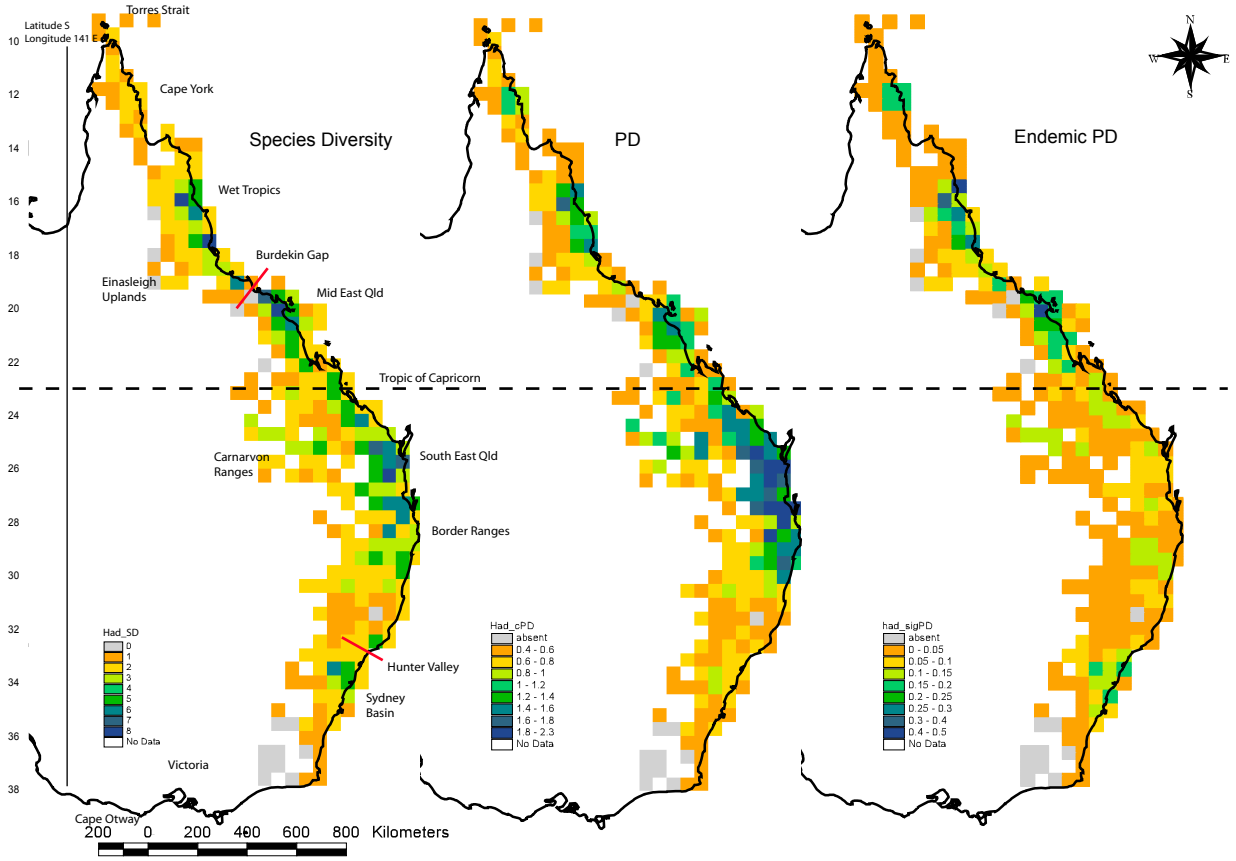
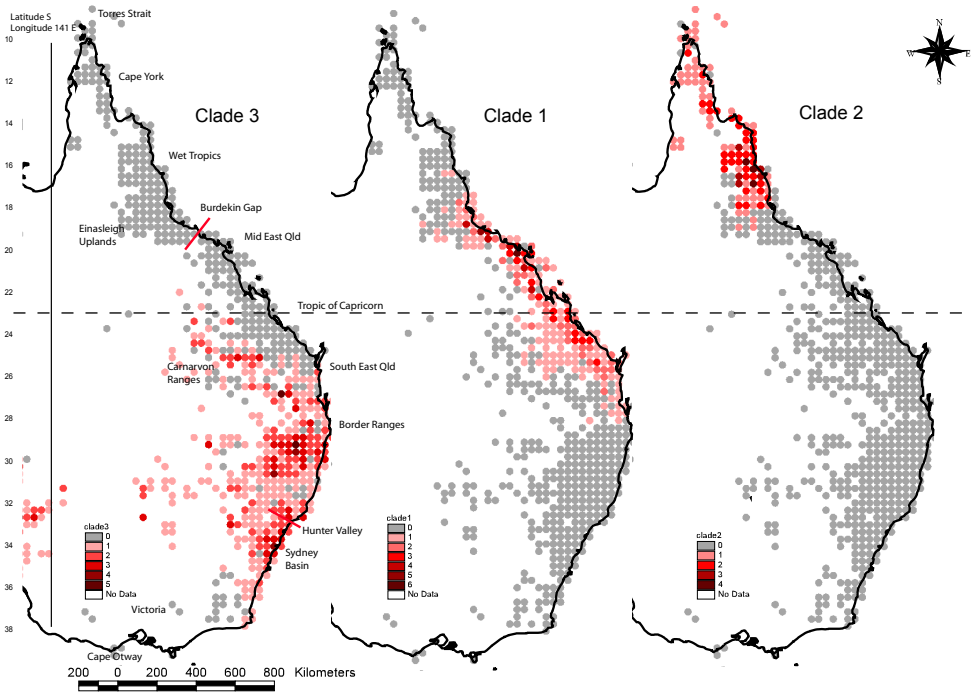
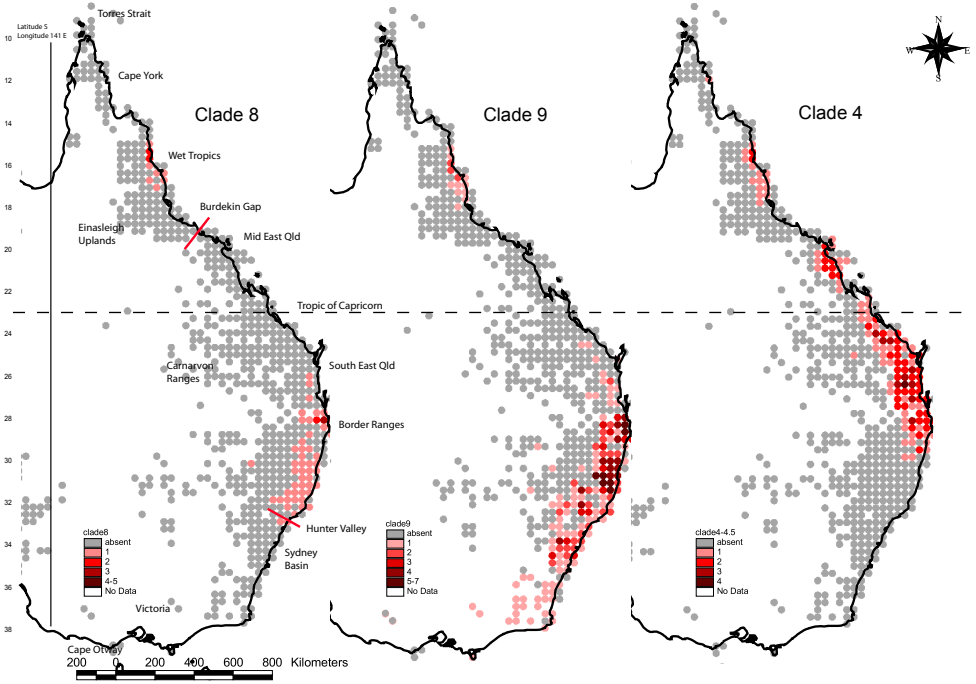


Figure 9. Species and phylogenetic diversity for the hadroid group. Endemic PD based on the sigmoidal function (a=0.4, b=1). Values are shown for the same 1/2 degree grids used for all east coast camaenids, even though hadroid distribution data is sufficient for more grids (and finer scale) to meet the adequate sampling criteria. Note that some regions can be recorded as having no hadroid species (e.g. Southern and New England highlands).





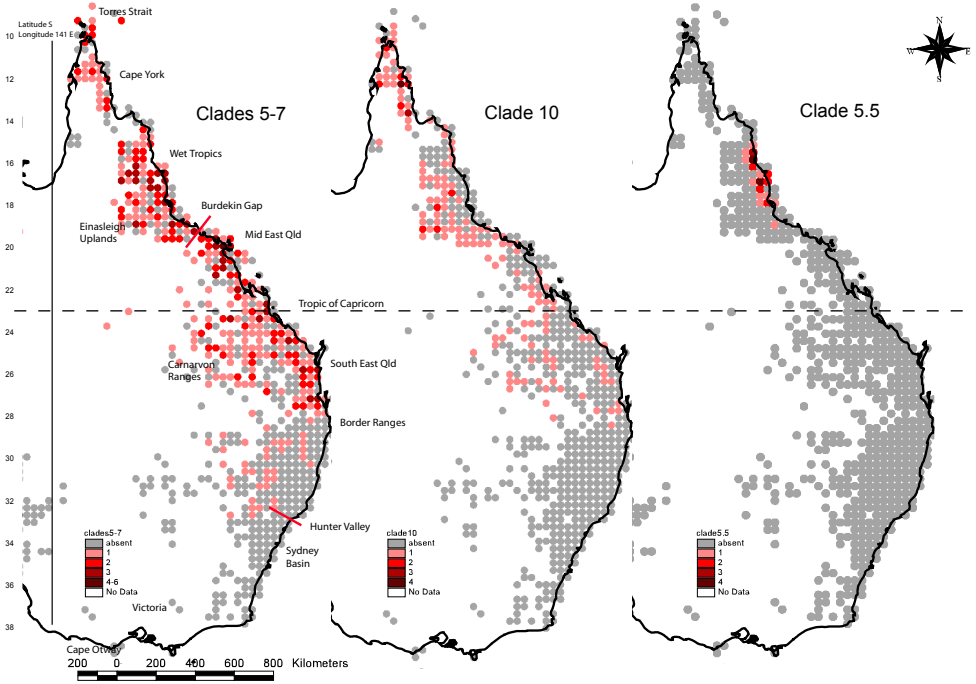


Figure 10. Distribution of some major clades contributing to phylogenetic diversity in eastern Australia. Based on the 327 taxa tree: see Chapter 2 for further details on definition and distribution of clades. Plotted using all records by circles of 1/3 degree diameter (c.f. grid cells with species-complete sampling). Clade 3 is distributed more widely across the Flinders, Lofty and Gawler Ranges in South Australia. Distribution of clades 5-7 south of 26 °S is due to the very widespread *Neveritis aridorum*. Distribution of clade 10 south of 26 °S is due to the very widespread *Trachiopsis mucosa*. Clade 8 also has a disjunct distribution in the Red Centre.

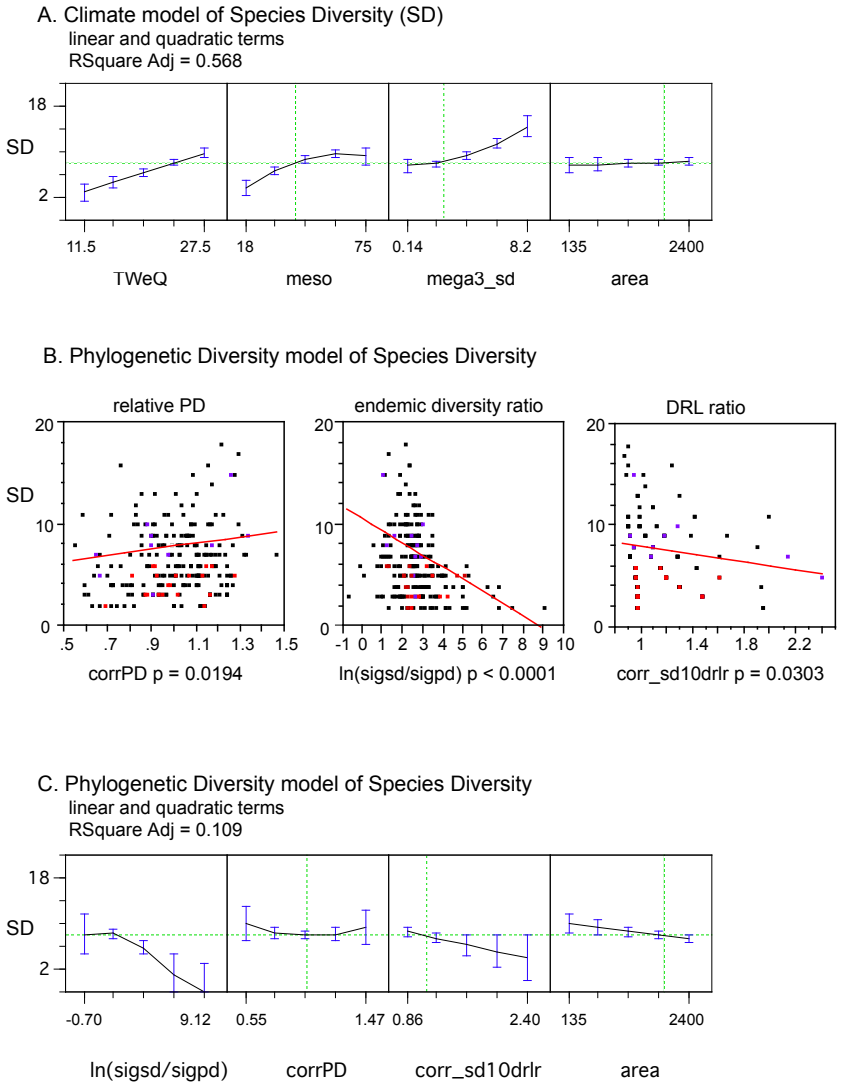


Figure 11. Climate and phylogenetic diversity modelling of species diversity. Top: prediction profile for climate model. Middle: linear correlations of three phylogenetic diversification indices. Bottom: prediction profile for PD model. Standard least squares using linear and quadratic terms, with observations weighted by the proportion of the inverse function endemic PD. All analyses included area and intercept. See text right and Table 1 for additional details.

