Molecular systematics and biogeography of an Australian soil burrowing cockroach with polymorphic males, *Geoscapheus dilatatus* (Blattodea: Blaberidae: Geoscapheinae)

Running title: Molecular systematics of *Geoscapheus dilatatus*

Perry G. Beasley-Hall [a, b*], Harley A. Rose [a], Thomas Bourguignon [c], Nathan Lo [a]

[a] School of Life and Environmental Sciences, The University of Sydney, Sydney, New South Wales 2006, Australia

[b] School of Biological Sciences, The University of Adelaide, Adelaide, South Australia 5000, Australia

[c] Okinawa Institute of Science & Technology Graduate University, 1919-1 Tancha, Onna-son, Okinawa 904-0495, Japan

* Corresponding author

Perry Beasley-Hall

Postal Address: School of Biological Sciences, Darling Building, The University of Adelaide, Adelaide, South Australia 5000, Australia

Email: perry.beasley-hall@adelaide.edu.au

Abstract

An iconic group of arid-adapted insects is the Australian soil burrowing cockroaches (Blaberidae: Geoscapheinae), large, wingless insects that evolved burrowing behaviour and associated forms in parallel from wood feeding ancestors in the subfamily Panesthiinae. A particularly problematic taxon within the Geoscapheinae is *Geoscapheus dilatatus* (Saussure, 1864), which might represent a species complex and whose delimitation has been complicated for decades by the species harbouring polymorphic males. Males can be divided into two main morphs: individuals possessing horn-like protrusions on the anterior margin of the pronotum ("tuberculate") and those without these characters ("non-tuberculate"). A less common, third form consists of individuals that possess tubercles but are far larger than other tuberculate males and occur solely to the north of the species' distribution ("atypical" tuberculates). Here, we make use of whole mitochondrial genomes and nuclear ribosomal RNA data from individuals across the range of *G. dilatatus* to conduct the first phylogenetic analysis of this species to date. We recover all tuberculate males (including atypical forms) as monophyletic and the derived form of *G. dilatatus*, having evolved only once in this species, whereas non-tuberculate forms are paraphyletic. Fossil-calibrated molecular clock analysis revealed the divergence between these two forms occurred during the late Miocene approximately 6.7 million years ago, concurrent with an expansion of the continent's drier biomes. Environmental niche modelling suggests tuberculate male forms are more climatically tolerant than their more restricted non-tuberculate counterparts and both forms' predicted fundamental niches are strongly limited by rainfall. Three species delimitation analyses implemented here failed to consistently delimit *G. dilatatus* beyond a single species. Ultimately, population genetics approaches paired with additional sampling will be necessary to determine these findings more concretely, but at present we do not consider the results presented here sufficient to delimit *G. dilatatus* based on morphological differences found in the species' polymorphic males.

Key words: Blattodea, biogeography, systematics, sexual dimorphism, environmental niche modelling, Geoscapheinae, Blaberidae

Introduction

The impact of ancient climatic fluctuations on the present-day distributions of species has been well-characterised in taxa found in the Northern Hemisphere due to a biogeographic bias in the literature (Byrne *et al.* 2011; Riddle 2016). A comparable body of knowledge is lacking for Southern Hemisphere taxa, which might have been exposed to a different array of environmental conditions (e.g. differing genetic impacts of aridification *vs.* glaciation; Byrne *et al*. 2008). This is of particular relevance for Australian taxa, which have been subjected to a gradual drying and aridification of the continent since the land mass separated from Antarctica and drifted north during the Eocene (McLoughlin et al. 2001).

How has ancient environmental change in Australia impacted speciation processes? To date, the majority of studies that have addressed this question have focused on species either outside arid biomes or to the exclusion of eastern Australian arid and temperate semi-arid biomes (Chapple *et al.* 2011; Eldridge et al. 2014; Edwards et al. 2017; Ansari et al. 2019). In addition, studies examining taxa in these latter regions have focused almost entirely on vertebrates (e.g. James and Shine 2000; Schäuble and Mortiz 2001; Rabosky et al. 2014). There remains a lack of knowledge concerning arid-adapted invertebrate taxa and their phylogeography in eastern Australia as a result. Understanding the evolution of arid-adapted species—and being able to delimit potential species complexes—is also pertinent given predicted expansions of the Australian arid zone under current climate change modelling, which suggest an increase in thermal extremes in the east, e.g. the South West and Far North regions of Queensland and New South Wales (New South Wales Government 2019; Syktus *et al.* 2020).

Iconic residents of contemporary arid and temperate semi-arid regions of eastern Australia are the soil burrowing cockroaches (Blaberidae: Geoscapheinae), a group of large, wingless insects endemic to the continent. Members of the subfamily construct permanent underground burrows in sandy soil and feed almost exclusively on plant material such as dry leaf litter. The Geoscapheinae are derived from the wood feeding Panesthiinae (Maekawa *et al.* 2003) and acquired burrowing behaviour repeatedly and independently from these ancestors, presumably due to the aridification of the Australian continent during the Miocene (Lo *et al.* 2016; Beasley-Hall *et al.* 2018). Recent molecular work has shown the existing morphology-based taxonomic framework of these two subfamilies is inadequate (Lo *et al.* 2016) and all four genera within the Geoscapheinae emerge as polyphyletic under current classifications. Here, we investigate a poorly-understood species within the Geoscapheinae found across south-eastern Australia.

Geoscapheus dilatatus (Saussure 1864) is a widespread species that occurs from the east of South Australia extending to southern Queensland. Although its phylogeography has not been characterised, *G. dilatatus* is known from arid, temperate semi-arid, sub-tropical mesic, and temperate mesic biomes, and potentially represents a species complex (Roth 1977; Rickard 1998). It has historically been mistaken for its close relative *Geoscapheus robustus* Tepper, 1893: the two species have overlapping distributions and extremely similar morphologies to the extent that many of Tepper's type specimens of *G. robustus* were subsequently found to be *G. dilatatus* by Roth (1977). This has in part been due to *G. dilatatus* harbouring two forms of males, one with horn-like structures (tubercles) at the anterior margin of the pronotum paired with pronotal thickening (hereafter "tuberculate" males) and others without ("non-tuberculate" males), leading to confusion not only between these sister species–as *G. robustus* lacks such tubercles–but between *G. dilatatus* females and non-tuberculate males (Fig. 1) (Saussure 1895; Roth 1977). An "atypical" form of *G. dilatatus* restricted to Queensland also exists, containing tuberculate males larger than that of the two major forms and occurring further to the north of the species' distribution (Brown 1997; Rickard 1998). Despite these differences, genitalic morphology cannot be used to distinguish between males of these three forms (HAR, pers. obs.).

Tuberculate populations are the most common within *G. dilatatus*, occurring to the south of the species' distribution in the Maranoa region of Queensland, though the two major forms are sympatric in the Moonie area of that state. The remaining members of the species are largely distributed throughout central and western New South Wales, northern Victoria, and eastern South Australia. It is possible the tubercles of *G. dilatatus* play or once played a role in sexual competition as such structures are not present in females and geoscapheine males are known to fight by butting their pronota, although the structures are considerably reduced compared to those of other blaberids such as *Macropanesthia rhinoceros* Saussure, 1895 (Rugg and Rose 1991). It is unknown as to why these traits are polymorphic in *G. dilatatus* and why this morphology differs from other geoscapheine species, as most possess tubercles but none in a *dilatatus*-like arrangement.

Given most morphological characters in the Geoscapheinae are known to be phylogenetically uninformative (Lo *et al*. 2016; Beasley-Hall in prep.) and *G. dilatatus* has been suspected to comprise two or more cryptic species, additional methods have been used to attempt to delimit the species. These include examining cuticular hydrocarbons (Brown *et al.* 1997), allozyme allele frequencies (Humphrey *et al.* 1998), and chromosome counts (Olime 1988; Rickard 1998). In all of these cases tuberculate and non-tuberculate populations

have been distinguished from one another to varying degrees, but delimiting the species has been complicated by atypical tuberculate forms alternately grouping with one of the major male morphs. Allozyme data presented by Humphrey *et al.* (1998) suggested the two major forms are distinguishable genetically, and recent molecular phylogenies show considerable amounts of genetic variation between non-tuberculate and tuberculate individuals. Within the species itself, tuberculate and non-tuberculate forms are thought to have been isolated since the late Miocene (~5 Mya) based on molecular evidence from a small number of representatives (Lo *et al.* 2016; Beasley-Hall *et al.* in prep.). However, the precise timing of this split remains unresolved due to insufficient sampling, as does the ancestral state of the species. It is also unclear as to whether *G. dilatatus* represents a species complex or if its male forms are reflective of intraspecific variation.

Here, we present whole mitochondrial genomes (hereafter mitogenomes) and nuclear ribosomal RNA data from 28 individuals of *G. dilatatus* throughout the species' range. We specifically investigate: 1) the phylogenetic position of the three morphs of the species, and therefore how many times *G. dilatatus* has evolved tubercles; 2) when the three morphs of *G. dilatatus* diverged from one another; 3) whether evidence exists for the species representing more than one distinct taxon; and 4) the biogeography of this species in the context of its climatic tolerances and dispersal capabilities. We also discuss the taxonomic implications of the molecular phylogeny presented here.

Materials and Methods

Taxon sampling, DNA sequencing, and assembly

Specimens were collected from across New South Wales, Victoria, Queensland, and South Australia between 1987 and 2012 (Table 1). In total, 28 representatives of *G. dilatatus* and one representative of *G. robustus*, its sister species, were sequenced in this study. Specimens of *G. dilatatus* were selected to ensure their localities reflected as much of the species' known range as possible based on museum records and extensive sampling by HAR. Outgroups representing mitogenomes and nuclear *ITS1+18S* rRNA data from other members of the Geoscapheinae and wider Blaberidae were obtained from GenBank and Beasley-Hall *et al.* (in prep.) (Table 1).

DNA was extracted from cockroach fat bodies using a QIAGEN DNeasy Blood and Tissue Kit following the manufacturer's instructions. Genomic DNA libraries were prepared using an Ultra FS II Library Preparation Kit (New England Biolabs), as described in the manufacturer protocol, but with all reagent volumes divided by 10. We used the Unique Dual

Indexing Kit (New England Biolabs) to minimize contamination by index hopping. Pooled libraries were sequenced on a single Illumina HiSeq4000 lane and we obtained between 0.23 and 2.87 gigabases of reads for the resulting libraries.