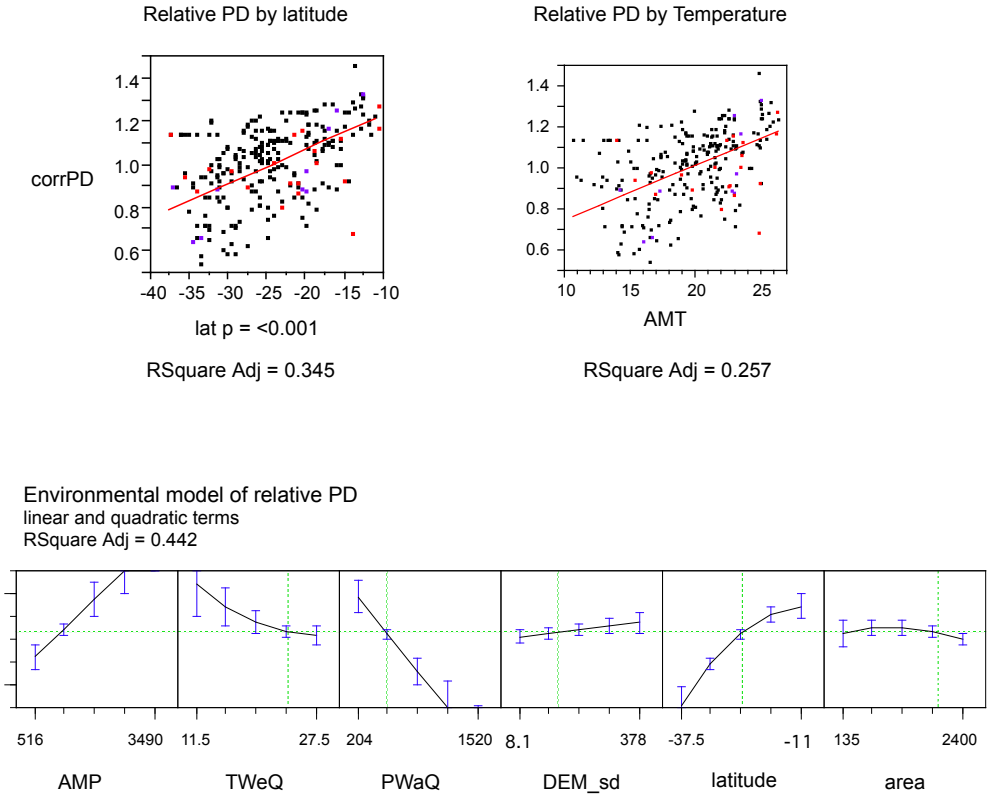


Figure 12 Environmental correlates of relative PD. Top: linear correlations to latitude and annual mean temperate. Bottom: prediction profile of environmental model. Standard least squares after stepwise selection of linear and quadratics terms for eight variables: AMT, AMP, TWeQ, PWaQ, DEM_sd, meso_sd, latitude, area. Details of methods as before.

Climate variables:
1) climate (meso) = mesothermal PGI
2) seasonality (TWeQ) = mean temperature of the wettest quarter
3) grid heterogeneity (mega_sd) = standard deviation of the megathermal PGI within the grid
4) AMT = annual mean temperatue
5) AMP = annual mean precipitation
6) PWaQ = precipitation of the warmest quater
7) DEM_sd = standard deviation in elevation within a grid

Phylogenetic diversification indices:
1) relative PD (corrPD) = root PD - null mean
2) endemic diversity ratio [ln(sigsd/sigpd)] = natural log using sigmoidal endemicity
3) corrected DRL ratio (corr_sd10drlr) = tip to 0.1 divergence - null mean

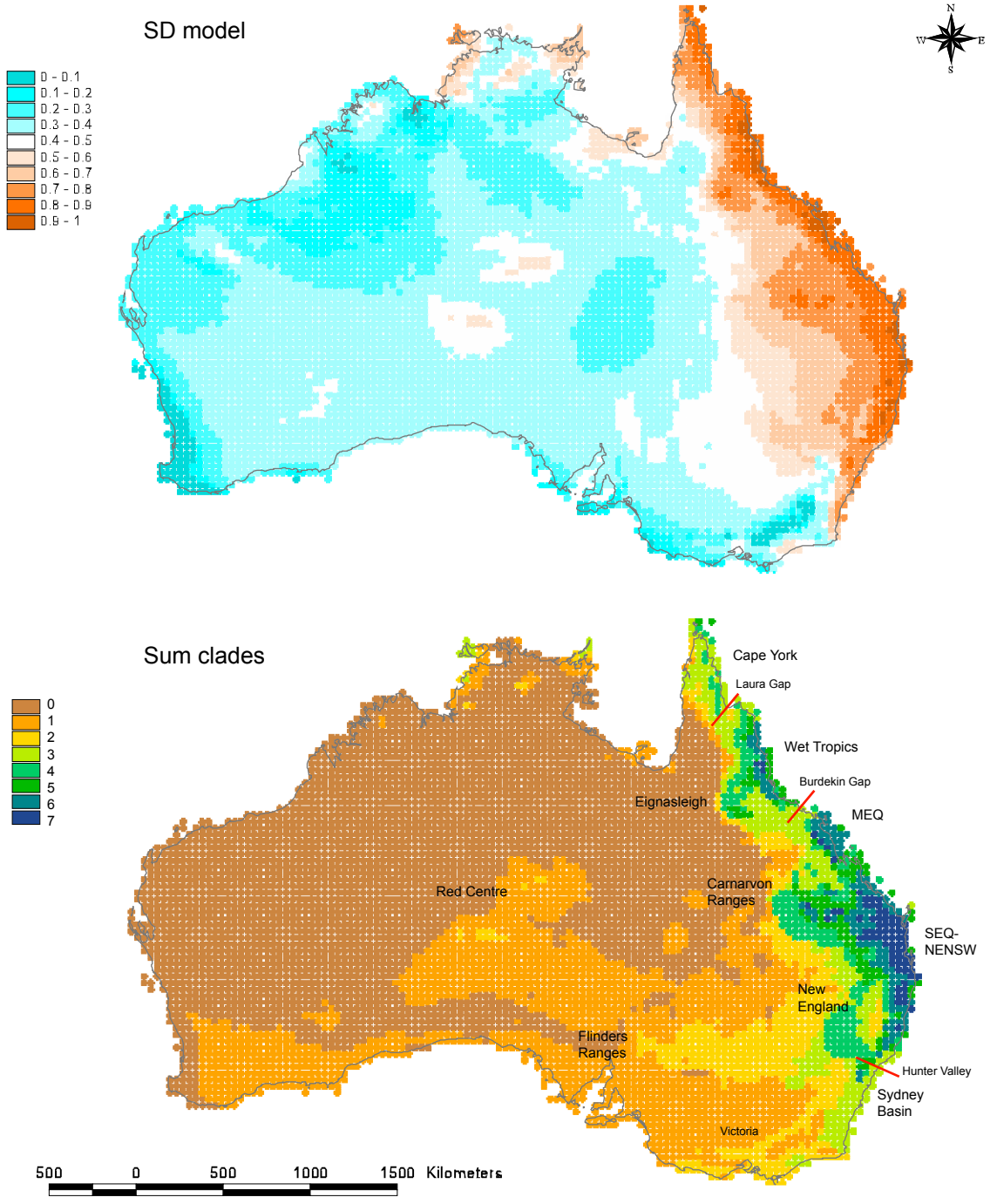


Figure 13. Two bioclimatic models of diversity. **Top**: general linear model of species diversity based on the east coast 50km grid cells with SD>1. The scale is relative probability of presence versus, therefore below 0.5 absence is more likely than presence. **Bottom**: Sum of logistic bioclimatic models for seven major groups. Models use four climate variables: annual mean temperature, mean annual precipitation, precipitation of the warmest quater, precipitation of the coldest quarter. The seven groups are: clades 1+2, clade 3, clade 4, clades 5+6+7, clade 8, clade 9, clade 10. In both case the models have been projected onto climate surfaces rescaled to $^1/_3$ degree averages.

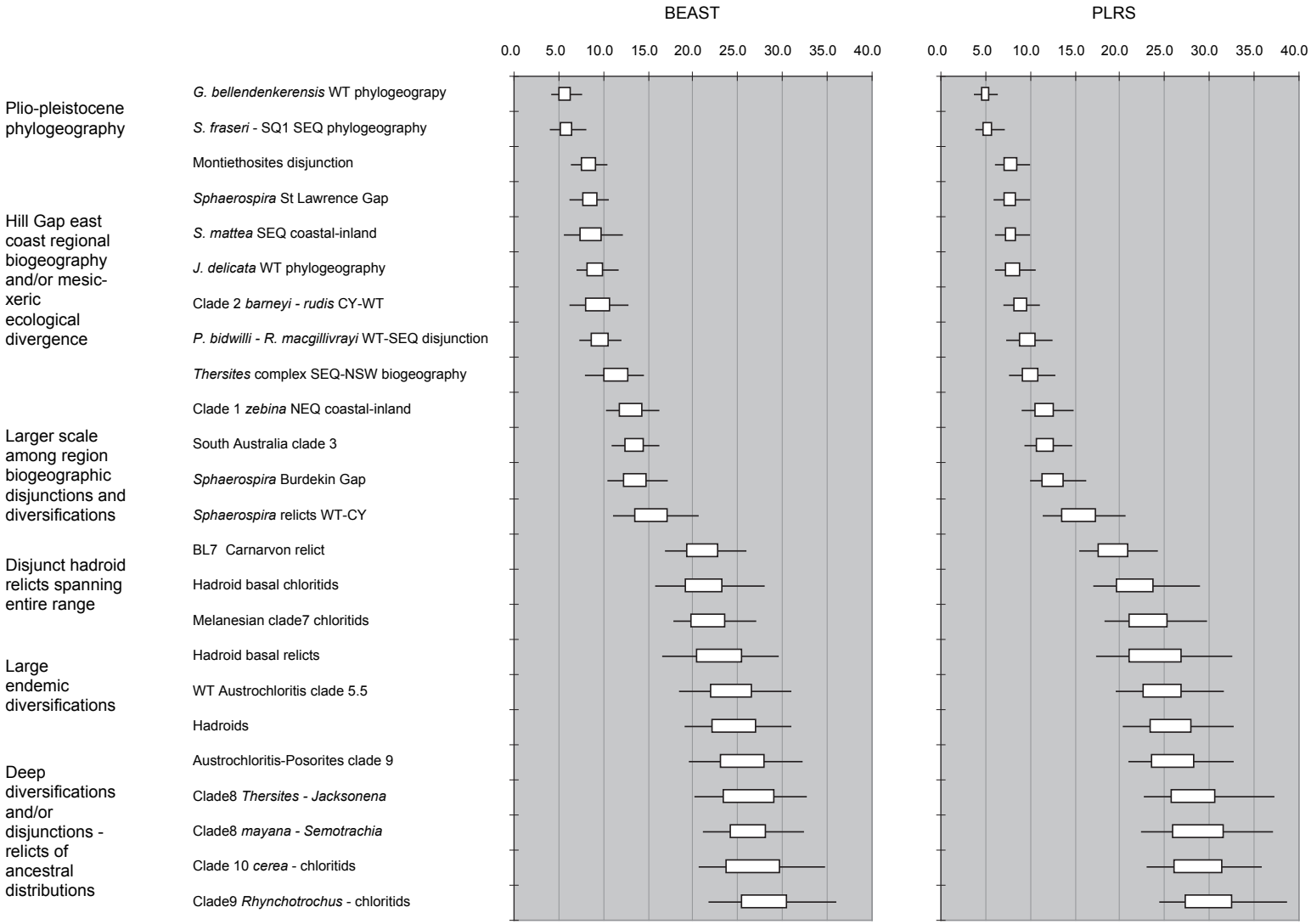


Figure 14. Relaxed-clock relative dating of selected camaenid divergences. The plots show quartile and 95% confidence intervals for two types of relaxed-clock methods. Both analyses use the 147 taxon dataset (see Chapter 2) where each taxon has at least two of three mtDNA gene (total maximum 1542 sites) with aGTR-G-inv model for each gene (=3 partitions). Both use the same calibration scheme: mean rate of 0.017 with stdev 0.002, combined with uniform constraint of 10-20 mya for the *Sphaerospira* lineage Burdekin Gap split (See Chapter 4). BEAST = bayesian analysis with BEAST v1.4.8; PLRS = penalized likelihood rate-smoothing using r8s v1.7.1; see text for details.

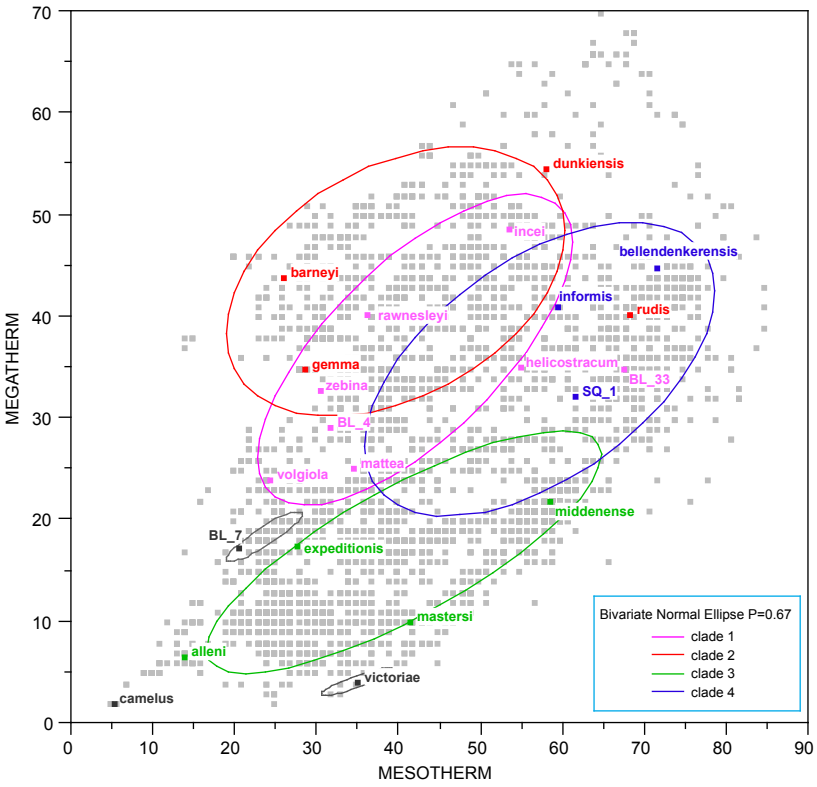


Figure 15. Bioclimatic distribution of hadroid clades and selected species. Grey points are all east-coast (east of longitude 142) hadroid distribution data points (n=4346), with corresponding two-thirds majority bivariate density ellipses for each clade. Select species points are species means. The axes are the Nix (1982) mesothermal and megathermal (c3) plant growth indices (see Chapter 1 for details).