Our raw reads were filtered against cockroach mitochondrial and nuclear ribosomal RNA reference sequence libraries using SAMtools and BWA (Li et al. 2009; Li and Durbin 2009) to exclude microbial reads from the cockroach endosymbiont *Blattabacterium*. Our reference libraries included the mitochondrial genomes of the Geoscapheinae from Beasley-Hall *et al.* (in prep.) and nuclear data from Lo *et al.* (2016) and Mukha *et al.* (2000). Filtered reads were then mapped against one tuberculate and one non-tuberculate *G. dilatatus* mitogenome from Beasley-Hall *et al.* (in prep., listed in Table 1) using Geneious' Read Mapper with "medium sensitivity" default settings, a 90% minimum overlap identity cut-off, and 25 fine-tuning iterations. Mitogenome annotation was performed using MITOS with default settings (Bernt *et al.* 2013), duplicated genes were corrected manually by merging multiple annotations into a single annotation per gene, and annotations were cross-checked by aligning the given mitogenome against non-geoscapheine outgroup taxa in Geneious. Per-gene alignments were constructed using MUSCLE (Edgar 2004) and our dataset was concatenated to include all coding genes, rRNAs, and tRNAs with the exclusion of intergenic regions and the mitochondrial control region, where recovered. Nuclear markers were recovered using the methods above and a reference sequence of the entire *18S*, *ITS1*, *5.8S*, *ITS2*, and *28S* rRNA genes sourced from *Diploptera punctata* (Eschscholtz, 1822) (Mukha *et al.* 2000). Where necessary, *de novo* assembly was performed using the SPAdes assembler (Bankevich *et al.* 2012) with default settings, sampling *k* values of 33, 55, 77, 91, and 121, and the resulting nuclear contigs were used to correct the output of our mapping step. Our mitochondrial and nuclear alignments represented a total of 22,339bp.

Phylogenetic analyses

Maximum-likelihood and Bayesian phylogenetic methods were used to infer relationships between our samples. We used PartitionFinder (Lanfear *et al.* 2012) to find the best partitioning scheme (1^{st} + 2^{nd} codon positions, 3^{rd} codon position, rRNAs, tRNAs) for our mitogenome dataset with the Bayesian Information Criterion and a greedy search algorithm. Maximum-likelihood analyses were performed on our mitogenome and nuclear datasets using RAxML (Stamatakis 2014) with 1000 bootstrap replicates and the GTR+G nucleotide substitution model for all partitions.

Bayesian analyses were performed in BEAST2 (v2.4.5, Bouckaert *et al.* 2014) using the uncorrelated relaxed lognormal clock and a birth-death tree prior to accommodate inter- and intraspecific sampling in our dataset (Ritchie *et al.* 2017). We used the package bModelTest to infer the nucleotide substitution model for each partition, the presence of invariant sites, and gamma rate heterogeneity (Bouckaert and Drummond 2017) and ran three independent chains of 100 million steps sampling every 5000 generations. Convergence to stationarity and effective sample size values of model parameters were checked using TRACER v1.7.1 (Rambaut 2014) and the maximum-clade credibility tree was inferred using TreeAnnotator v2.4.5 with a 10% burn-in (Bouckaert *et al.* 2014).

Few appropriate cockroach fossil calibrations exist for a divergence as apparently recent as that suggested by Lo *et al.* (2016) and Beasley-Hall *et al.* (in prep.) for this species, and as such we chose to include distantly related blaberid outgroups so the stem Blaberidae could be calibrated using *"Gyna" obesa* Piton 1940 per the best-practice recommendations of Evangelista *et al*. (2017). We implemented this calibration with an exponential distribution and soft maximum bounds, with a minimum age of 57.7 Mya and 145 Mya as a soft maximum bound to represent the first modern cockroach (Lin 1980; Bourguignon *et al.* 2018). Because this calibration point is distant from the Geoscapheinae, for the sake of robustness we performed a second molecular clock analysis using uncalibrated mitochondrial substitution rates. These values were sourced from Allegrucci *et al.* (2011) for the Mediterranean cave cricket genus *Dolichopoda* that began to diversify in the late Miocene based on the well-dated separation and isolation of two islands in the Tyrrhenian and Aegean Seas. Cave crickets have similar biology to geoscapheines in that they are subterranean and apterous with a limited dispersal capability. In addition, *Dolichopoda* began to diversify approximately 7 Mya, a comparable timescale to the estimated 5-6 My age of *G. dilatatus* per Lo *et al.* (2016). As the 12S and 16S genes had separate substitution rates available we split our single rRNA partition into two for this subsequent analysis. Calibrations were applied as lognormal priors on the clock rate of each partition in BEAST2, with the measure of uncertainty for each rate corresponding to the 2.5% and 97.5% quantiles of the distribution. As this latter calibration set could only be applied to mitochondrial data, both calibration methods implemented here were only applied to our mitogenome dataset to allow for meaningful comparisons between trees.

Environmental niche modelling

In order to assess whether different environmental factors determine the distribution of the widespread tuberculate males compared to their relatively restricted non-tuberculate counterparts, we constructed environmental niche models (ENMs) in MAXENT v3.3.3k (Phillips *et al*. 2006). ENMs seek to characterise the fundamental niche of species and predict their probability of occurrence in the absence of other limiting factors of their distributions, such as geographic barriers or competition. Our analysis considered localities from HAR's personal collection of 52 non-tuberculate and 158 tuberculate male individuals using ten jackknife replicates per male form with 75% of the data and the other 25% used to calculate probabilities. We retained duplicate localities due to the close proximity of some samples and assessed model performance using area-under-the-curve values. We selected 23 environmental variables pertaining to temperature, precipitation, and soil content sourced from the CliMond Archive and CSIRO's Australian Soil Resource Information System for our ENMs and excluded highly correlated variables following Beasley-Hall *et al.* (2018) (Hutchinson *et al.* 2009; ASRIS 2011; Kriticos *et al.* 2012). More detailed information for each of these variables is listed in Table S1.

To assess whether any climatic variables are consistently associated with the presence of pronotal tubercles in *G. dilatatus*, we also performed ancestral niche reconstructions (ANRs) using the *phyloclim* package in R, which collates the mean environmental tolerances of the ancestors of samples in a phylogenetic tree (R Core Team 2019; Evans *et al.* 2009). These reconstructions are commonly used to investigate the climatic preferences of ancestors shared by a certain node in a phylogeny but are also useful for visually comparing predicted mean climatic tolerances of lineages. As this method is limited by sample size, we only considered major differences between the two major male forms, sampling the species' known range of each of these morphs and not clades within each of these groups. For the sake of comparison, we also included published MAXENT probability surfaces of select geoscapheine species paired with the most up to date phylogenetic framework for the subfamily (Beasley-Hall *et al*. 2018; in prep.). Our ANRs considered all 23 environmental variables.

Species delimitation

We made use of three species delimitation methods applied to our *COX1* alignment and the maximum-clade credibility tree derived from our mitogenome analysis in BEAST2 (Fig. 2), as well as an additional set of analyses based off our combined phylogeny (Fig. S4). These represented one distance-based and two tree-based methods, as follows: automated barcode

gap discovery (ABGD) (Puillandre *et al.* 2012), the generalized mixed Yule coalescent (GMYC) (Pons *et al.* 2006), and the multi-rate implementation of the Poisson tree processes (mPTP) (Kapli *et al.* 2017).

For our distance-based analysis, we used the command line implementation of ABGD (Puillandre *et al.* 2012). GMYC was implemented using the *splits* package in R v3.6.2 with default settings (Ezard *et al.* 2009; R Core Team 2020). We performed mPTP analysis using both maximum likelihood and MCMC delimitation methods with differing rates of coalescence among species (the default *--multi* option). Our MCMC mPTP analysis was run for 100 million generations, sampling every 10,000 steps, with a burn-in of the first 10 million steps. We ran the MCMC analysis starting sampling from the ML species delimitation estimate, a random delimitation, and the null model.

Results

Phylogenetic analyses

The Bayesian and ML phylogenies inferred here produced near-identical topologies for our mitogenome dataset (Figs. 2, S1) and suggest non-tuberculate (NT) male forms are the ancestral state of *G. dilatatus*, with NT samples recovered as paraphyletic with respect to typical and "atypical" tuberculate (T) morphs. The earliest branching member within the phylogeny is the Injune sample (NT) sourced from Queensland, which diverged ~8.6 Mya (95% HPD 6.03–14.07 Mya), with the remaining NT individuals forming a monophyletic group with clade D and diverging from the more derived T morphs ~6.72 Mya (95% HPD 4.83–10.47 Mya; Fig. 2). The tuberculate clade began to diversify ~6.06 Mya (95% HPD 4.34–9.32 Mya) and form three major clades: one containing both typical and "atypical" tuberculates in southwest Queensland that is ~3.71 My old (95% HPD 3.06–7.55 Mya; clade C), a second with samples sourced from near the Queensland-New South Wales border (~4.36 My old, 95% HPD 3.01–6.74 Mya; clade B), and the third and largest clade containing individuals from New South Wales, Victoria, and South Australia (~4.72 My old, 95% HPD 3.41–7.16 Mya; clade A). All of the major groupings recovered in our mitogenome and combined phylogenies formed clusters with respect to their geographic origin. Our phylogeny calibrated using orthopteran substitution rates yielded younger dates to those above with narrower 95% HPD values. Under this alternate scenario, *G. dilatatus* began to diversify ~5.05 Mya (95% HPD 4.27–5.91 Mya).

We also inferred the evolutionary history of *G. dilatatus* from separate markers comprised of *18S*, *ITS1*, *5.8S*, *ITS2*, and *28S* nuclear rRNA genes (Fig. S3). There is discordance between our mitochondrial and nuclear phylogenies, potentially due to phenomena such as incomplete lineage sorting, hybridisation, and introgression in *G. dilatatus*, though such questions are beyond the scope of this study. Given phylogenetic incongruence is not uncommon between mitochondrial and nuclear datasets, we cannot rule out the validity of our nuclear topology in the present study simply because it differs from our mitochondrial tree (Fig. 2) or has poor node support. Under this alternate scenario, tubercules could have been acquired on up to four separate instances if secondary losses have not occurred, and only once if tubercules have subsequently been lost in the Condamine, Roma, Yuleba, Miles, Mitchell, and Kumbarilla samples (Fig. S3, pink clade). Finally, we also performed a concatenated phylogenetic analysis combining both mitochondrial and nuclear markers, and this tree yielded a topology identical to that of our dated mitogenome tree (Fig S4).