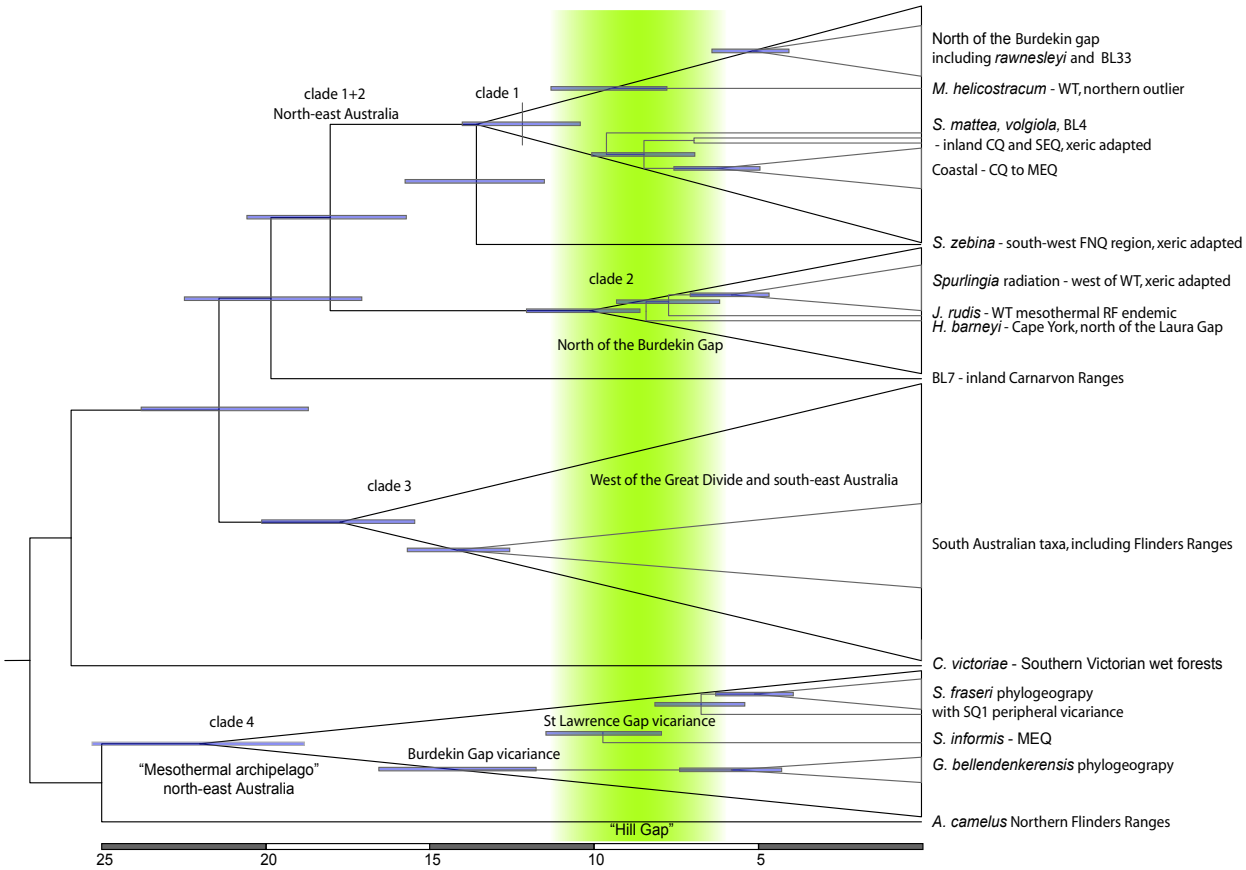


Figure 16. Maximum clade credibility tree from BEAST v1.4.8 relaxed-clock analysis of hadroid mtDNA. Methods and calibrations as for Figure 14 but using the all taxa dataset (hadroid group excluding *Vidumelon watti,* plus five outgroups drawn from the 327 taxa dataset; see Chapter 2). Tree shows the four main clades, collapsed proportional to the number of tips (usingFigTree v1.0; total of 163 ingroup tips), with selected sub-lineages super-imposed. See Chapter 2 for details on sampling across the hadroids, and topological support. Node bars indicate relative age 95% confidence intervals. Sampling for South Australian taxa only covers generic level diversity (11 of 32 species). Shaded interval denotes the mid-late Miocene era of change known as the "Hill Gap" (see Byrne et al. 2008).

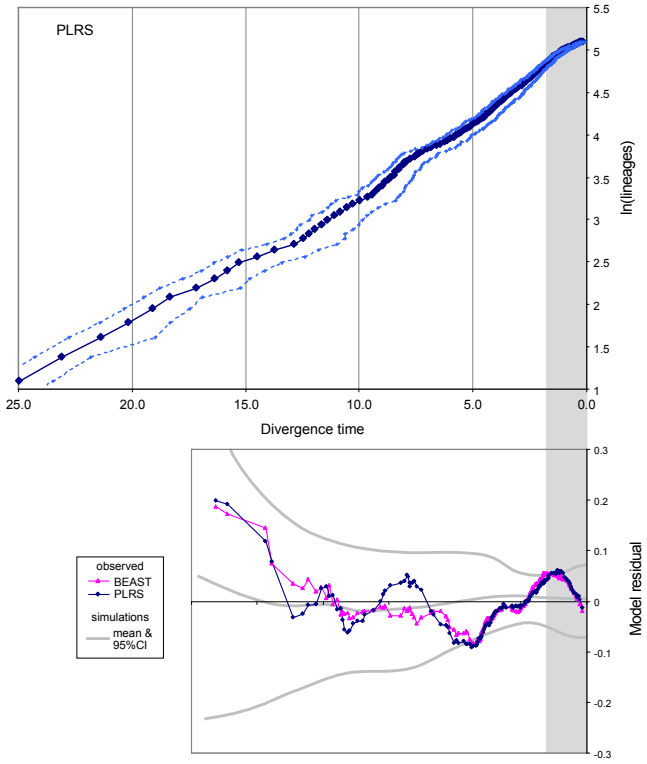


Figure 17. Hadroid relaxed-clock Lineage accumulation plots. **Top**: PLRS lineage-through-time plot. **Bottom**: Deviation from pure birth model. Mean and 95%CI of lineage accumulation (LTT) drawn from 100 samples of the two relaxed-clock analyses, PLRS and BEAST (see Appendix 5H). The bottom panel shows residuals of fit to a birth model (log-linear) excluding the basal 10 nodes: observed is the mean of 10 trees, compared to residual deviation seen across the simulations. Fitting was done in JMP v3.1.5 and 200 simulations done in Phylo-Gen v1.0. Mean and 95%CI lines spline fit smoothed for presentation. Figures 16 to 19 share a common divergence scale. The grey zone roughly indicates the lineages excluded in the diversification rates analyses (Table 3).

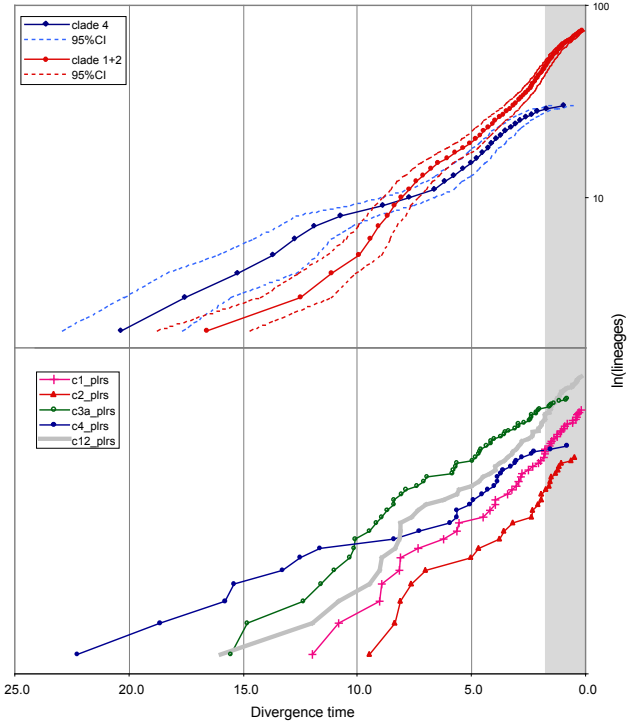
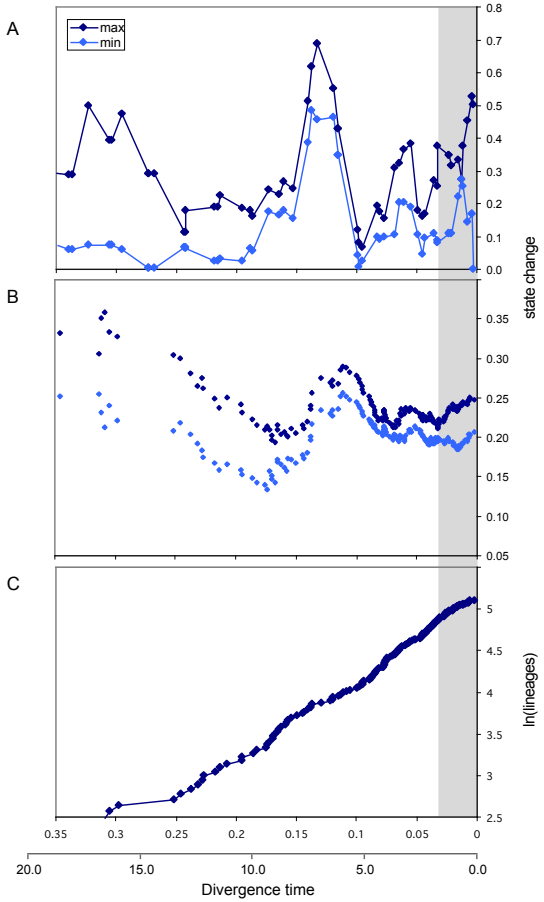


Figure 18. Lineage acccumulation plots for clades within the hadroids. Upper lineage plot shows mean and 95% CI of the BEAST analysis (from posterior 100 samples) for clade 4 and clade 1+2. Lower lineage plots are from the PLRS tree. Clade 3 is shown for interest but incomplete sampling of South Australian and some western NSW taxa limit comparison. Plots show only diversification up to the most recent split only (and hence stop before zero divergence). Clade 4 and clade 1+2 are broadly co-distributed across north-eastern Australia but tend to occupy different bioclimatic domains (Figure 15 and see Chapters 1 & 2).



D

categorization	xeric = 0	mesic = 1
meso_2C	92	72
M60.40.30_2C	85	79
AMP1000_2C	67	97
AMP1100and3m50	99	65
AMP1000and3mCI	102	62

Figure 19. Mesic-xeric bioclimatic category state changes across the hadroid PLRS phylogeny. Change is measured as the absolute value of the difference in proportional likelihood of state 0 between the parent and both daughter lineages. Five different categorizations of species using combinations of PGI and AMP giving slightly different numbers in each category (**D**; see also Appendix 5J). Tree node states inferred using symmetric Mk1 and assymetric Mk2 models in Mesquite v2.6. **A**: average change per node in divergence bins; **B**: cumulative average change per node from the base of the tree to the tips. Both plots show the minimum and maximum values across the five categorizations by two stochastic models. **C**: hadroid PLRS lineage accumulation plot with the same timescale as previous figures.

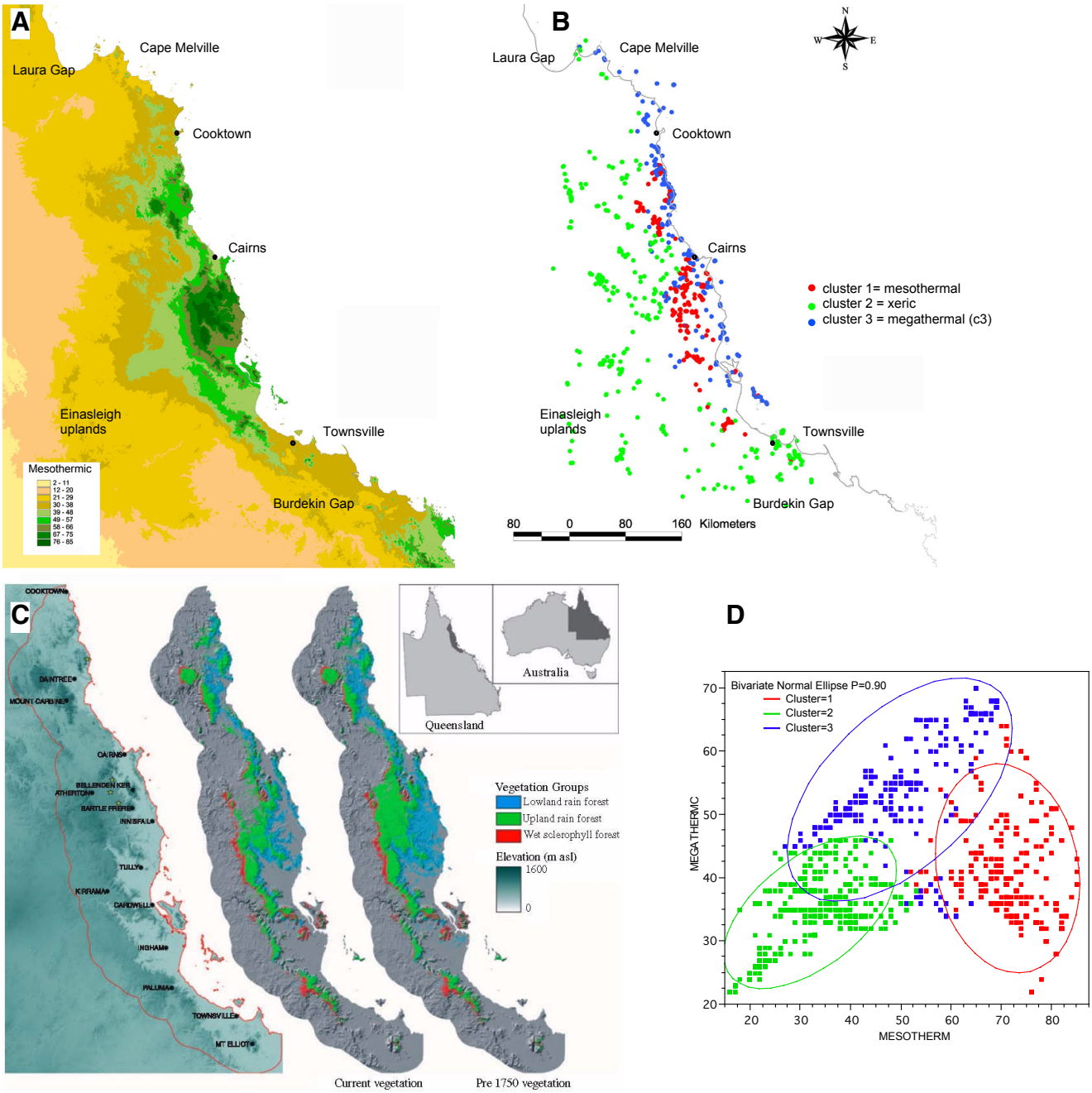


Figure 1 Elevational profile and distributional differences of vegetation for current and pre-clear time periods for lowland and upland rain forest, and wet sclerophyll forest classifications of the Australia Wet Tropics region used in this study. The area delineated in red is the analysis region. The stars on the left map represent major palaeoecological sites including, from north to south, Thornton Range escarpment, Lake Euramoo, Lake Barrine and Lynch's Crater.