Environmental niche modelling and species delimitation

We used area-under-the-curve values to gauge how well our two environmental niche models (ENMs) perform compared to those computed from random background data. These values were high for ENMs of both NT and T male forms when all abiotic variables were used (0.95 and 0.98 for T and NT; models shown in Fig. 3). Non-tuberculate forms were most limited by the highest temperature during the warmest week of the year ("BIO05"), whereas tuberculate forms were limited by rainfall during the driest week of the year ("BIO14"). However, the latter variable had the most useful information when provided in isolation to the model for *both* forms when all variables in Table S1 were considered (instead of excluding those that were highly correlated), and so these differences might reflect limitations of the models themselves as opposed to limiting factors for each form. Alternate measures of variable importance, such as jackknife tests, can also be useful in assessing which variables are best able to predict test data from the training dataset. In this case, the two models shared rainfall during the driest week of the year ("BIO14") as the variable most useful for predicting the distribution of morphs of *G. dilatatus*.

Although there was notable overlap in the predicted fundamental niche of non-tuberculate and tuberculate taxa in southern QLD and northern NSW, that of tuberculate taxa was found to extend much further south-west into south-western NSW, north-western Victoria, and southeastern SA (Fig. 3). The tuberculate male morphs assessed here also

appear to have suitable habitat, in the absence of other limiting factors, in the Nullarbor region of West Australia.

To assess whether tuberculate forms have evolved to survive in more arid conditions overall than their non-tuberculate counterparts, we also performed ancestral niche reconstructions (ANRs) to compare the evolution of climatic preferences between these morphs and other Geoscapheinae and Panesthiinae sister-pairs (Fig. 4). The mean tolerance of tuberculate *G. dilatatus* forms for a given environmental variable often aligns with *G. robustus*, reflecting the similarities in their modelled fundamental niches. We note that the 80% central density of tolerance also tends to be much wider–or in the case of temperature and precipitation, higher and lower, respectively–than in tuberculate forms, suggesting non-tuberculate males are comparatively more tolerant of arid and extreme conditions.

The species delimitation methods employed here were not able to reliably delimit *G. dilatatus* (Fig. 2) or all grouped *G. dilatatus* into a single species (combined mitochondrial and nuclear analysis, Fig. S4). These results are in disagreement with our initial species hypothesis based on morphology, which posited that the non-tuberculate and tuberculate samples represented two distinct taxa. Finally, a Mantel test for isolation-by-distance found a significant correlation (Rxy = -0.357, $p = 0.01$) between increasing geographic distance and decreasing genetic identity among our samples, reflective of clades generally clustering by geographic location in our phylogeny (Fig. S6).

Discussion

Phylogenetic analyses

Here we present the first molecular phylogeny for members of the species *G. dilatatus*, originally thought to represent a species complex due to morphological variation in its males. The results of our phylogenetic analyses did not clearly split taxa in accordance with their morphological characteristics, with NT males recovered as paraphyletic with respect to T counterparts. This paraphyly was solely caused by the divergent Injune sample from Queensland (Fig. 2), though it was nonetheless grouped with the remainder of *G. dilatatus* specimens sampled in our species delimitation analyses. “Atypical” tuberculate male forms were not recovered as a monophyletic group as expected, indicating these individuals likely reflect variation among tuberculate populations as opposed to the former morphs constituting a discrete taxon.

That northern, non-tuberculate populations represent the ancestral state of *G. dilatatus*, with tuberculate populations to the west and south being derived, is consistent with previous

molecular systematic work on the Geoscapheinae. This subfamily is known to be polyphyletic within the larger Panesthiinae (Lo *et al.* 2016) and it is likely panesthiine ancestors of the Australian fauna migrated via the South-East Asian archipelago and arrived on the Australian continent ~25 Mya prior to dispersing south, as is inferred to be the case for *G. dilatatus* here (Fig. 2). The subsequent parallel evolution of burrowing forms occurred at least seven times from this wood feeding ancestor and is thought to have been spurred by the aridification of the Australian continent, which constituted bursts of expansion of drier, open habitats ~15 and ~7 Mya prior to an onset of severe aridity in the Pliocene (McLoughlin 2001; Byrne *et al.* 2011; Lo et al. 2016; Beasley-Hall *et al.* 2018). Both scenarios presented by our fossil and rate-calibrated trees suggest *G. dilatatus* began to diversity once these open habitats had expanded during the late Miocene (Figs. 2, S2). Nonetheless, the *Dolichopoda* substitution rates used as calibrations in Fig. S2 were themselves sourced from geological calibrations (Allegrucci *et al.* 2011) and their use here relied on a biased assumption of the timescale of diversification of G. *dilatatus*. As such, we consider the fossil calibration implemented in Fig. 2 a more reliable source of these diversification dates.

***Historical biogeography of* Geoscapheus dilatatus**

The biogeography of *G. dilatatus* is of particular note within the Geoscapheinae given its incidence in arid environments paired with a wide geographic distribution that is second only to its sister species *G. robustus*. The localities inhabited by these two species represent the hottest and driest known conditions occupied by any other member of the Geoscapheinae (Beasley-Hall *et al.* 2018) and these preferences are reflected in our biogeographic analyses. As stated previously, tuberculate male forms not only have wide known geographic distributions but also wide predicted fundamental niches, indicating they are potentially able to tolerate a robust variety of environmental niches in the absence of other limiting factors such as food availability, biogeographic barriers, or competition (Figs. 4, 5). The predicted area of occupancy for non-tuberculate males is also wide but does not extend as far west as that belonging to tuberculate forms and does not significantly diverge from the forms' current known distribution. Whether this might have been facilitated by an adaptive advantage on the part of tuberculate males–given their tolerance of more arid environments per our ancestral state reconstructions (Fig. 4)–is unclear.

At face value, the biogeography of this species appears to be a particularly striking instance of long-distance dispersal into Australia's arid and semi-arid zones by an apterous insect with presumably limited dispersal capabilities. The presence of early branching

lineages (Injune sample, clades C and D) in the northern part of the distribution *G. dilatatus*, with the more derived lineages (clades A and B) being found in more southern parts of the distribution is suggestive of a gradual dispersal of the species from the north to the south. These results suggest the species might be capable of migrating longer distances than once assumed given sufficient time on an evolutionary scale. Accordingly, *G. dilatatus* and *G. robustus* have been documented as being more active than other species within the Geoscapheinae: though males in many geoscapheine taxa are known to wander outside of burrows after periods of rainfall, *G. dilatatus* has been documented moving above-ground en masse in "spectacular migrations" (Roth 1977). While it is unclear if this behaviour reflects an ability to disperse long distances on an individual scale, the activity rhythms in this species, paired with a higher tolerance of aridity in tuberculate forms per our ENMs, might explain the wide distribution of these morphs (Roth 1977, HAR, pers. obs.).

The phylogeographic patterns we observe in *G. dilatatus* are broadly consistent with the evolutionary history of other Australian species in the east of the (semi-)arid zone, specifically lizards (James and Shine 2000; Rabosky *et al.* 2014; Ansari *et al*. 2019) and frogs (Schauble and Moritz 2001). These taxa all display a general distinction between populations in southern Queensland and western New South Wales, Victoria, and South Australia, similar to the split between our two major tuberculate clades A and B (Fig. 2) and tuberculate and non-tuberculate male morphs, though we note the latter have a less clear geographic distinction between them. Such study systems might indicate the presence of a past biogeographic barrier between these populations, though this question is beyond the scope of the present study.

Species delimitation

Only a small handful of cockroach delimitation studies exist to date, but those using molecular data have largely relied on *COXI* barcoding gap analyses to form species hypotheses (Che *et al*. 2017; Trotter *et al*. 2017; Yang *et al.* 2019). In contrast, the series of delimitation methods we employ here are based on a variety of different methods of calculating divergences between taxa, recommended as best practice in species delimitation studies (Carstens *et al*. 2013). ABGD seeks to partition samples based on their pairwise genetic distance, with the goal of finding a "gap" in this distance distribution representing the threshold between inter- and intraspecific variation. PTP methods directly measure the number of substitutions between samples to model speciation processes, with mPTP accounting for differing rates of evolution on each branch in a tree (Kapli *et al.* 2017). GMYC

seeks to find the maximum likelihood solution for a model that takes into account between- and within-species diversification on a time-calibrated ultrametric tree (Pons *et al.* 2006). Whereas distance-based methods like ABGD do utilise an explicit species concept, PTP and GMYC rely on the phylogenetic species concept, with the assumption of reciprocal monophyly of species in gene trees.

The methods we used provided conflicting signals regarding the taxonomic status of *Geoscapheus dilatatus* (Fig. 2). Our species delimitation analyses failed to reliably delimit *G. dilatatus* into more than a single species (Fig 2), in disagreement with our initial species hypothesis based on morphology, which posited that the non-tuberculate and tuberculate samples represented two distinct taxa. In our concatenated species delimitation analysis (Fig. S4), all three methods inferred that *G. dilatatus* was a single species. We recovered these results in spite of biases inherent in the three methods used here: for instance, ABGD is known to under-split species (Pentinsaari *et al.* 2017; Xu *et al.* 2019) whereas GMYC is known to over-split both simulated and empirical datasets (Miralles and Vences 2013; Zhang *et al.* 2013; Xu *et al.* 2019).

Our inability to clearly separate *G. dilatatus* into more than one species is reflected in previous studies that did not rely on molecular data. The earliest such study was that of Olime (1988), who assessed six tuberculate populations of the species against an equal number of *G. robustus* populations with respect to their chromosome counts; at this time the taxonomic status of the two species was in doubt. Olime (1988) demonstrated that, while the diploid chromosome number (2N) ranged between 45 and 61 for *G. dilatatus* and 47 to 53 for *G. robustus*, the total number of major chromosome arms–assuming most chromosomal changes have occurred via centric fusions–within the two species remained constant at 90 and 94, respectively. Rickard (1998) later focused on two tuberculate and two non-tuberculate populations of *G. dilatatus* and found a 2N of 57 to 61 in males.

Several of the localities examined in these studies correspond to samples in this study: members of clade A within the tuberculate group (Fig. 2) have either 2N = 45 (Bourke and Byrock; Olime 1988) or 61 (Wyandra; Rickard 1998), with atypical forms in clade C possessing either 2N = 57 (Augathella; Rickard 1998) or 61 (Charleville; Olime 1988). Overall, no clear divergences between the two major male forms could be identified on the basis of chromosome number in these studies.

Brown *et al.* (1997) and Humphrey *et al.* (1998) were able to distinguish tuberculate and non-tuberculate populations of *G. dilatatus* more readily, but they were nonetheless unable to find a clear distinction between tuberculate and non-tuberculate forms. Brown

(1997) assessed cuticular hydrocarbons across the species' range and found all tuberculate samples examined, as well as two atypical tuberculate individuals from Charleville and Eulo, belonged to the same phenotype, whereas the atypical Augathella sample grouped with the remaining non-tuberculate individuals. As cuticular hydrocarbons are used for interspecific recognition and social interactions in cockroaches and their allies (Lihoreau and Rivault 2009; Funaro *et al.* 2018), differing compositions in the two major male forms might indicate reproductive isolation. Humphrey *et al.* (1998) assessed allele allozyme frequencies in the Geoscapheinae as a whole and sampled 17 individuals of *G. dilatatus* to do so, including all three male forms. This study recovered all tuberculate samples, including atypical forms, as sister to non-tuberculate males, similar to the results presented here.

Ultimately, given the paraphyletic nature of non-tuberculate samples in our phylogenetic analyses and all three of our delimitation analyses failing to delimit *G. dilatatus*, we did not find evidence that would support splitting G. dilatatus on the basis of tubercle morphology. However, given the relatively deep divergences we recovered between the five major lineages presented here (Fig. 2) and known chromosome number variability between different populations, the question of whether the *G. dilatatus* is a species complex remains unresolved and requires further investigation.

Conclusions

Here, we sought to infer a phylogenetic framework for individuals in *Geoscapheus dilatatus* for the first time, a potential species complex within the Australian endemic soil burrowing cockroaches (Blaberidae: Geoscapheinae) that contains tuberculate, non-tuberculate, and "atypical" tuberculate male forms. Past studies have failed to delimit this species based on chromosome counts, cuticular hydrocarbons, and allozyme allele frequency data. We constructed a fossil-calibrated molecular phylogeny using mitochondrial genomes and nuclear data that recovered the non-tuberculate male morphs as paraphyletic with respect to the remaining tuberculate forms; these two major groupings diverged from one another approximately 6.7 million years before the present in the late Miocene and the species itself began to diversify ~8.6 Mya based on a fossil-calibrated mitogenome phylogeny presented here. An equally valid, yet less well-supported, phylogeny constructed from our nuclear dataset suggests that tubercules could have been acquired on up to four separate instances depending on whether secondary losses have occurred or not, and we cannot rule out this hypothesis in the present study. Both distance and tree-based species delimitation methods were unable to consistently delimit *G. dilatatus* into more than one species within the

constraints of our phylogeny. These findings suggest tuberculate morphology in this species is not representative specific variation in these cockroaches.

We also sought to further investigate the climatic tolerances of the two major male forms in our phylogeny by performing environmental niche modelling and ancestral niche reconstructions. The morphs of *G. dilatatus* have different predicted fundamental niches to one another as modelled in MAXENT, and tuberculate forms (which are much more widespread geographically) appear to be more tolerant of broader environmental conditions than their non-tuberculate counterparts, particularly with respect to variables linked to aridity. Whether this is due to an adaptive advantage on the part of tuberculate males remains to be seen.

Overall, the widespread distribution of *G. dilatatus* is a remarkable instance of dispersal into Australia's arid zone by an apterous species that might be expected to have quite limited dispersal capabilities. The long evolutionary history (i.e. >8 Myr) of this species and its unique activity rhythms within the Geoscapheinae might explain its wide distribution. The results presented here contribute further to the understudied phylogeography of invertebrates in the east of Australia's arid and semi-arid biomes, and do not support *G. dilatatus* being delimited into more than one taxon based on morphological variation alone. Further studies would benefit from increased sampling across the species' range and the use of genome-wide nuclear markers to better examine gene flow between different populations. Subsequent environmental niche modelling could potentially take advantage of this to model the response of *G. dilatatus* to thermal extremes predicted under current climate change projections.

Acknowledgments

The authors wish to thank Crystal Clitheroe for library preparation and the Okinawa Institute of Science and Technology (OIST) DNA sequencing section for Illumina sequencing. This research was supported by the Australian Research Council (grant number FT160100463) and by subsidiary funding to OIST.

References

- Allegrucci, G, Trucchi, E & Sbordoni, V. 2011. Tempo and mode of species diversification in *Dolichopoda* cave crickets (Orthoptera, Rhaphidophoridae). *Molecular Phylogenetics and Evolution* **60**, 108-121.

- Ansari, MH, Cooper, SJB, Schwarz, MP, Ebrahimi, M, Dolman, G, Reinberger, L, Saint, KM, Donnellan, SC, Bull, CM & Gardner, MG. 2018. Plio-Pleistocene diversification and biogeographic barriers in southern Australia reflected in the phylogeography of a widespread and common lizard species. *Molecular Phylogenetics and Evolution* **133**, 107-119.
- ASRIS. 2011. ASRIS Australian Soil Classification. *CSIRO Land and Water*. Available from: http://asris.csiro.au [Accessed August 4 2020]
- Bankevich, A, Nurk, S, Antipov, D, Gurevich, AA, Dvorkin, M, Kulikov, AS, Lesin, VM, Nikolenko, SI, Prjibelski, AD, Pyshkin, AV, Sirotkin, AV, Vyahhi, N, Tesler, G, Alekseyev, MA & Pevzner, PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**, 455-477.
- Beasley-Hall, PG, Lee, TRC, Rose, HA & Lo, N. 2018. Multiple abiotic factors correlate with parallel evolution in Australian soil burrowing cockroaches. *Journal of Biogeography* **45**, 1515–1528.
- Bernt, M, Donath, A, Jühling, F, Externbrink, F, Florentz, C, Fritzsch, G, Pütz, J, Middendorf, M & Stadler, PF. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* **69**, 313-319.
- Bouckaert, R, Heled, J, Kühnert, D, Vaughan, T, Wu, C-H, Xie, D, Suchard, MA, Rambaut, A & Drummond, AJ. 2014. BEAST2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**, e1003537.
- Bouckaert, R & Drummond, AJ. 2017. *bModelTest: Bayesian phylogenetic site model averaging and model comparison*. *BMC Evolutionary Biology* **17**, 42.
- Bourguignon, T, Qian, T, Ho, SYW, Juna, F, Wang, Z, Arab, DA, Cameron, SL, Walker, J, Rentz, D, Evans, TA & Lo, N. 2018. Transoceanic and plate tectonics shaped global cockroach distributions: evidence from mitochondrial phylogenomics. *Molecular Biology and Evolution* **35**, 970-983.
- Brown, WV, Rose, HA & Lacey, MJ. 1997. The cuticular hydrocarbons of the soil burrowing cockroach Geoscapheus dilatatus (Saussure) (Blattodea: Blaberidae: Geoscapheinae) indicate species dimorphism. *Comparative Biochemistry and Physiology B* **118**, 549-562.
- Byrne, M, Steane, DA, Joseph, L, Yeates, DK, Jordan, GJ, Crayn, D, Aplin, K, Cantrill, DJ, Cook, LG, Crisp, MD, Keogh, S, Melville, J, Moritz, C, Porch, N, Sniderman, JMK, Sunnucks, P & Weston, PH. 2011. Decline of a biome: evolution, contraction,

fragmentation, extinction and invasion of the Australian mesic zone biota. *Journal of Biogeography* **38**, 1635-1656.

- Byrne, M, Yeates, DK, Joseph, L, Kearney, M, Bowler, J, Williams, MAJ, Cooper, S, Donnellan, SC, Keogh, JS, 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. *Molecular Ecology* **17**, 4398-4417.
- Carstens, BC, Pelletier, TA, Reid, NM & Satler, JD. 2013. *How to fail at species delimitation*. Molecular Ecology **22**, 4369-4383.
- Chapple, DG, Chapple, SNJ & Thompson, MB. 2011. Biogeographic barriers in south-eastern Australia drive phylogeographic divergence in the garden skink, Lampropholis guichenoti. *Journal of Biogeography* **38**, 1761-1775.
- Che, Y, Gui, S, Lo, N, Ritchie, A & Wang, Z. 2017. Species delimitation and phylogenetic relationships in ectobiid cockroaches (Dictyoptera, Blattodea) from China. *PLoS ONE* **12**, e0169006.
- Dumans, ATN, Grimaldi, DB, Furtado, C, Machado, Machado, EA, & Prosdocimi, F. 2017. The complete mitochondrial genome of the subsocial cockroach Nauphoeta cinerea and phylogenomic analyses of Blattodea mitogenomes suggest reclassification of superfamilies. *Mitochondrial DNA Part B* **2**, 76-78.
- Edgar, RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797.
- Edwards, DL, Melville, JM, Joseph, L & Keogh, JS. 2015. Ecological divergence, adaptive diversification, and the evolution of social signalling traits: an empirical study in arid Australian lizards. *The American Naturalist* **186**, E144-E161.
- Edwards, RD, Crisp, MD, Cook, DH & Cook, LG. 2017. Congruent biogeographical disjunctions at a continent-wide scale: Quantifying and clarifying the role of biogeographic barriers in the Australian tropics. *PLoS ONE* **12**, e0174812.
- Eldridge, MDB, Potter, S. Johnson, CN & Ritchie, EG. 2014. Differing impact of a major biogeographic barrier on genetic structure in two large kangaroos from the monsoon tropics of Northern Australia. *Ecology and Evolution* **4**, 554-567.
- Evangelista, DA, Djernæs, M, & Kohli, MK. 2017. Fossil calibrations for the cockroach phylogeny (Insecta, Dictyoptera, Blattodea), comments on the use of wings for their identification, and a redescription of the oldest Blaberidae. *Palaeontologia Electronica* 20.3.1FC: 1-23.