Figure 20. Far North Queensland bioregion (incorporating the Wet Tropics) showing camaenid snail distribution and bioclimatic classification into three categories. **A:** Map of the Mesothermal PGI for the FNQ region, bounded by the Laura and Burdekin Gaps, and west to the Einasleigh uplands; **B:** Camaenid snail distribution points, classified into three types based on the Nix (1982) plant growth indices; **C:** Figure 2 from VanDerWal et al. (2008) showing Wet Tropics region and extent of upland (mesothermal) and lowland (megathermal) rainforests; **D:** distribution of mesothermal and megathemal PGI values for the snail distribution points and three categories.

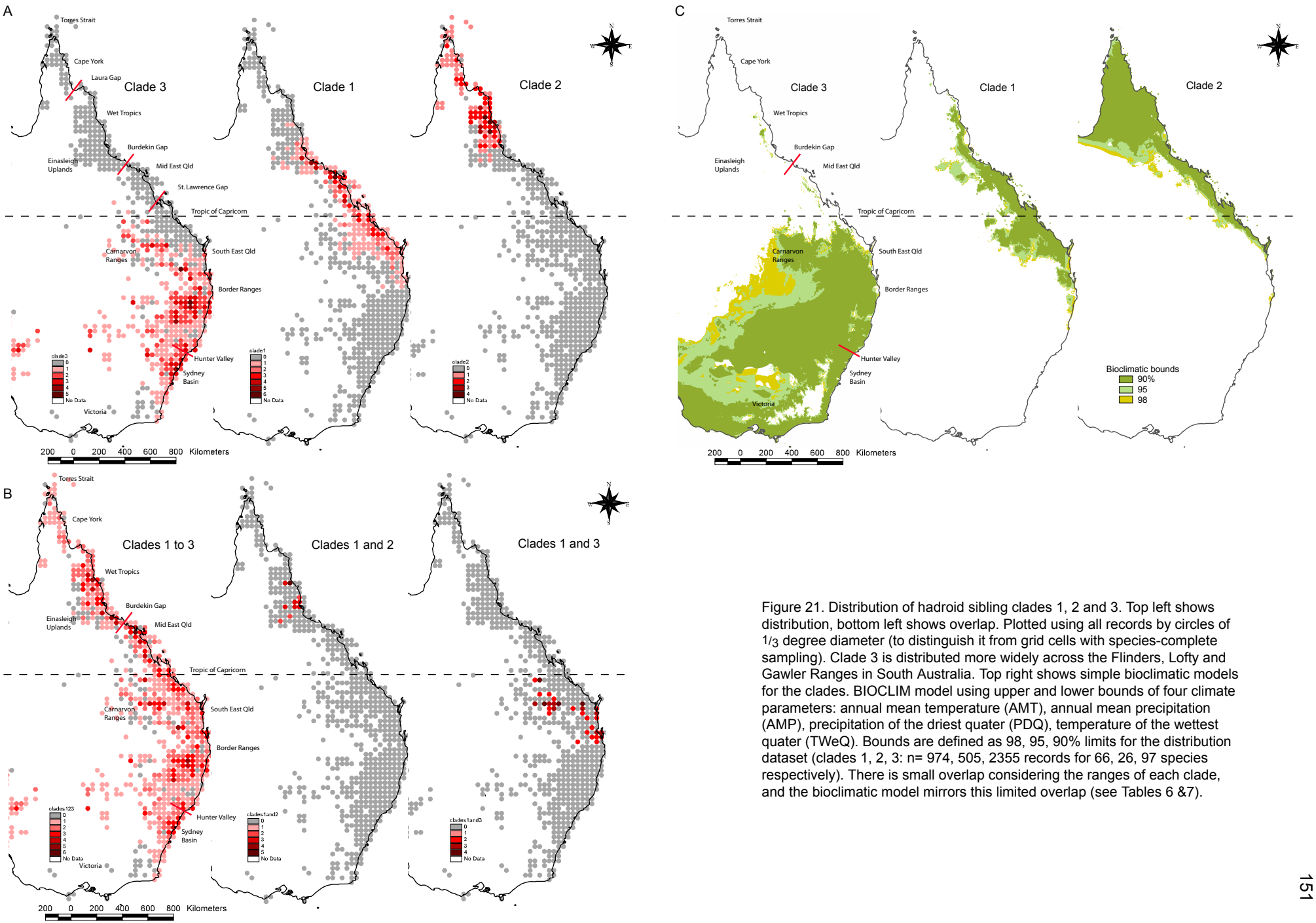


Figure 21. Distribution of hadroid sibling clades 1, 2 and 3. Top left shows distribution, bottom left shows overlap. Plotted using all records by circles of 1/3 degree diameter (to distinguish it from grid cells with species-complete sampling). Clade 3 is distributed more widely across the Flinders, Lofty and Gawler Ranges in South Australia. Top right shows simple bioclimatic models for the clades. BIOCLIM model using upper and lower bounds of four climate parameters: annual mean temperature (AMT), annual mean precipitation (AMP), precipitation of the driest quater (PDQ), temperature of the wettest quater (TWeQ). Bounds are defined as 98, 95, 90% limits for the distribution dataset (clades 1, 2, 3: n= 974, 505, 2355 records for 66, 26, 97 species respectively). There is small overlap considering the ranges of each clade, and the bioclimatic model mirrors this limited overlap (see Tables 6 &7).

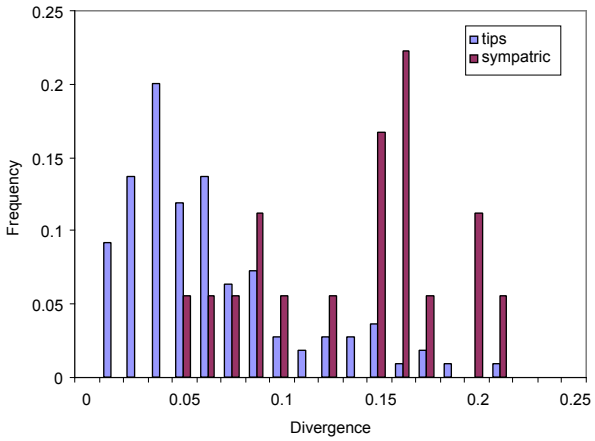


Figure 22. Histogram of sister species (tips) and sympatric lineage divergences for east coast hadroids. Divergences from PLRS Bayesian consensus in substitutions per site.

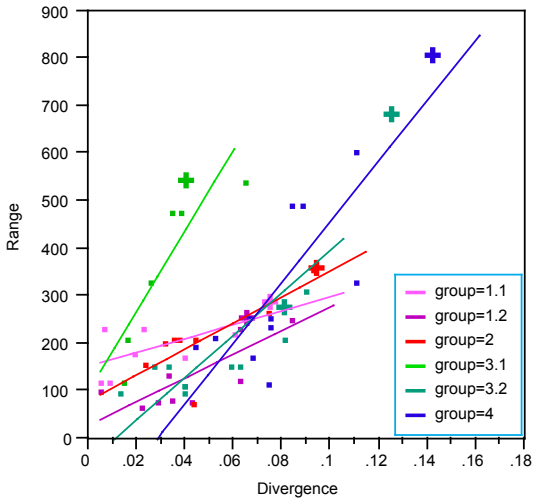


Figure 24. Geographic range by divergence for allopatric and sympatric subclades subtended by the eight most closely related hadroid sympatric species. Geographic range is measured as maxumim linear range in km. Divergence is from the PLRS Bayesian phylogeny. Large crosses indicate the most closely related nodes that subtend each of the symparic species pairs (see Table 8). A linear regression is fitted for each clade as a visual aid but the relationship between divergence and range is not expected to be so simple.

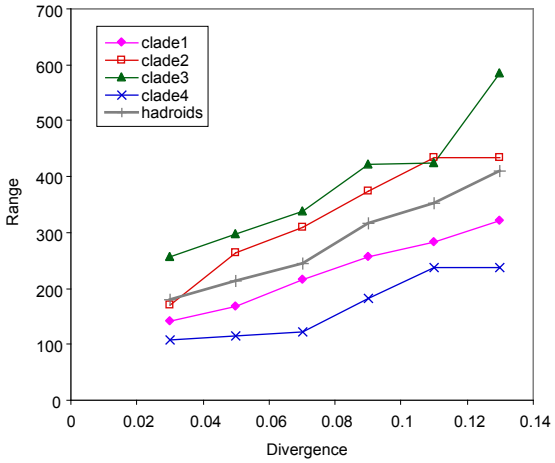


Figure 23. Geographic range by phylogenetic divergence for hadroid clades. Geographic range is measured as maxumim linear range. Averages of lineages at given divergence depths in the PLRS phylogeny.General trend in km per 1% divergence:

clade 1	18
clade2	27
clade3	30
clade4	15
hadroids	23

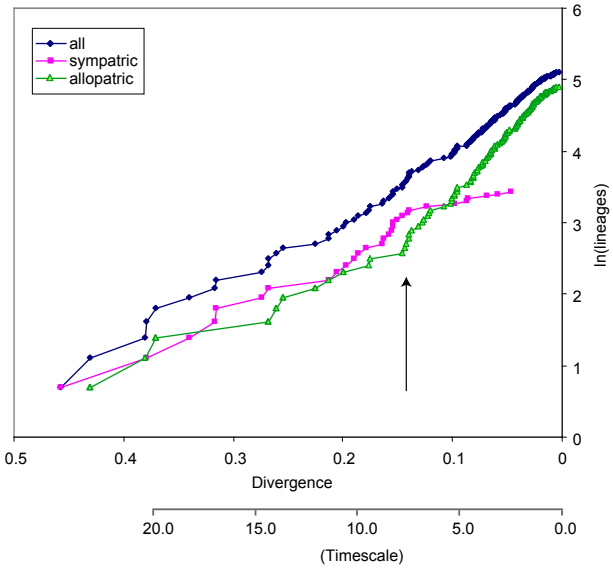


Figure 25. Cumulative plot of lineage by divergence for hadroids.The tree can be partitioned into the nodes defining daughter lineages than contain all instances of sympatric diversity and the remainder, which are allopatric. PLRS Bayesian consensus tree (Appendix 5M). Plot of all nodes is equivalent to those in Figures 17 & 19, with matching approximate timescale.

References for Chapters 1 & 5

Adam P. 1992. *Australian Rainforests,* Clarendon Press, Oxford.

AFD website: Australian Fauna Directory website: (http://www.environment.gov.au/biodiversity/abrs/online-resources/fauna/afd/taxa/CAMAENIDAE).

Arajuo MB & New M. 2006. Ensemble forecasting of species distributions. TREE 22:42-47.

Araujo MB, David Nogues-Bravo D, Diniz-Filho DAF, Haywood AM, Valdes PJ & Rahbek C. 2008. Quaternary climate changes explain diversity among reptiles and amphibians. *Ecography* 31:8-15.

Archer M, Hand SJ & Godthelp H. 1991. *Riversleigh. The story of animals in Ancient Rainforest of Inland Australia.* Reed Books Balgowlah NSW.

Avise JC, Walker D & Johns GC. 1998. Speciation durations and pleistocene effects on vertebrate phylogeography. *Phil. Trans. R. Soc. Lond. B* 265:1707-1712.

Avise JC. 2008. Phylogeography: retrospect and prospect. *J. Biogeoraph.* 36:3-15.

Balke M, Ribera I, Hendrich L, Miller MA, Sagata K, Posman A, Vogler AP & Meier R. 2009. New Guinea highland origin of a widespread arthropod supertramp. *Proc. R. Soc. Lond. B* 276:2359-2367.

Barker GM. 2002. Phylogenetic diversity: a quantitative framework for measurement of priority and achievement in biodiversity conservation. *Biol. J. Linn. Soc.* 76:165-194.

Barraclough TG, Vogler AP & Harvey PH. 1998. Revealing the factors that promote speciation. *Phil. Trans. R. Soc. Lond.* B 353:241-249.

Barraclough T G & Nee S. 2001. Phylogenetics and speciation. *Trends Ecol. Evol*. 16:391–399.

Baselga A. 2010. Partitioning the turnover and nestedness components of beta diversity. *Global Ecol. Biogeogr.* 19:134-143.

Beaumont MA & Rannala B. 2004. The Bayesian revolution in genetics. *Nat. Rev. Genet.* 5:251-261.

Bickford SA, Laffan SW, de Kok RPJ & Orthia LA. 2004. Spatial analysis of taxonomic and genetic patterns and their potential for understanding evolutionary histories. *J. Biogeography* 31:1715–1733

Blackburn T.M. & Gaston KJ. 2004. Macroecology: Concepts and Consequences: 43rd Symposium of the British Ecological Society. Cambridge University Press.

Bowler JM. 1982. Aridity in the late Tertiary and Quaternary of Australia. In: Barker WR, Greenslade PJM, eds. *Evolution of the flora and fauna of arid* Australia. Norwood, South Australia: Peacock Publications, 35-45.

Brooks DR. & McLennan DA. 1991. *Phylogeny, ecology, and behaviour. A research program in comparative biology.* The University of Chicago Press, Chicago.

Brown JH. 1995. *Macroecology.* The University of Chicago Press, Chicago.

Burbidge NT. 1960. The phytogeography of the Australian region. *Aust. J. Botany* 8:75-210.

Butlin RK, Bridle JR & Schluter D. 2009. Speciation and patterns of biodiversity. In: R Butlin, J. Bridle and D. Schluter, eds. *Speciation and patterns of diversity*. Cambridge University Press, Cambridge, pp 1-14.