- Evans, MEK, Smith, SA, Flynn, RS & Donoghue, MJ. 2009. Climate, niche evolution, and diversification of the "bird-cage" evening primroses. *The American Naturalist* **173**, 225-240.
- Ezard, T, Fujisawa, T & Barraclough, TG. 2009. SPLITS: Species' Limits by Threshold Statistics. Available from: http://r-forge.r-project.org/projects/splits [Accessed 4 August 2020]
- Funaro, CF, Böröczky, K, Vargo, EL & Schal, C. 2018. Identification of a queen and king recognition pheromone in the subterranean termite Reticulitermes flavipes. *Proceedings of the National Academy of Sciences* **115**, 3888-3893.
- Humphrey, M, Rose, HA & Colgan, DJ. 1998. Electrophoretic studies of cockroaches of the Australian endemic subfamily Geoscapheinae. *Zoological Journal of the Linnean Society* **124**, 209-234.
- Hutchinson, M, Xu, T, Houlder, D, Nix, H & McMacon, J. 2009. *ANU-CLIM 6.0 user's guide*. Canberra: Australian National University, Fenner School of Environment and Society.
- James, CD & Shine, R. 2000. Why are there so many coexisting lizards in Australian deserts? *Oecologia* **125**, 127-141.
- Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-1405.
- Kapli, T, Lutteropp, S, Zhang, J, Kobert, K, Pavlidis, P, Stamatakis, A, & Flouri, T. 2016. Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics* **33**, 1630-1638.
- Kriticos, DJ, Webber, BL, Leriche, A, Ota, N, Macadam, I, Bathols, J, & Scott, JK. 2012. CliMond: global high resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution* **3**, 53-64.
- Kumar, S, Stecher, G, Li, M, Knyaz, C, & Tamura, K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* **35**, 1547-1549.
- Lanfear, R, Calcott, B, Ho, SYW, & Guindon, S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29**, 1695–1701.
- Li, H & Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **25**, 1754-1760.

- Li, H, Handsaker, B, Wysoker, A, Fennell, T, Ruan, J, Homer, N, Marth, G, Abecasis, G, Durbin, R, & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.
- Lihoreau, M & Rivault, C. 2009. Kin recognition via cuticular hydrocarbons shapes cockroach social life. *Behavioural Ecology* **20**, 46-53.
- Lin, Q. 1980. Fossil insects from the Mesozoic of Zhejiang and Anhui provinces. In: Nanjing Institute of Geology and Palaeontology, Chinese Academy of Science (eds). *Division and correlation of the mesozoic volcano-sedimentary strata in Zhejiang and Anhui Provinces*, p. 211-234. Beijing: Science Press.
- Lo, N, Beninati, T, Stone, F, Walker, J, & Sacchi, L. 2007. Cockroaches that lack Blattabacterium endosymbionts: the phylogenetically divergent genus Nocticola. *Biology Letters* **3**, 327-330.
- Lo, N, Tong, K. J, Rose, HA, Ho, SYW, Beninati, T, Low, DLT, Matsumoto, T, & Maekawa, K. (2016). Multiple evolutionary origins of Australian soil burrowing cockroaches driven by climate change in the Neogene. *Proceedings of the Royal Society B: Biological Sciences* **283**, 20152869.
- McLeish, ML, Miller, JT, & Mound, LA. (2013). Delayed colonisation of Acacia by thrips and the timing of host-conservatism and behavioural specialisation. *BMC Evolutionary Biology* **13**, 188.
- McLoughlin, S. 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany* **49**, 271-300.
- Miralles, A & Vences, M. 2013. New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in Madascincus lizards. *PLoS ONE* **8**, e68242.
- Mukha, D, Wiegmann, BM, & Schal, C. 2000. Evolution and phylogenetic information content of the ribosomal DNA repeat unit in the Blattodea (Insecta). *Insect Biochemistry and Molecular Biology* **32**, 951-960.
- New South Wales Government. 2019. New South Wales climate change snapshot. Office of Environment and Heritage. Available from: https://climatechange.environment.nsw.gov.au/~/media/8C0EFD5C6C584BA488DE0AF7DF67635C.ashx [Accessed 4 August 2020]
- Olime, MT. 1988. Population cytogenetics of two soil burrowing cockroaches Geoscapheus dilatatus and Geoscapheus robustus (Blattodea: Geoscapheinae). Unpublished thesis.

- Pentinsaari, M, Vos, R, & Mutanen, M. 2017. Algorithmic single-locus species delimitation: effects of sampling effort, variation and nonmonophyly in four methods and 1870 species of beetles. *Molecular Ecology Resources* **17**, 393-404.
- Phillips, SJ, Anderson, RP, & Schapire, RE. 2006. Maximum entropy modelling of species geographic distributions. *Ecological Modelling* **190**, 231-259.
- Pons, J, Barraclough, TG, Gomez-Zurita, J, Cardoso, A, Duran, DP, Hazell, S, Kamoun, S, Sumlin, WD, & Vogler, AP. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* **55**, 595-609.
- Puillandre, N, Lambert, A, Brouillet, S, & Achaz, G. 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* **21**, 1864-1877.
- R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/ [Accessed 4 August 2020]
- Rabosky, DL, Hutchinson, MN, Donnellan, SC, Talaba, AL, & Lovette, IJ. 2014. Phylogenetic disassembly of species boundaries in a widespread group of Australian skinks (Scincidae: Ctenonus). *Molecular Phylogenetics and Evolution* **77**, 71-82.
- Rambaut, A, Drummond, AJ, Xie, D, Baele, G, & Suchard, MA. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67**, 901-904.
- Rickard, R. 1998. Genetic discrimination of Geoscapheus dilatatus through karyotypic characterisation. Unpublished thesis.
- Riddle, B. R. 2016. Comparative phylogeography clarifies the complexity and problems of continental distribution that drove A. R. Wallace to favor islands. *Proceedings of the National Academy of Sciences* **113**, 7970-7977.
- Ritchie, AM, Lo, N, & Ho, SYW. 2017. The impact of the tree prior on molecular dating of data sets containing a mixture of inter- and intraspecific sampling. *Systematic Biology* **66**, 413-425.
- Rugg, D & Rose, HA. 1984. Reproductive biology of some Australian cockroaches (Blattodea: Blaberidae). *Journal of the Australian Entomological Society* **23**, 113-117.
- Rugg, D & Rose, HA. 1991. Biology of Macropanesthia rhinoceros Saussure (Dictyoptera: Blaberidae). *Annals of the Entomological Society of America* **84**, 575-582.
- Saussure, H. de. 1864. Blattarum novarum species aliquot. *Revue et magasin de zoologie pure et appliqueé* **16**, 305-326.

- Saussure, H. de 1895. Révision de la tribu des Panesthiens et de celle des Épilampriens (Orthoptères de la famille des Blattides). Revue Suisse de Zoologie **3**, 299-364.
- Schäuble, CS, & Moritz, C. 2001. *Comparative phylogeography of two open forest frogs from eastern Australia*. Biological Journal of the Linnean Society 74: 157-170.
- Syktus, J, Trancoso, R, Ahrens, D, Toombs, N, & Wong, K. 2020. Queensland Future Climate Dashboard: Downscaled CMIP5 climate projections for Queensland. Available from: https://www.longpaddock.qld.gov.au/qld-future-climate/ [Accessed 4 August 2020]
- Trotter, AJ, McRae, JM, Main, DC, & Finston, TL. 2017. Speciation in fractured rock landforms: towards understanding the diversity of subterranean cockroaches (Dictyoptera: Nocticolidae: Nocticola) in Western Australia. *Zootaxa* **4250**, 143-170.
- Xiao, B, Chen, A-H, Zhang, Y-Y, Jiang, G-F, Hu, C-C, Zhu, C-D. Complete mitochondrial genomes of two cockroaches, *Blattella germanica* and *Periplaneta americana*, and the phylogenetic position of termites. *Current Genetics* **58**, 65-77.
- Xu, X, Kuntner, M, Bond, JE, Ono, H, Ho, SYW, Liu, F, Yu, L, & Li, D. 2019. Molecular species delimitation in the primitively segmented spider genus Heptathela endemic to Japanese islands. *Molecular Phylogenetics and Evolution* **151**, 106900.
- Yang, R, Wang, Z, Zhou, Y, Wang, Z, & Che, Y. 2019. Establishment of six new Rhabdoblatta species (Blattodea, Blaberidae, Epilamprinae) from China. *ZooKeys* **851**, 27-69.

Table 1: sampling locations of *G. dilatatus* individuals in this study. Male forms of *G. dilatatus* are either denoted as non-tuberculate (NT), tuberculate (T), or atypical (A). QLD = Queensland; NSW = New South Wales; VIC = Victoria; SA = South Australia. NA = sequence not recovered. Taxa in grey were not sequenced as part of this study.

			GenBank accession no.	
Species	**♂ form**	**Collection locality/source**	**Mitogenome**	**Nuclear genes**
Geoscapheus dilatatus	NT	Condamine, QLD (20km ENE)	MW600980	MW365851
	NT	Injune, QLD (36km SW)	MW600979	MW365845
	NT	Kogan, QLD (5.8km NNW)	MW600977	NA
	NT	Kumbarilla, QLD (5km N)	MW600982	MW365849
	NT	Mitchell, QLD (Beasley-Hall *et al.* in prep)	MW354074	18S: MW365870; ITS1: MW365806
	NT	Roma, QLD (26km NE)	MW600978	MW365850
	NT	Yuleba, QLD (14km ESE)	MW600981	MW365842
	T	Blackall, QLD (85km WSW)	MW600984	MW365853
	T	Bourke, NSW (36km N)	MW600991	MW365846
	T	Broken Hill, NSW (101km S)	MW600989	MW365866
	T	Byrock, NSW (1.8km SSE)	MW600990	NA
	T	Chinkapook, VIC (6.3km S)	MW601001	MW365864
	T	Cockburn, SA (42km NNW)	MW601003	MW365862
	T	Dimboola, VIC (1.2km NNE)	MW601000	MW365859
	T	Eulo, QLD	NA	MW365841
	T	Gilgandra, NSW (11km NE)	MW600997	MW365863
	T	Goondiwindi, QLD (42km NW)	MW600983	MW365848
	T	Gwabegar, NSW (7km S)	MW600996	MW365858
	T	Hattah, VIC (1.1km E)	MW600999	MW365840
	T	Mendooran, NSW (18km SSW)	MW600998	MW365857
	T	Menindee, NSW (20km S)	MW600992	MW365868
	T	Miles, QLD (27km NNW)	NA	MW365852
	T	Moonie, QLD (16km S)	NA	MW365847
	T	Patchewollock, VIC (Beasley-Hall *et al.* in prep)	MW354075	18S: MW365871; ITS1: MW365807
	T	Renmark, SA (19km NW)	MW601004	MW365865
	T	Ungarie, NSW (6.5km SE)	MW600994	MW365843
	T	Urana, NSW (15km NE)	MW600993	MW365860
	T	Walpeup, VIC (1.6km SE)	MW601002	MW365867
	T	Wyandra, QLD (18km NNE)	MW600985	MW365856
	T	Yenda, NSW (4.3km ENE)	MW600995	MW365861
	T (A)	Augathella, QLD (11km N)	MW600987	MW365854
	T (A)	Charleville, QLD (52km NNE)	MW600986	MW365855
Outgroup				

Geoscapheus robustus	-	Wentworth, NSW (35km NNE)	MW600988	MW365844
Macropanesthia lithgowae	-	Beasley-Hall *et al.* (in prep)	MW354066	18S: MW365878; ITS1: MW365814
Macropanesthia mutica	-		MW354067	18S: MW365881; ITS1: MW365817
Macropanesthia rothi	-		MW354068	18S: MW365882; ITS1: MW365818
Panesthia australis	-		MW354070	18S: MW365887; ITS: MW365823
Panesthia matthewsi	-		MW354071	18S: MW365889; ITS1: MW365825
Panesthia ancaudellioides	-		MW354069	18S: MW365886; ITS1: MW365822
Panesthia tryoni tryoni	-		MW354072	18S: MW365901; ITS1: MW365835
Parapanesthia gigantea	-		MW354073	18S: MW365904; ITS1: MW365838
Diploptera punctata (Diplopterinae)	-	Bourguignon *et al.* (2018)	MG882143	-
Blattella germanica (Ectobiidae)	-	Xiao *et al.* (2012)	NC_012901.1	-
Epilampra maya (Epilamprinae)	-	Bourguignon *et al.* (2018)	MG882194	-
Galiblatta cribosa (Epilamprinae)	-		MG882232	-
Gyna capucina (Gyninae)	-		MG882152	-
Nauphoeta cinerea (Oxyhaloinae)	-	Dumans *et al.* (2017)	KY212743	-
Neolaxta mackerrasae (Perisphaerinae)	-	Bourguignon *et al.* (2018)	MG882201	-
Paranauphoeta circumdata (Paraneuphoetinae)	-		MG882225	-
Pycnoscelus sp. (Pycnoscelinae)	-		MG882200	-
Rhabdoblatta sp. (Epilamprinae)	-		MG882228	-
Schultesia lampyridiformis (Zetoborinae)	-		MG882163	-

1

Fig. 1: dorsal and front-on views of tuberculate (A) and non-tuberculate (B) male forms of *Geoscapheus dilatatus*. Tuberculate forms are typified by horn-like protrusions (tubercles) on the anterior margin of the pronotum, denoted here by arrows. "Atypical" tuberculate forms are identical in tubercle morphology to typical tuberculates but tend to have a larger body size. All females in the species are non-tuberculate and cannot be distinguished between tuberculate and non-tuberculate populations. Photo credit: Yi-Kai Tea.

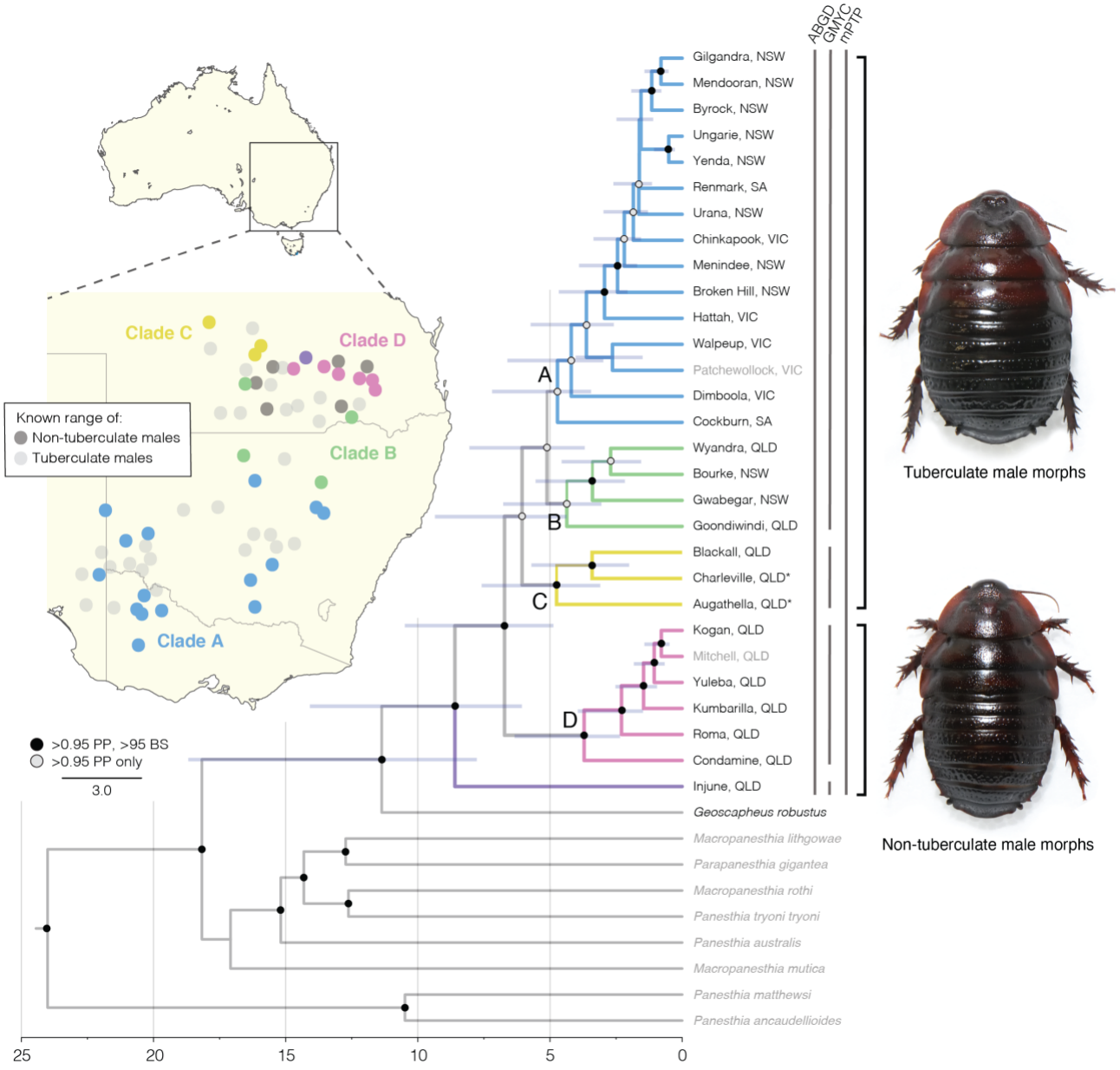


Fig. 2: fossil-calibrated phylogeny of *Geoscapheus dilatatus* male morphs inferred using BEAST2 and RAxML from whole mitogenomes. "Atypical" tuberculate male morphs are denoted by asterisks. The result of our three species delimitation analyses are shown to the right of tips. Taxa in grey were not sequenced in this study and were retrieved from Beasley-Hall *et al.* (in prep). Distant outgroups used for node calibration are not shown. Node support symbols refer to both our Bayesian and ML analyses; PP = posterior probability, BS = bootstrap support. Node bars denote 95% highest posterior density (HPD) values of divergence times at the respective node. Timescale is shown in millions of years and the scale bar denotes substitutions/site/My. Photo credit: Yi-Kai Tea.