Byrne M, Yeates DK, Joseoh L, Kearney M, Bowler J, Williams MAJ, Cooper S, Donnellan SC, Keogh S, Leys R, Melville J, Murphy DJ, Porch N & Wyrwoll K-H.2008. Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. *Mol. Ecol.* 17:4398-4417.

Cain AJ. 1954. *Animal species and their evolution.* Hutchinson House, London, W1.

Cardillo M. 1999. Latitude and rates of diversification in birds and butterflies. *Proc. Roy. Soc. Lond. B* 266:1221–1225.

Cardillo M, Orme CDL & Owens IPF. 2005. Testing for latitudinal bias in diversification rates: an example using New World birds. *Ecology* 86:2278–2287.

Cavender-Bares J, Kozak KH, Fine PVA & Kembel SW 2009. The merging of community ecology and phylogenetic biology. *Ecol. Lett*. 12:693-715.

Clark SA. 2005. Systematics, spatial analysis and conservation genetics of *Meridolum corneovirens* (Pfeiffer, 1851) and related forms (Gastropoda: Camaenidae) from the Sydney region of Australia. PhD thesis, University of Western Sydney.

Clark SA. 2009. A review of the land snail genus *Meridolum* (Gastropoda: Camaenidae) from central New South Wales, Australia. *Molluscan Research* 29:61–120.

Colwell RK & Lees DC. 2000. The mid-domain effect: geometric constraints on the geography of species richness. *Trends Ecol. Evol.* 15:70-76.

Condit R, Pitman N, Leigh Jr EG, Chave J, Terborgh J, Foster RB, Nunez V P, Aguilar S, Valencia R, Villa G, Muller-Landau HC, Losos E & Hubbell SP. 2002. Beta-diversity in tropical forest trees. *Science* 295:666-669.

Cracraft J. 1986. Origin and evolution of continental biotas: speciation and historical congruence within the Australian avifauna. *Evolution*: 40:977-996.

Crisp MD, Linder HP & Weston PH. 1995 Cladistic biogeography of plants in Australia and New Guinea: congruent pattern reveals two endemic tropical tracks. *Syst. Biol.* 44:457–473.

Crisp MD, Laffan S, Linder HP,& Monro A. 2001. Endemism in the Australian Flora. *J. Biogeoraphy.* 28:183-198.

Crisp M. Cook L & Steane, D. 2004. Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present-day communities? *Phil. Trans. Roy. Soc. Lond. B* 359;1551-1571.

Crisp MD & Cook LG. 2009. Explosive radiation or mass extinction? Interpreting signatures in molecular phylogenies. *Evolution* 63:2257–2265.

Crisp MD, Arroyo MTK, Cook LG, Gandolfo MA, Jordan GJ, McGlone MS, Weston PH, Westoby M, Wilf P & Linder HP. 2009. Phylogenetic biome conservatism on a global scale. *Nature* 458:754-756.

Darwin C. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.

Davies RG, Orme CD, Webster AJ, Jones KE, Blackburne TM & Gaston KJ. 2007. Environmental predictors of global parrot (Aves: Psittaciformes) species richness and phylogenetic diversity. *Global Ecol. Biogeog*. 16:220-233.

Davies TJ, Fritz SA, Grenyer R, Orme CDL, Bielby J, Bininda-Emonds ORP, Cardillo M, Jones KE, Gittleman JL, Mace GM & Purvis A. 2008. Phylogenetic trees and the future of mammalian biodiversity. *Proc. Natl. Acad. Sci. USA.* 105:11556-11563.

Dolman G & Hugall AF. 2008. Combined mitochondrial and nuclear data enhance resolution of a rapid radiation of Australian rainbow skinks (Scincidae: *Carlia*). Mol. Phylo. Evol. 49:782-794.

Dony JG. & Denholm I. 1985. Some quanttiative methods for assessing the conservation value of ecologically similar sites. *J. Applied Ecol.* 22, 229-238.

Drummond AJ, Ho SYW, Phillips MJ & Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 4:699–710.

Drummond AJ & Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214.

Egan AN & Crandall KA. 2008. Divergence and diversification in North American Psoraleeae (Fabaceae) due to climate change. *BMC Biology* 6:55.

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE, 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129–151.

Emberton KC, Kuncio GS, Davis GM, Phillips SM, Monderewicz KM & Guo YH. 1990. Comparison of recent classifications of Stylommatophoran land-snail Families, and evaluation of large-ribosomal-RNA sequencing for their phylogenetics. *Malacologia* 31:327-352.

Faith DP. 1994. Phylogenetic pattern and the quantification of organismal biodiversity. *Phil. Trans. Roy. Soc. Lond. B* 345:45-58.

Faith DP. 2002. Quantifying biodiversity: a phylogenetic perspective. *Conservation Biology* 16:248-252.

Faith DP, Reid CAM & Hunter J. 2004. Integrating Phylogenetic Diversity, Complementarity, and Endemism. *Conservation Biology* 18:255-261.

Felsenstein J. 2004. Inferring Phylogenies. Sinauer, Sunderland MA.

Férussac, A.E.J.P.J.F. d'A. de 1823. *Histoire Naturelle Générale et Particulière des Mollusques Terrestres et Fluviatiles.* (1819–1851). Paris : J.-B. Baillière, Libraire de l'Academie Naturale de Médecine 19-21 pp. plus plates. [27 Sept. 1823]

FitzJohn RG, Maddison WP, Otto SP. 2009. Estimating traitdependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* 58:595–611.

Fujioka T, Chappell J, Honda M, Yatsevich I, Fifield K & Fabel D. 2005. Global cooling initiated stony deserts in central Australia 2-4 Ma, dated by cosmogenic ^{21}Ne-^{10}Be. *Geology* 33:993-96.

Gaston KJ. 1994. Measuring geographic range sizes. *Ecography* 17:198-205.

Gaston KJ. 2000. Global patterns in biodiversity. *Nature* 405:220-227.

Gaston KJ & Fuller RA. 2009.The sizes of species' geographic ranges. *J. Applied Ecol.* 46:1-9.

Goldberg EE, Roy K, Lande RA & Jablonski D. 2005. Diversity, endemism, and age distribution in macroevolutionary sources and sinks. *Am. Nat.* 165: 623–633.

Graham CH & Hijmans RJ. 2006. A comparison of methods for mapping species ranges and species richness. *Global Ecol. Biogeogr.* 15:578–587.

Graham CH, Moritz C & Williams SE. 2006. Habitat history improves prediction of biodiversity in rainforest fauna. *Proc. Natl. Acad. Sci. USA* 103: 632-636.

Graham CH & Fine PVA. 2008. Phylogenetic beta diversity: linking ecological and evolutionary processes across space and time. *Ecol. Lett.* 11:1265-1277.

Graham CH, Parra JL, Rahbek C. & McGuire JA. 2009. Phylogenetic structure in tropical hummingbird communities. *Proc. Natl. Acad. Sci. USA* 106:19673-19678.

Greenwood DR & Christophel DC. 2005. The origins and Tertiary history of Australian 'tropical' rainforests. In: Bermingham E, Moritz C, Dick CW eds. *Tropical rainforests past, present and future*. University of Chicago Press, Chicago, pp 336-373.

Guisan A & Zimmermann NE. 200. Predictive habitat distribution models in ecology. Ecol. Modelling 135:147-186.

Hardy OJ & Sentierre B. 2007. Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. *J. Ecology* 95:493-506.

Hardy OJ. 2008. Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. *J. Ecology* 96:914–926.

Harmon LJ, Schulte JA, Larson A. & Losos JB. 2003. Tempo and mode of evolutionary radiation in iguanian lizards. *Science* 301:961-964.

Hartmann K & Steel M. 2007. Phylogenetic diversity: from combinatorics to ecology. In: O. Gascuel, M Steel eds. *Reconstructing Evolution, New mathematical and computational advances*. Oxford University Press, pp 171-192.

Harvey PH, May RM & Nee S. (1994). Phylogenies without fossils. *Evolution* 48:523-529.

Heard SB & Cox GH. The shapes of phylogenetic trees of clades, faunas, and local assemblages: exploring spatial pattern in differential diversification. *Am. Nat.* 169:E000.

Hey J. 1992. Using phylogenetic trees to study speciation and extinction.*Evolution* 46:627-640.

Hey J. 2009. On the arbitrary identification of real species. In: R Butlin, J. Bridle and D. Schluter, editors. *Speciation and patterns of diversity*. Cambridge University Press, Cambridge, pp 15-28.

Hill RS. 1994. The history of selected Australian taxa. In: Hill RS ed. *History of Australian Vegetation: Cretaceous to Recent*. Cambridge University Press, Cambridge, UK, pp. 390–419.

Hocknull SA, Zhao J-x, Feng Y-x & Webb G.E. 2007. Responses of Quaternary rainforest vertebrates to climate change in Australia, *Earth Planetary Sci. Lett.* 264:317–331.

Houlder D, Hutchinson M, Nix H & McMahon, J 2000. ANUCLIM User's Guide v.5.1. Centre for Resource and Environmental Studies, Australian National University, Canberra.

Hu XS, He FL & Hubbell SP. 2009. Community differentiation on landscapes: drift, migration and speciation. *Oikos* 118:1515-1523.

Hubbell SP. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography.* Princeton University Press, Princeton.

Hugall AF, Foster R., Hutchinson M & Lee MSY. 2008. Phylogeny of Australasian agamid lizards based on nuclear and mitochondrial genes: implications for morphological evolution and biogeography. *Biol. J. Linn. Soc.* 93:343–358.

IBRA 6.1 website: http://www.environment.gov.au/parks/nrs/science/ibra.html

Iredale T. 1937. A basic list of the land Mollusca of Australia. *Aust. J. Zoology* 8:287-333.

Isaac NJB, Mallet J & Mace GM. 2004. Taxonomic inflation: its influence on macroecology and conservation. *TREE* 19:464–469.

Jansson R & Davies TJ. 2008. Global variation in diversification rates of flowering plants: energy vs. climate change. *Ecology Letters* 11: 173–183

Johnson CN. 2006 *Australia's mammal extinctions: a 50,000 year history*. Cambridge, UK: Cambridge University Press.

Joseph L, Moritz C & Hugall A. 1993. A mitochondrial perspective on the historical biogeography of mideastern Queensland rainforest birds. *Memoirs Qld Museum* 34:201-214.

Kershaw AP. 1994. Pleistocene vegetation of the himid tropics of northeastern Queensland, Australia. *Palaeogeog. Palaeoclim. Palaeoecol.* 109:399-412.

Kershaw AP, Moss PT & Wild R. 2005. patterns and causes of vegetation change in the Australian Wet tropics region over the last 10 million years. In: Bermingham E, Moritz C, Dick CW eds. *Tropical rainforests past, present and future*. University of Chicago Press, Chicago, pp 374-400.

Kikkawa J & Pearse K. 1969. Geographical distribution of land birds in Australia: A numerical analysis. *Aust. J. Zoology* 17:821-840.

Kikkawa J, Monteith GB & Ingram G. 1981. Cape York Peninsula: Major region of faunal interchange pp 1695-1742 in *Ecological biogeography of Australia* (J. A. Keast ed.) W. Junk, The Hague.

Koleff P, Lennon JJ & Gaston KJ. 2003. Are there laitudinal gradients in species turnover? *Global Ecol. Biogeog*. 12:483-498.

Kubo T. & Iwasa Y. 1995. Inferring the rates of branching and extinction from molecular phylogenies. *Evolution* 49:694-704.

Laffan SW. & Crisp MD. 2003 Assessing endemism at multiple spatial scales, with an example from the Australian vascular flora. *J. Biogeography* 30:511–520.

Legendre P. 1993. Spatial autocorrelation: Trouble or new paradigm? Ecology 74:1659-1673.

Legendre P. & Legendre L. 1998. *Numerical Ecology*, 2nd English edn. Elsevier, Amsterdam.

Legendre P, Borcard D & Peres-Neto PR 2005. Analyzing β diversity: Partitioning the spatial variation of community composition data. *Ecol. Monogr*. 75:435–450.

Lemmon AR, Brown JM, Stanger-Hall K, Lemmon EM. 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Systematic Biology* 58: 130–145.

Lewis LA & Lewis PO. 2005. Unearthing the molecular phylodiversity of desert soil green algae (Chlorophyta). *Syst. Biol.* 54:936-947.

Loader SP, Pisani D, Cotton JA, Gower DJ, Day JJ & Wilkinson M. 2007. Relative time scales reveal multiple origins of parallel disjunct distributions of African caecilian amphibians. *Biol. Lett.* 3:505–508.

Lobo JM, Jimenez-Valverde A & Real R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* 17:145–151.

Lomolino MV, Riddle BR & Brown JH. 2006. *Biogeography*, 3rd edn. Sinauer Associates, Inc., Sunderland, MA.

Losos JB. &. Schluter D. 2000. Analysis of an evolutionary species-area relationship. *Nature* 408:847-850

Low T. 1993. Last of the rainforests: rainforest refuges of the Mackay region. *Wildlife Australia* 4:18-21.

Lynch JD. 1989. The gauge of speciation. In *Speciation and its consequences* (D. Otte & J.A. Endler eds). Sinauer Associates, Inc., Sunderland Massachusetts.

MacArthur RH, & Wilson EO. 1967. *The theory of island biogeography*. Princeton University Press, Princeton, New Jersey, USA.

Maddison WP & Maddison DR. 2009. Mesquite: A modular system for evolutionary analysis. Version 2.6. http://mesquiteproject.org

Maddison WP, Midford PE & Otto SP 2007. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* 56:701–710.

Magallon S & Sanderson MJ. 2001. Absolute diversification rates in angiosperm clades. *Evolution* 55:1762-1780.

Malay MCD & Paulay G. 2010. Peripatric speciation drives diversification and distributional pattern of reef hermit crabs (Decapoda: Diogenidae: *Calcinus*). Evolution 64:634–662.

Mallet J. 1995. A species definition for the modern synthesis. Trends Ecol. Evol. 10:294–299.

Martin HA. 1994. Australian Tertiary phyogeography: evidence from palynology. In: Hill RS ed. *History of Australian Vegetation: Cretaceous to Recent*. Cambridge University Press, Cambridge, UK, pp. 104-142.

Martin HA. 2006. Cenozoic climatic changes and the development of the arid vegetation of Australia. *J Arid Environments*, 66:533–563.

Mayr E. 1976. *Evolution and the diversity of life: selected essays.* Pp 479-528. Belknap Press Cambridge, Massachusetts.