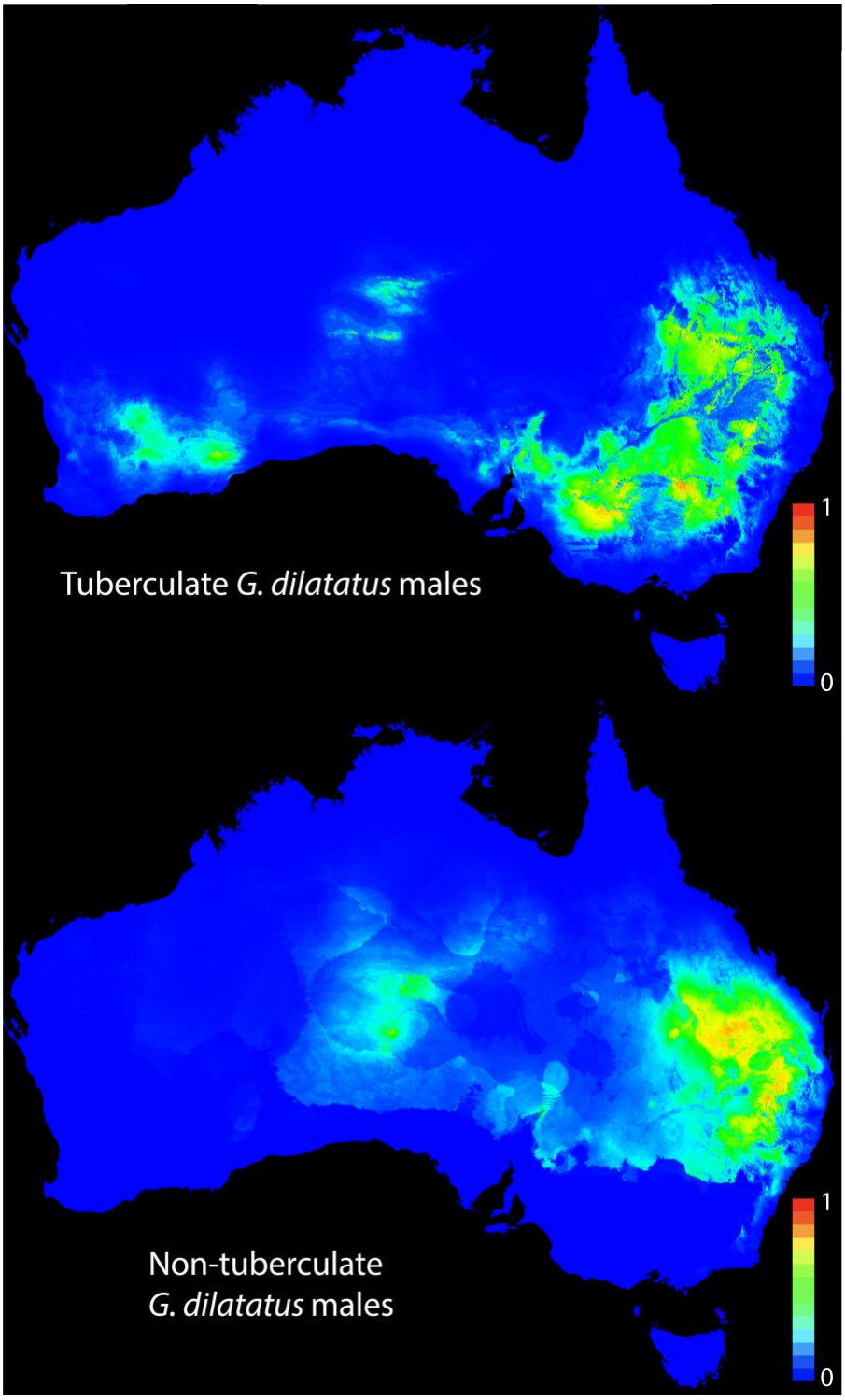


Fig. 3: environmental niche models of the two major male forms of *Geoscapheus dilatatus*. Red and blue in the heat maps represent a 100% and 0% probability, respectively, of occurrence within the constraints of the supplied abiotic variables. ENMs represent the predicted fundamental niche of taxa as opposed to the realised niche; other barriers or biological processes are likely to have existed that have prevented morphs from occurring in a given location that were not considered in the present study.

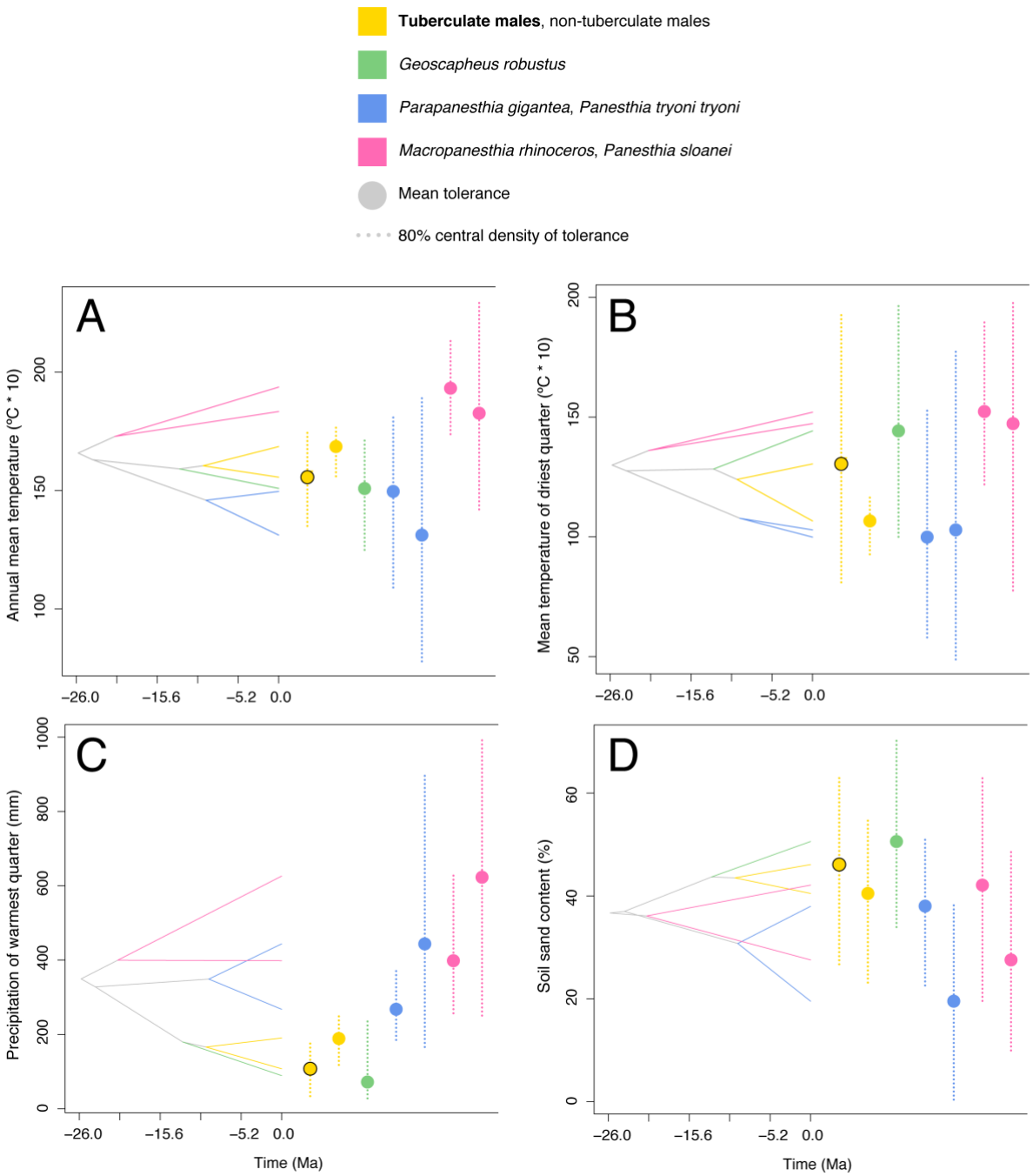


Fig. 4: ancestral niche reconstructions (ANRs) for select environmental variables of the two major male forms of *Geoscapheus dilatatus*, showing their predicted climatic tolerances compared to other members of the Geoscapheinae retrieved from Beasley-Hall *et al.* (2018). Phylogenetic trees are shown to the left of each reconstruction. Tips are positioned at the predicted mean tolerance of a given variable for each taxon, denoted by a circle within their 80% central density (range) of tolerance (dotted line). Tuberculate male forms show similarities with the mean tolerances of *G. robustus* (A, C) and are potentially able to tolerate hotter (B) and drier (C, D) environments overall with respect to their more restricted non-tuberculate counterparts. The entire set of 23 environmental variables tested here can be found in Fig. S5.

1

SUPPLEMENTARY MATERIAL

Table S1: environmental variables sourced from the BioClim and ASRIS databases used in environmental niche modelling (bold) and ancestral niche reconstructions (all variables) in this study.

Variable	Description
BIO01	Annual mean temperature (ºC)
BIO02	Mean diurnal temperature range (ºC)
BIO03	Isothermality (BIO02 / temperature annual range)
BIO04	Temperature seasonality (coefficient of variation)
BIO05	Max. temperature of warmest week (ºC)
BIO06	Min. temperature of coldest week (ºC)
BIO07	Temperature annual range (BIO05 - BIO06) (ºC)
BIO08	Mean temperature of wettest quarter (ºC)
BIO09	Mean temperature of driest quarter (ºC)
BIO10	Mean temperature of warmest quarter (ºC)
BIO11	Mean temperature of coldest quarter (ºC)
BIO12	Annual precipitation (mm)
BIO13	Precipitation of wettest week (mm)
BIO14	Precipitation of driest week (mm)
BIO15	Precipitation seasonality (coefficient of variation)
BIO16	Precipitation of wettest quarter (mm)
BIO17	Precipitation of driest quarter (mm)
BIO18	Precipitation of warmest quarter (mm)
BIO19	Precipitation of coldest quarter (mm)
Soil clay content	< 2 um mass fraction of < 2 mm soil material (%)
Soil sand content	20 um - 2 mm mass fraction of < 2 mm soil material (%)
Soil silt content	2-20 um mass fraction of < 2 mm soil material (%)
Soil bulk density	Bulk density of whole soil (including coarse fragments) in g/cm^3