Mayr E. 1997. *This is biology. The science of the living world.* Pp 64-78. Belknap Press Cambridge, Massachusetts.

Mayr E & Diamond J. 2001. *The birds of northern Melanesia: speciation, ecology and biogeography*. Oxford University Press, Oxford, UK, and New York, New York, USA.

Meredith, RW, Westerman M & Springer, MS. 2009. A phylogeny of Diprotodontia (Marsupialia) based on sequences for five nuclear genes. *Mol. Phylo. Evol.* 51:554–571.

McGowran B, Holdgate GR, Li Q & Gallagher SJ. 2004. Cenozoic stratigraphic succession in southeastern Australia. *Aust. J.Earth Sci.* 51:459-496.

McPeek, M. A. 2008. The ecological dynamics of clade diversification and community assembly. *Am. Nat*. 172:E270–E284.

Monaghan MT, Wild R, Elliot M, Fujisawa T, Balke M, Inward DJ, Lees DC, Ranaivosolo R, Eggleton P, Barraclough TG & Vogler AP. 2009. Accelerated Species Inventory on Madagascar Using Coalescent-Based Models of Species Delineation. *Syst. Biol.* 58:298-311.

Moritz C. Joseph L, Cunningham M & Schneider C. 1997. Molecular perspectives on historical fragmentation of Australian tropical and subtropical rainforest: implications for conservation. In: *Tropical Rainforest Remnants: Ecology, Management, and Conservation of Fragmented Communities* (ed. W.F. Laurance and R.O. Bierregard). Chicago Univ. Press, New York, pp 442-454.

Moritz C, Patton JL, Schneider CJ & Smith TB. 2000. Diversification of rainforest faunas: An integrated molecular approach. *Ann. Rev. Ecol. Syst.* 31:533-563.

Moritz C, Hoskin C, Graham CH, Hugall A & Moussalli A. 2005. Historical biogeography, diversity and conservation of Australia's tropical rainforest herpetofauna. In: Gittleman JL, Brooks TM, Purvis A. eds. *Phylogeny and Conservation.* Zoological Society of London, Cambridge University Press, pp 243-266.

Moussalli A, Hugall, AF & Moritz C. 2005. A mitochondrial phylogeny of the rainforest skink genus *Saproscincus*, Wells and Wellington (1984). *Mol. Phylo. Evol*. 34:190–202.

Moussalli A, Moritz C, Williams SE & Carnaval AC. 2009. Variable responses of skinks to a common history of rainforest fluctuation: concordance between phylogeography and palaeo-distribution models. *Mol. Ecol.* 18:483-499.

Nabholz B, Glemin S & Galtier N. 2009. The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals BMC Evol. Bio. 9:54

Nee S, Holmes EC, May RM & Harvey PH. 1994a. Extinction rates can be estimated from molecular phylogenies. *Phil. Trans. R. Soc. Lond. B* 344:77-82.

Nee S. 2001. Inferring speciation rates from phylogenies. *Evolution* 55:661-668.

Nelson G. & Platnick N. 1981. *Systematics and Biogeography: Cladistics and Vicariance*. Columbia University Press, New York.

Nix HA. 1982. Environmental determinates of biogeography and evolution in Terra Australis. In: Barker WR & Greenslade PJM eds *Evolution of the Flora and Fauna of Arid Australia*). Peacock Publications, Frewville, South Australia, pp 47-66.

Nix H. 1991. Biogeography: pattern and process. In: Nix HA, Switzer MA eds. Rainforest animals: atlas of vertebrates endemic to Australia's wet tropics. Australian National Parks and Wildlife, Canberra, pp. 11–40.

Norman JA, Christidis L, Joseph L, Slikas B & Alpers D. 2002. Unravelling a biogeographical knot: origin of the 'leapfrog' distribution pattern of Australo-Papuan sooty owls (Strigiformes) and logrunners (Passeriformes). *Proc. Roy. Soc. Lond. B* 269:2127–2133.

Nott J. 2005. The origin and evolution of Australia's eastern highlands. In: Bermingham E, Moritz C, Dick CW eds. *Tropical rainforests past, present and future*. University of Chicago Press, Chicago, pp 322-335.

Phillips SJ, Anderson RP & Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190:231–259.

Pulquerio MJF & Nichols RA. 2007. Dates from the molecular clock: how wrong can we be? *TREE* 22:180-184.

Purvis A, Orme CDL, Toomey NH & Pearson PN. 2009. Temporal patterns in diversification. In: R Butlin, J. Bridle and D. Schluter, eds. *Speciation and patterns of diversity*. Cambridge University Press, Cambridge, pp 278-300.

Pybus OG, Harvey PH. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc. R. Soc. Biol. Lond. B* 267:2267–2272.

Qian H, Badgley C & Fox DL. 2009. The latitudinal gradient of beta diversity in relation to climate and topography for mammals in North America. *Global Ecol. Biogeogr*. 18:111–122.

Quek SP, Davies SJ, Itino T. & Pierce NE. 2004. Codiversification of an ant-plant mutualism: the phylogeny of host use in Crematogaster (Formicidae) associates of Macaranga (Euphorbiaceae). *Evolution* 58:554-570.

Rabosky D. 2006a. Likelihood methods for temporal detecting shifts in diversification rates. *Evolution* 60:1152-1164.

Rabosky DL. 2006b. LASER: a maximum likelihood toolkit for inferring temporal shifts in diversification rates. Evol. *Bioinformatics* 2:257–260.

Rabosky DL & Lovette IJ. 2008. Explosive evolutionary radiations:decreasing speciation or increasing extinction through time? *Evolution* 62:1866–1875.

Rabosky DL. 2009a. Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. *Ecol. Lett.* 12:735–743.

Rabosky DL. 2009b. Ecological limits on clade diversification in higher taxa. *Am. Nat.* 173:662–674.

Raboksy DL. 2009c. Heritability of extinction rates links diversification patterns in molecular phylogenies. *Syst. Biol.* 58:629-640.

Rabosky D.L. 2010. Extinction rates should not be estimated from molecular phylogenies. *Evolution* 64:1816–1824.

Rabosky DL & Glor RE. 2010. Equilibrium speciation dynamics in a model adaptive radiation of island lizards. *Proc. Nat. Acad. Sci. USA* 107:22178–22183.

Rahbek C. & Graves GR. 2001. Multiscale assessment of patterns of avian species richness. *Proc. Nat. Acad. Sci. USA* 98:4534–4539.

Rambaut A & Charleston M. 2000. Phylogenetic tree editor v1.0a4-61. University of Oxford, Oxford, UK.

Rambaut A. 2002. PhyloGen: phylogenetic tree simulator package version 1.1. Department of Zoology, University of Oxford, Oxford, UK.

Ribera I, Fresneda J, Bucur R, Izquierdo A, Vogler AP, Salgado JM & Cieslak A. 2010. Ancient origin of a Western Mediterranean radiation of subterranean beetles. *BMC Evol. Biol*. in press.

Ricklefs RE. 2006. Evolutionary diversification and the origin of the diversity–environment relationship. *Ecology* 87:S3–S13.

Ricklefs RE. 2007a. History and diversity: explorations at the intersection of ecology and evolution. *Am. Nat.* 170:S56–S70.

Ricklefs R. 2007b. Estimating diversification rates from phylogenetic information. TREE 22:601-610.

Ricklefs RE. 2008. Disintegration of the Ecological Community. *Am. Nat.* 172:741–750.

Ricklefs RE. 2009. Speciation, extinction and diversity. In: R Butlin, J. Bridle and D. Schluter, eds. *Speciation and patterns of diversity*. Cambridge University Press, Cambridge, pp 257-277.

Ricklefs RE, Bermingham E. 2001. Nonequilibrium diversity dynamics of the Lesser Antillean avifauna. *Science* 294:1522-1524.

Ricklefs RE & Schluter D. 1993. Species diversity: regional and historical influences. In *Species diversity in ecological communities:historical and geographical perspectives* (ed. R.E. Rickefs & D. Schluter). University of Chicago Press, Chicago.

Rosauer D, Laffan SW, Crisp MD, Donnellan SC & Cook LG. 2009. Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Mol. Ecol.* 18:4061-4072.

Rosenzweig ML. 1995. *Species diversity in space and time*. Cambridge University Press.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times a penalized likelihood approach. Mol. Biol. Evol. 19:101–109.

Sanderson MJ. 2003. r8s; inferring absolute rates of evolution and divergence times in the absence of a molecular clock. Bioinformatics 19, 301–302.

Sanderson MJ, McMahon MM, Steel M. 2010. Phylogenomics with incomplete taxon coverage: the limits to inference. *BMC Evol. Biol.* 10, 155.

Schemske D. 2009. Biotic interactions and speciation in the tropics. Pages 219–239 *in* R Butlin, J. Bridle and D. Schluter, editors *Speciation and patterns of diversity*. Cambridge University Press, Cambridge.

Sechrest W, Brooks TM, da Fonseca GAB, Konstant WR, Mittermeier RA, Purvis A, Rylands AB, & Gittleman JL. 2002. Hotspots and the conservation of evolutionary history. *Proc. Natl. Acad. Sci. USA* 99:2067–2071.

Sinnott RW. 1984. Virtues of the Haversine, *Sky and Telescope* 68:159.

Slatyer C, Rosauer D & Lemckert F. 2007. An assessment of endemism and species richness patterns in the Australian Anura. *J. Biogeography* 34:583–596.

Smith BJ. 1992. Non-Marine Mollusca. *In* Houston, W.W.K. (ed.) *Zoological Catalogue of Australia*. Canberra: AGPS Vol. 8 xii 405 pp.

Smith TB, Wayne RK. Girman DJ & Bruford MW. 1997. A role for ecotones in generating rainforest biodiversity. *Science* 276:1855-1857.

Soininen J, McDonald R & Hillebrand H. 2007. The distance decay of similarity in ecological communities. *Ecography* 30:3-12

Solem A. 1979a. Land-snail biogeography: A true snail's pace of change. *In* Nelson, G. and Rosen, D.E. eds. *Vicariance biogeography a critique*. Columbia University Press New York: pp. 197-221.

Solem A. 1979b. Camaenid land snails from Western and Central Australia (Mollusca: Pulmonata: Camaenidae) I Taxa with trans-Australian distribution. Records of the Western Australian Museum Supplement No. 10:1-142.

Solem A. 1992. Camaenid land snails from southern and eastern South Australia, excluding Kangaroo Island. Records of the South Australia Museum, Monograph Series 2:1-425.

Solem A. 1997. Camaenid land snails from Western and Central Australia (Mollusca: Pulmonata: Camaenidae) VII Taxa from Dampierland through the Nullabor. Records of the Western Australian Museum Supplement No. 50:1461-1899.

Spencer, W. B. 1896. Summary. In Spencer WB ed. *Report on the Work of the Horn Scientific Expedition to Central Australia Part 1*. Dulau, London, pp 137-199

Stanisic J. 1994. The distribution and patterns of species diversity of land snails in eastern Australia. *Memoirs of the Queensland Museum* 36:207-213.

Stanisic J, Shea M, Potter D & Griffiths O. *Australia land snails: Volume 1 a field guide to eastern Australian species.* Queensland Museum Press, in prep.

Swofford DL. 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4 (beta). Computer program. Sunderland MA: Sinauer Associates.

Thomson RC, Shaffer HB. 2010. Sparse Supermatrices for Phylogenetic Inference: Taxonomy, Alignment, Rogue Taxa, and the Phylogeny of Living Turtles. *Syst. Biol.* 59: 42-58.

Tofts R & Silvertown J. 1999. A phylogenetic approach to community assembly from a local species pool. *Proc. Roy. Soc. Lond. B* 267:363-369.

Truswell E. 1993. Vegetation changes in the Australian Tertiary in response to climatic and phytogeographic forcing factors. Aust. Syst. Bot. 6:533–557.

Tuomisto H. 2010. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography* 33:2-22

Tuomisto H. 2010. A diversity of beta diversities: straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. *Ecography* 33:23-45

Twidale CR. 2007. Ancient Australian Landscapes. Chapter 9, Eastern uplands. Rosenberg Publishing Pty Ltd, Dural, pp 90-97.

Vamosi SM, Heard SB, Vamosi JC & Webb CO. 2009. Emerging patterns int he comparative analysis of phylogenetic community structure. *Mol. Ecol.* 18:572-592.

VanDerWal J Shoo LP, Graham C & Williams SE. 2009a. Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological modelling* 220:589-594

VanDerWal J, Shoo LP & Williams SE. 2009b. New approaches to understanding late Quaternary climate fluctuations and refugial dynamics in Australian wet tropical rain forests. *J. Biogeography* 36:291-301.

Webb L & Tracey J. 1981. Australian rainforests: pattern and change. In: *Ecological biogeography of Australia* (ed. J.A. Keast). W. Junk, The Hague.

Webb CO, Ackerly DD, McPeek MA & Donoghue MJ. 2002. Phylogenies and community ecology. *Ann. Rev. Ecol. Syst.* 33:475-505.

Wertheim JO, Sanderson MJ, Worobey M & Bjork, A. 2010. Relaxed molecular clocks, the bias-variance trade-off, and the quality of phylogenetic Inference. *Syst. Biol.* 59:1-8.

Wellman P. 1979. On the Cainozoic uplift of the southeastern Australian highland. *J. Geol. Soc. Aust*. 26:1-10.

Weir JT & Schluter D. 2007. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science* 315:1574-1576.

White MJD & Key KHL. 1957. A cytotaxonomic study of the pusilla group of species in the genus *Austroicetes* U.V. (Orthoptera: Acrididae). *Aust. J. Zoology* 5:56-87.

Whittaker RH. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monogr*. 30:279-338.

Wiens JJ & Donoghue MJ 2004. Historical biogeography, ecology, and species richness. *Trends Ecol. Evol*. 19:639–644.

Wiens JJ, C. H. Graham CH, Moen DS, Smith SA & Reeder TW. 2006. Evolutionary and ecological causes of the latitudinal diversity gradient in hylid frogs: treefrog trees unearth the roots of high tropical diversity. *Am. Nat.* 168:579–596.

Wiens JJ, Sukumaran J, Pyron RA & Brown RM. 2009. Evolutionary and biogeographic origins of high tropical diversity in old world frogs (Ranidae). Evolution 63:1217-1231.

Williams JE. 1991. Biogeographic patterns of three sub-alpine eucalypts in south-east Australia with special reference to *Eucalyptus pauciflora* Sieb. Ex Spreng. *J. Biogeography* 18:223-230.

Williams SE, Pearson RG &. Walsh PJ. 1996. Distributions and biodiversity of the terrestrial vertebrates of Australia's Wet Tropics . *Pacific Cons. Biol*. 2:327-362.

Wilson MV & Shmida A. 1984. Measuring beta diversity with presence-absence data. *J. Ecology* 72:1055–1064.

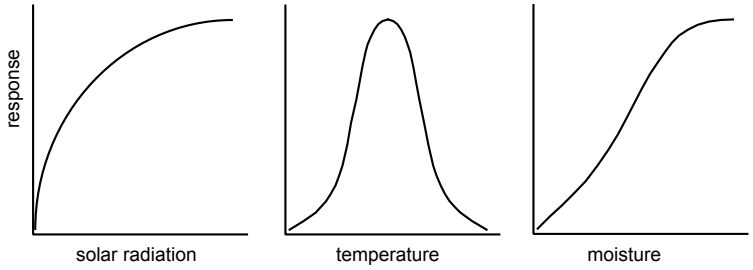
Winter JW. 1988. Ecological specialization of mammals in Australian tropical and sub-tropical rainforest: refugial and ecological determinism. In: Kitching ed. *The ecology of Australia's Wet Tropics*. Sydney, Surrey Beatty, pp 127-138.

Womble WH. 1951. Differential Systematics. *Science* 114:315-322.

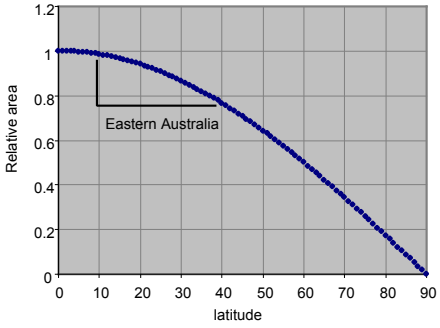
Zhang W & Shea M. 2008. A new genus and species of land snail of the family Camaenidae from New South Wales. *Molluscan Research* 28:123–132.

Appendix to Chapter 1

1A: Schematic of the Nix (1982) Plant Growth Index functions.

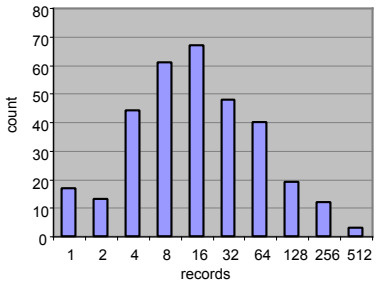


1B: Harvesine equal area function.

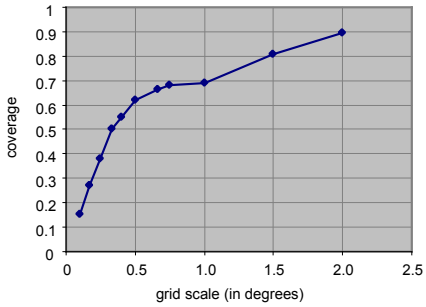


1C: Various characteristics of distribution records.

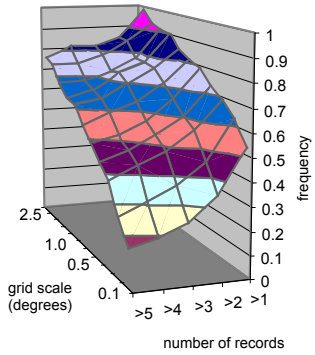
Histogram of records per taxa (eastern Australia) in the distribution dataset.



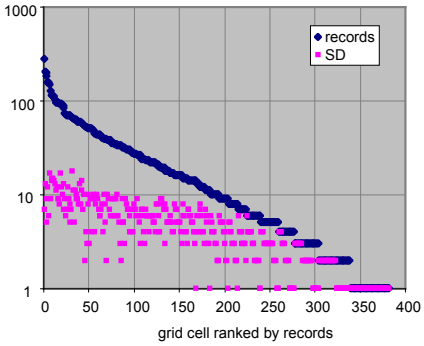
Grid cell scale and proportion of region covered.



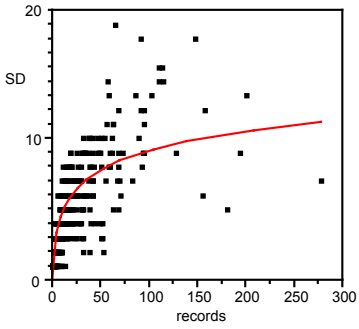
Grid scale and proportion of cells with few records



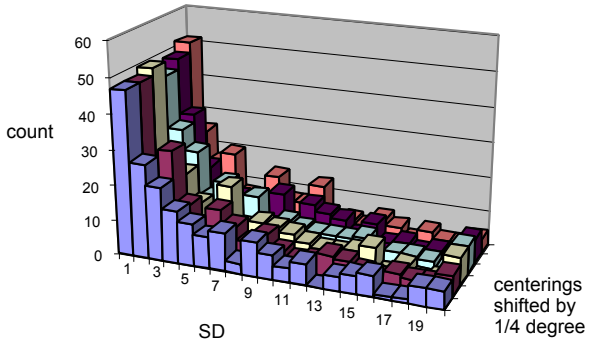
Relationship between records and SD for 1/2 degree grid cells



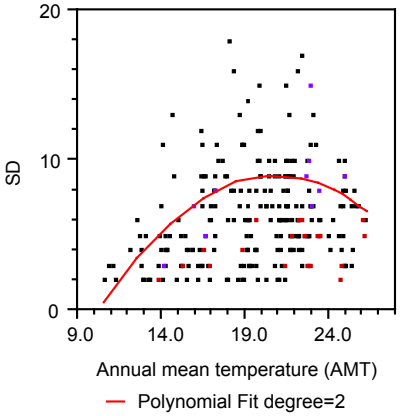
Species Diversity by records per 1/2 degree grid cell



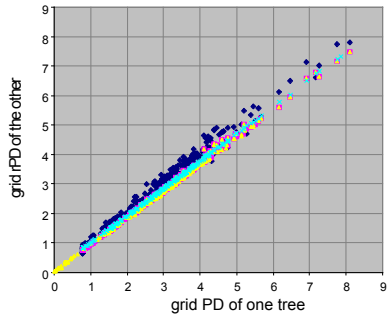
1D: Species Diversity of 1/2 degree grid cells with different centerings

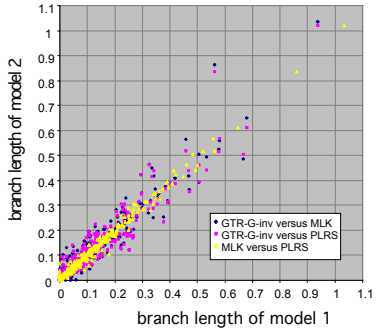


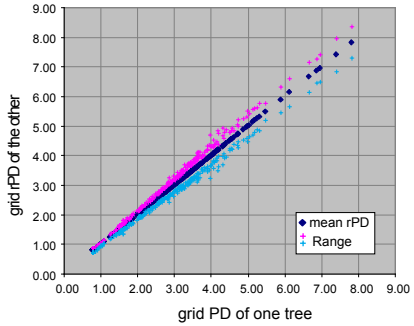
1E: Relationship between species diversity and temperature.



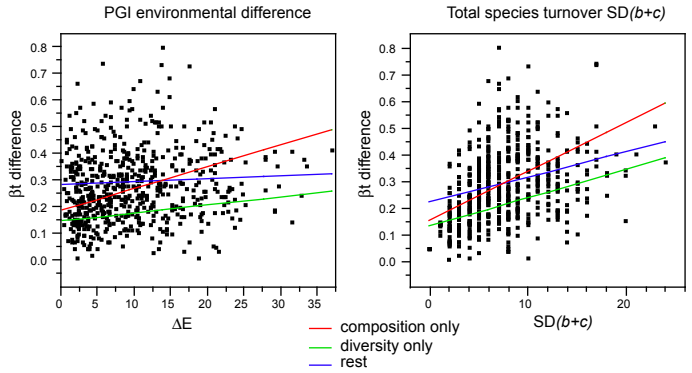
Appendix 5A. Assessing the effect on indexing PD of uncertainty in topology, branch length, and clock assumptions . The all compatible consensus tree (Chapter 2, Figure 7) and 10 randomly drawn MCMC samples were use to calculate PD of the set of 1/2 degree grid cells. This was done by: 1) using the unconstrained branch lengths of the original trees (GTR-G-inv); 2) PLRS transformed branch lengths; 3) strict-clock model transformed branch lengths using a single partition GTR-G-inv model branch lengths in PAUP 4.10b.

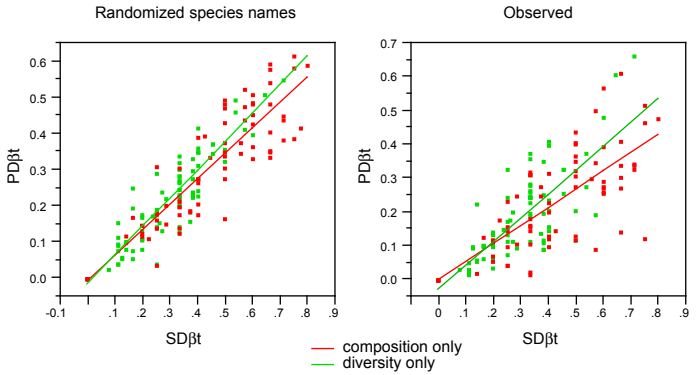


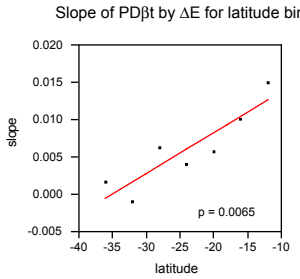




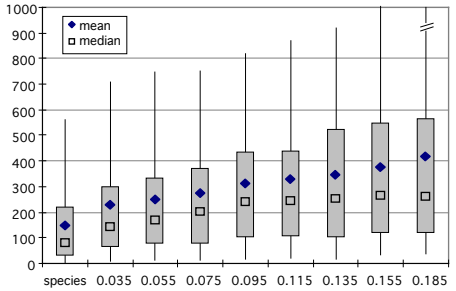
Appendix 5B. SD and PD bt turnover characteristics. Top: Difference between SDbt and PDbt by PGI difference DE. Adjacent grid cells only, separated into three categories: those with equal SD (composition only, n=77), those where one grid is a subset of the other (diversity only, n=74), and the remainder. Bottom: Intrinsic and observed relationship between PD and SD bt, adjacent grid cells composition and diversity categories.

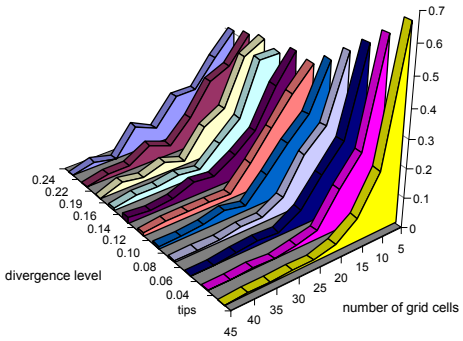






Appendix 5C. Range sizes of lineages. Top: Linear range distribution of divergence level lineages. Bottom: Distribution of divergence level lineages in 50km grid cells.



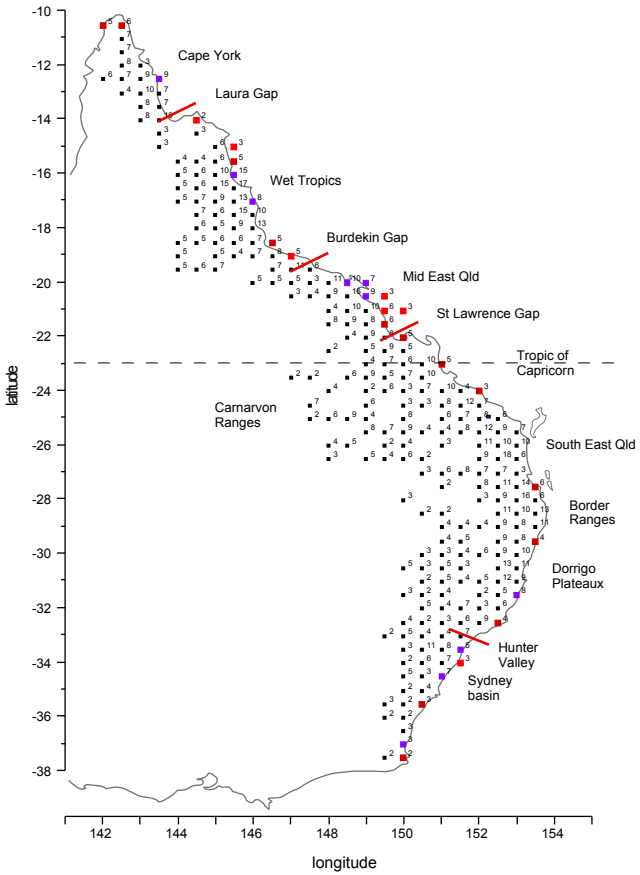


Appendix 5D. Correlations among various indices

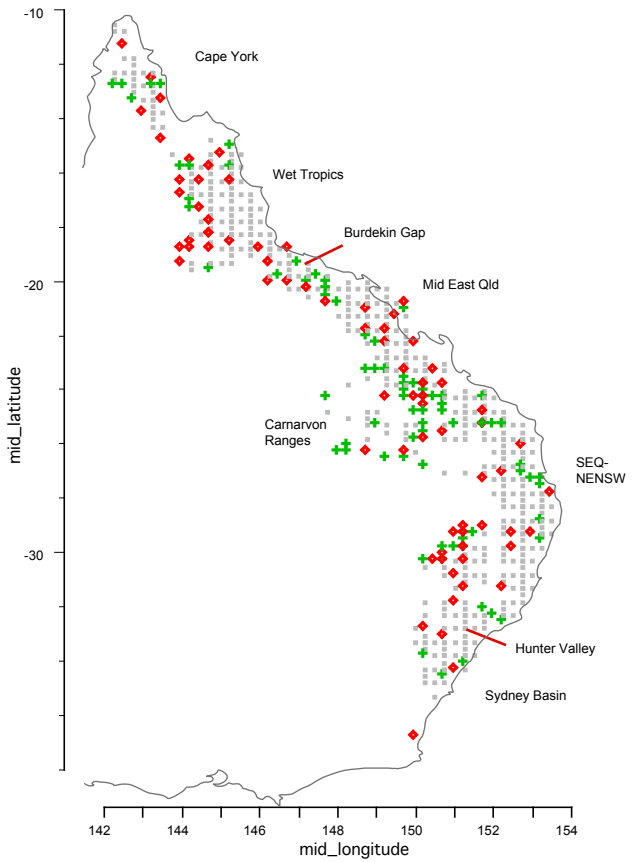
Variable	delta_rPD	delta_sigPDw	sd/pd	ln(invsd/invpd)	sd10drlr
delta_rPD	1.0000	0.2371	-0.6967	-0.5218	-0.6251
delta_sigPDw	0.2371	1.0000	0.1332	-0.2583	-0.0044
sd/pd	-0.6967	0.1332	1.0000	0.3521	0.6496
ln(invsd/invpd)	-0.5218	-0.2583	0.3521	1.0000	0.3854
sd10drlr	-0.6251	-0.0044	0.6496	0.3854	1.0000

Pearson product-moment correlation, included set of 231 50km grid cells
delta_rPD: relative PD as difference between observed PD and null PD
delta_sigPDw: relative endemic PD using sigmoidal function,
with null selection probability weighted by range
sd/pd: ratio of SD to PD
ln(invsd/invpd): natural log of the ratio of endemic SD to endemic PD using the inverse function
sd10drlr: ratio of tip lineages per to 0.1 divergent lineages

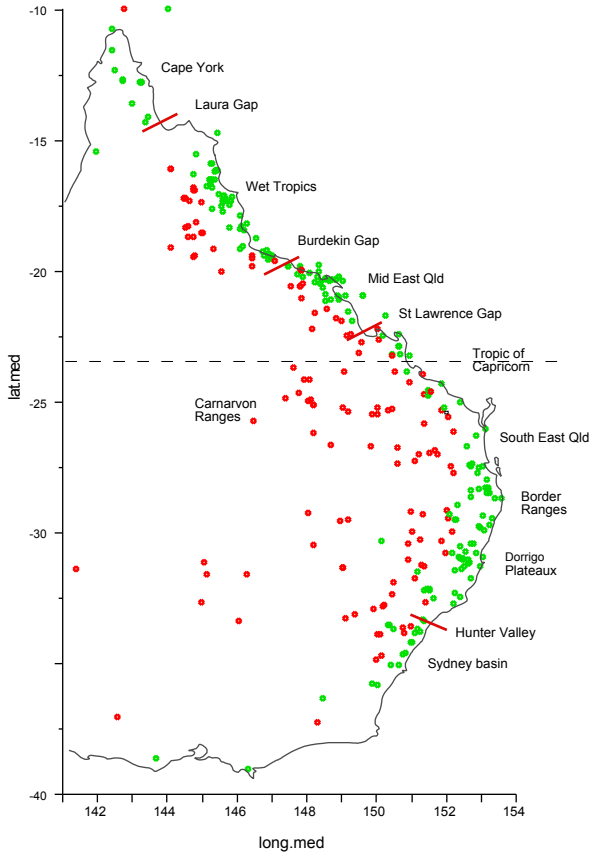
Appendix 5E. Set of grids used in multivariate diversity analyses. Starting with 266 grids, then excuding SD<2 and 20 grids with insufficient area (due to coastline) = 231 grids. Excluded grids shown in red, with grids of area< 40% of maximum shown in purple.



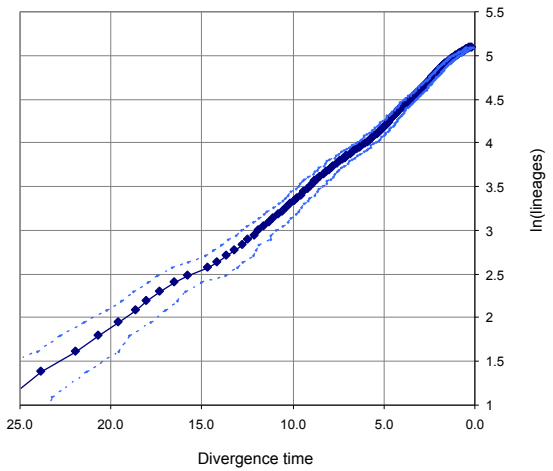
Appendix 5F. Geographic mid-points of turnover values used in adjacent grid cell analyses: red for change in composition only, green for change in number only, rest in grey.



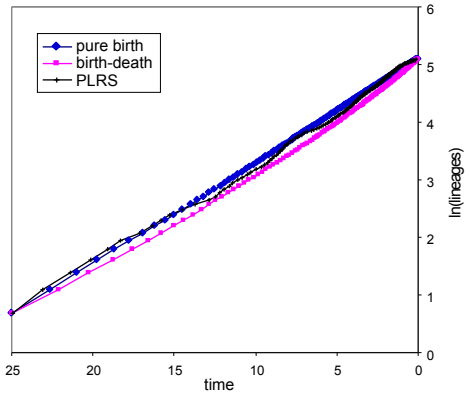
Appendix 5G. Example of mesic-xeric bioclimatic categorization. Based on species means, classification by AMP>1000 with PGI minimum cutoffs of mesothermal =48, megathermal =40, microthermal =43. Species median lat and long positions.



Appendix 5H. Hadroid BEAST
relaxed-clock lineage accumulation plot.



Appendx 5I. Lineage through time plots for simulated and observed data. Observed data is hadroid clade of the PLRS tree Appendix 5M. Other plots are mean of 200 simulations of pure birth (Yule) and birth-death (*b/d*) rate constant models that are the best fit to the hadroid data (see Chapter 5 Table 3A): pure birth with *r*=0.19 (*b*=0.19, *d*=0.0); birth-death with *r*=0.14 *a*=0.40 (*b*=0.23 *d*=0.09). Simulated in Phylo-Gen 1.1 (Rambaut 2002).



Appendix 5J. Multivariate model results using equal grid cell weighting. Observations (or Sum Wgts) = 231.

Climate model of Species Diversity (SD)

	Source	Estimate	F Ratio	Prob>F
linear and quadratic terms	c8_twq	0.0391	51.1227	<.0001
RSquare Adj = 0.517	meso	0.3442	20.2803	<.0001
	meso*meso	-0.0026	9.9669	0.0018
	mega3_sd	0.0645	0.0343	0.8532
	mega3_sd*mega3_sd	0.0832	2.9559	0.0869
	area	0.0004	1.0215	0.3133

Phylogenetic Diversity model of SD

	Source	Estimate	F Ratio	Prob>F
linear and quadratic terms	ln(sigsd/sigpd)	-0.2330	0.2104	0.6469
RSquare Adj = 0.105	ln(sigsd/sigpd)^2	-0.0831	1.6604	0.1989
	corrPD	9.2386	0.5380	0.4640
	corrPD*corrPD	-4.7586	0.5913	0.4427
	corr_sd10drlr	-0.9518	0.4700	0.4937
	area	-0.0010	4.1464	0.0429

Environmental model for relative PD

	Source	Estimate	F Ratio	Prob>F
linear and quadratic terms	c12_amp	-0.0004	1.9362	0.1655
RSquare Adj = 0.333	c12_amp*c12_amp	0.0000	6.0801	0.0144
	c8_twq	-0.0151	12.3646	0.0005
	c8_twq*c8_twq	0.0000	11.9418	0.0007
	c18_pwq	0.0006	0.8463	0.3586
	c18_pwq*c18_pwq	0.0000	4.7551	0.0303
	meso_sd	0.0267	3.8444	0.0512
	meso_sd*meso_sd	-0.0017	2.2836	0.1322
	latitude	0.0137	13.0934	0.0004

Appendix 5K. FNQ results using the full null phylogenetic pool.
Top: Phylogenetic clustering. Bottom: Phylogenetic turnover.

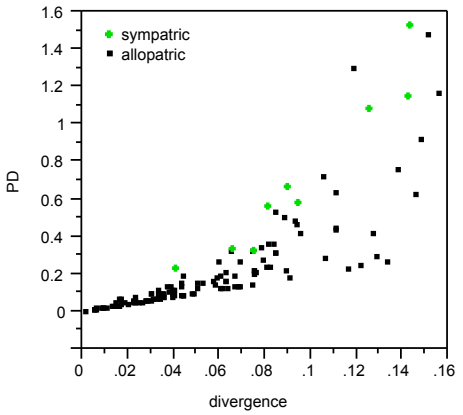
	all			hadroids		
category	n	relPD	p	n	relPD	CI
meso	30	0.87	*0.06*	13	0.74	*<0.01*
mega	16	0.91	*0.16*	9	0.64	*<0.01*
xeric	34	0.69	*0.01*	20	0.52	*<0.01*

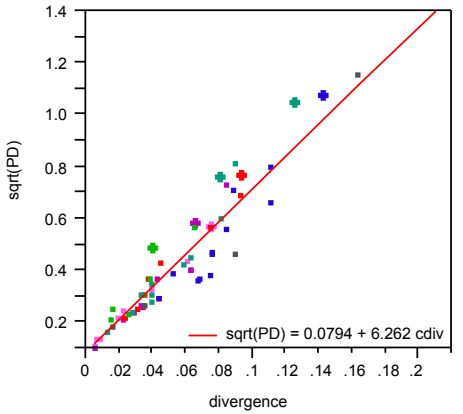
n = number of taxa in category
relPD = relative PD = observedPD/null PD
p = proportion of null randomizations with PD exceeding observed PD
Null determined using eastern Australian pool of taxa
200 replicates, equal probability of inclusion
Non-hadroids (rest) not done

	meso			mega		
	βt	relβt	*p*	βt	relβt	*p*
all						
mega	0.72	1.15	*0.98*			
xeric	0.82	1.31	*>0.99*	0.75	1.17	*>0.99*
hadroids						
mega	0.62	1.09	*0.74*			
xeric	0.67	1.17	*0.96*	0.42	0.94	*<0.01*

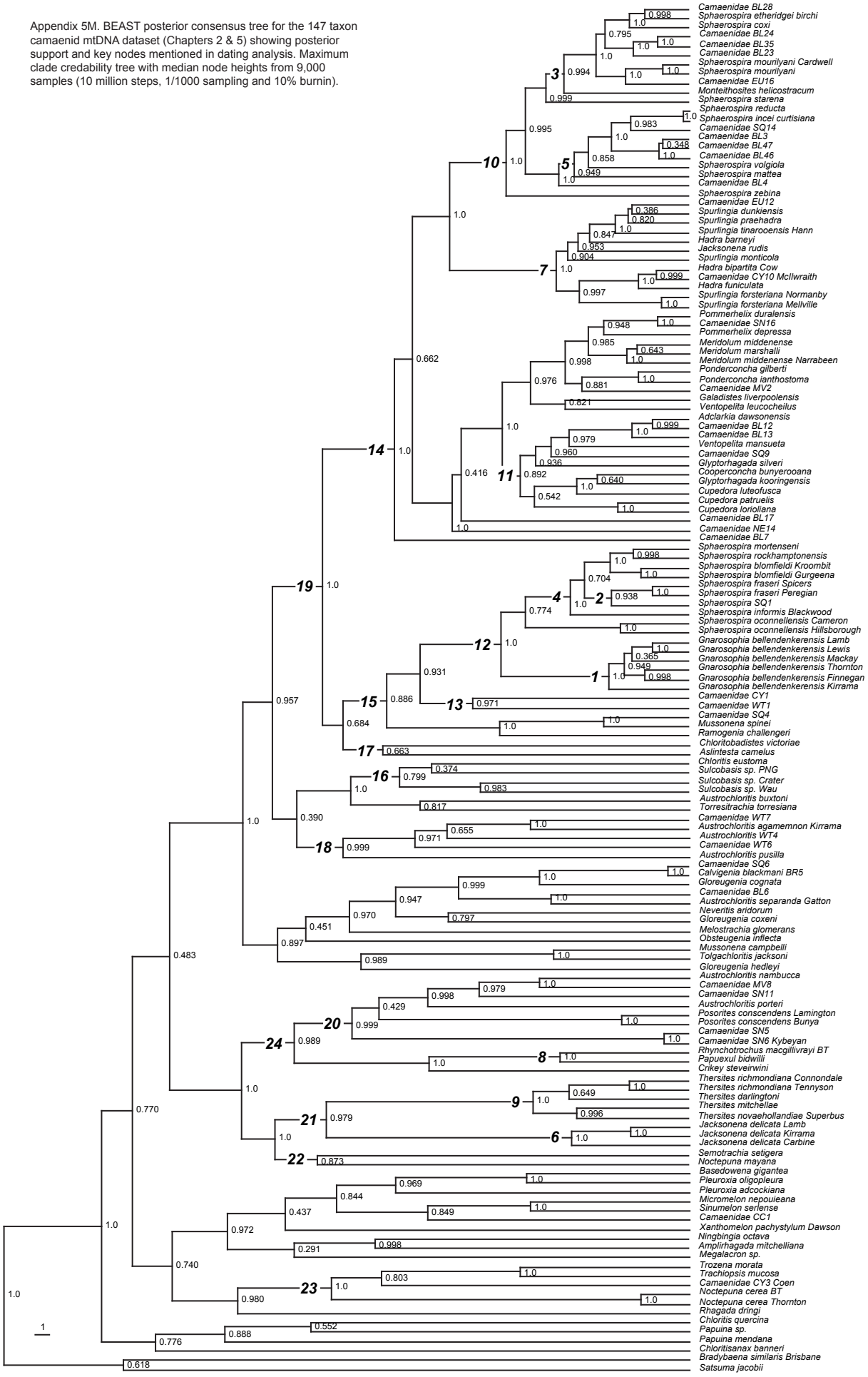
As for relative PD but relβt is βt/null
p = proportion of null randomizations with βt exceeding observed

Appendix 5L. Plots of divergence and PD in the hadroid clade. Top: Each point represents the divergence depth of a node in the phylogeny of east coast hadroids, with the sum PD of the subtended clade with sympatric nodes marked by crosses. The linear floor to the plot indicates clades with the minimim number of taxa (=2). Bottom: Nodes subtending the eight most closely related hadroid sympatric species groups. There is quite a reasonable fit to the square root of PD due to the general log linear LTT profile for much of the hadroids but this is only intended as a visula aid. At least six of the eight have relatively more PD that the general trend.





Appendix 5M. BEAST posterior consensus tree for the 147 taxon camaenid mtDNA dataset (Chapters 2 & 5) showing posterior support and key nodes mentioned in dating analysis. Maximum clade credability tree with median node heights from 9,000 samples (10 million steps, 1/1000 sampling and 10% burnin).



Appendix 5N. Penalized Likelihood Rate-Smoothed 279 taxa tree used for east coast diversity analyses, colour coded for the four categories of Far North Queensland taxa.

meso
mega
xeric
not east coast