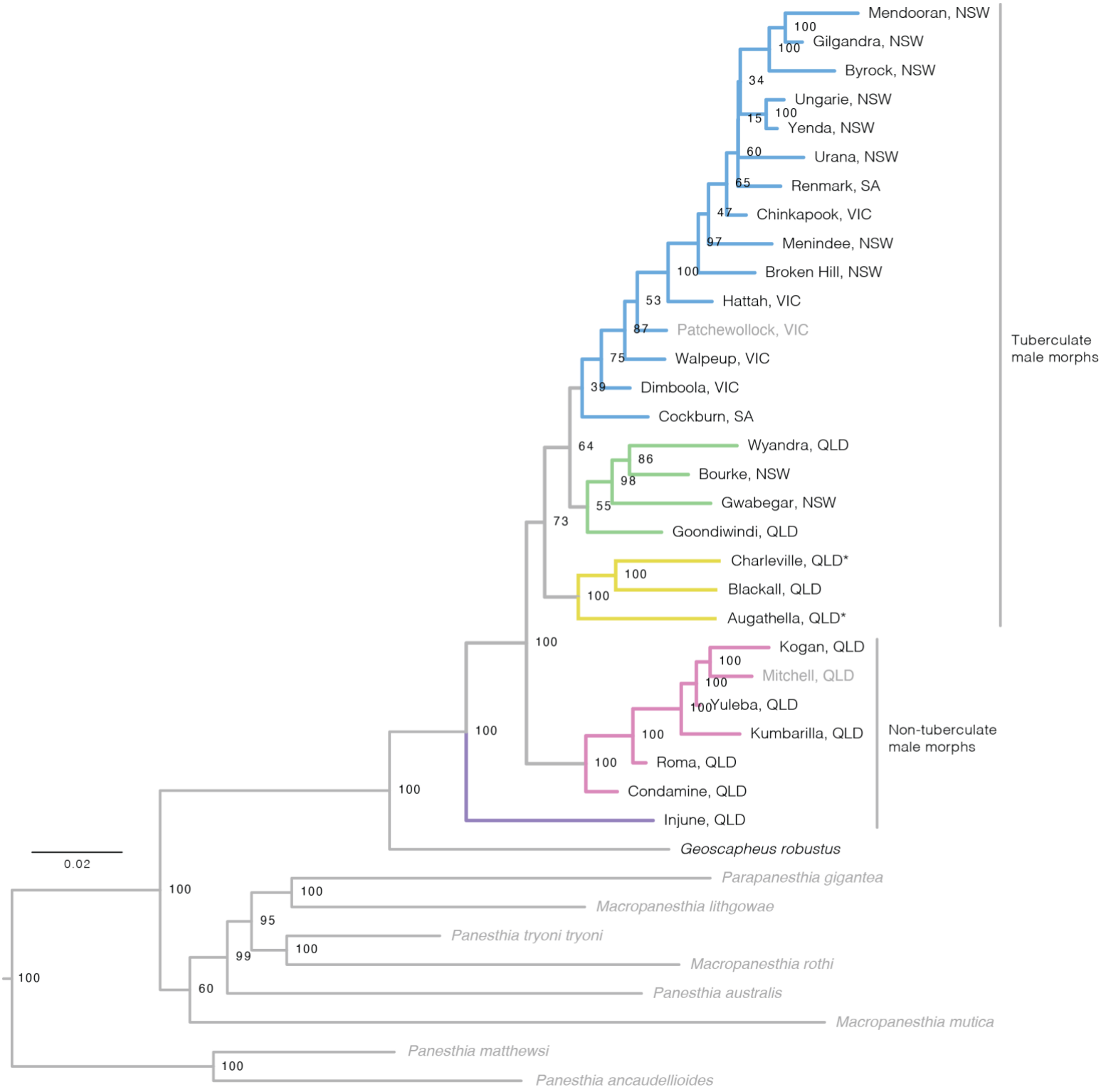


Figure S1: Phylogeny of *Geoscapheus dilatatus* male morphs inferred using RAxML with whole mitochondrial genomes. Branches and tip labels are colour-coded according to the scheme shown in Figure 2.

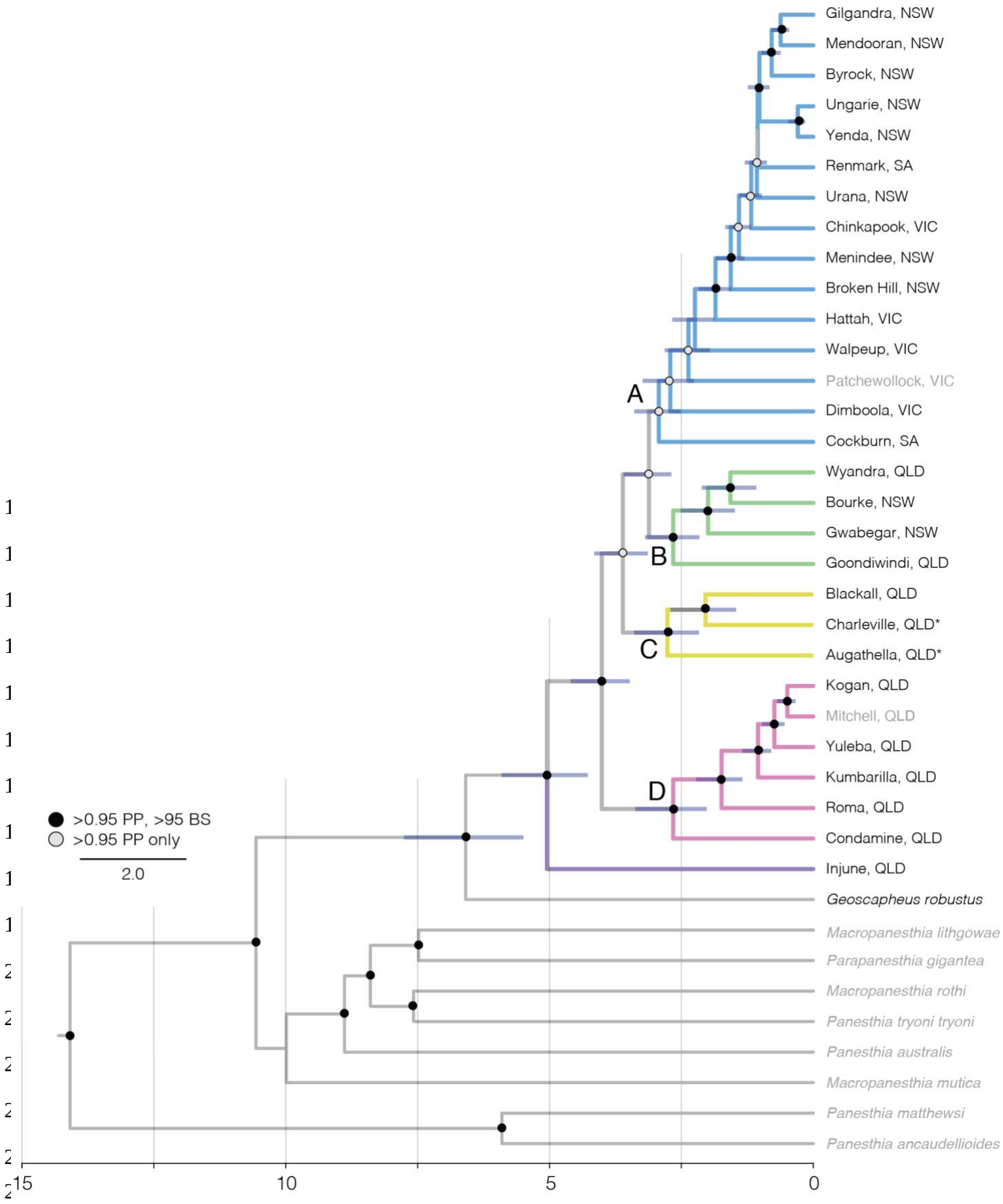


Figure S2: phylogeny of *Geoscapheus dilatatus* male morphs inferred using BEAST2 and RAxML from whole mitogenomes and uncalibrated orthopteran substitution rates. Timescale is shown in millions of years and the scale bar denotes substitutions/site/My. Branches and tip labels are coded according to the scheme shown in Figure 2.

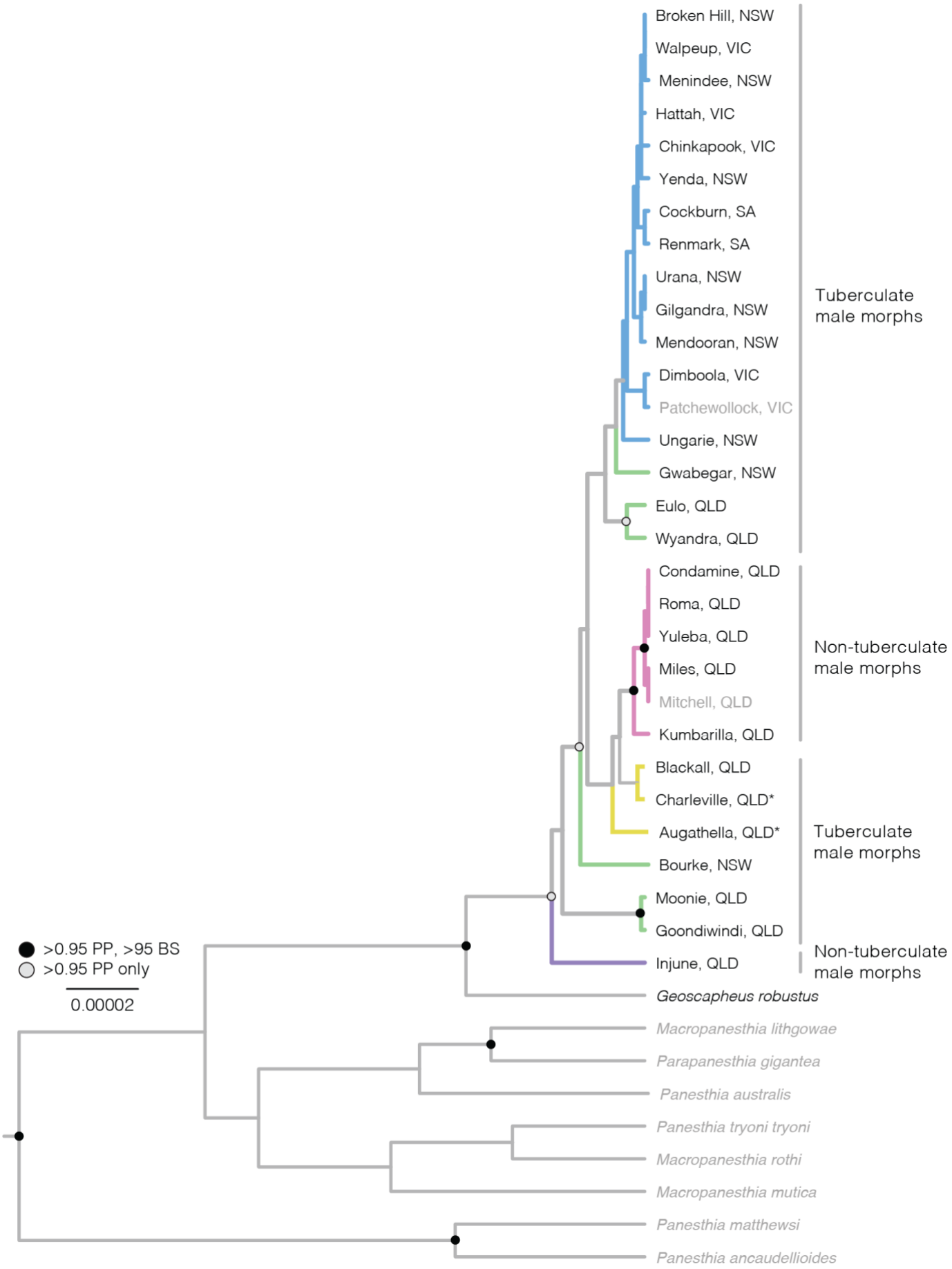


Figure S3: Phylogeny of *Geoscapheus dilatatus* male morphs inferred using BEAST and RAxML with nuclear markers *18S*, *ITS1*, *5.8S*, *ITS2*, and *28S*. Branches and tip labels are colour-coded according to the scheme shown in Figure 2.

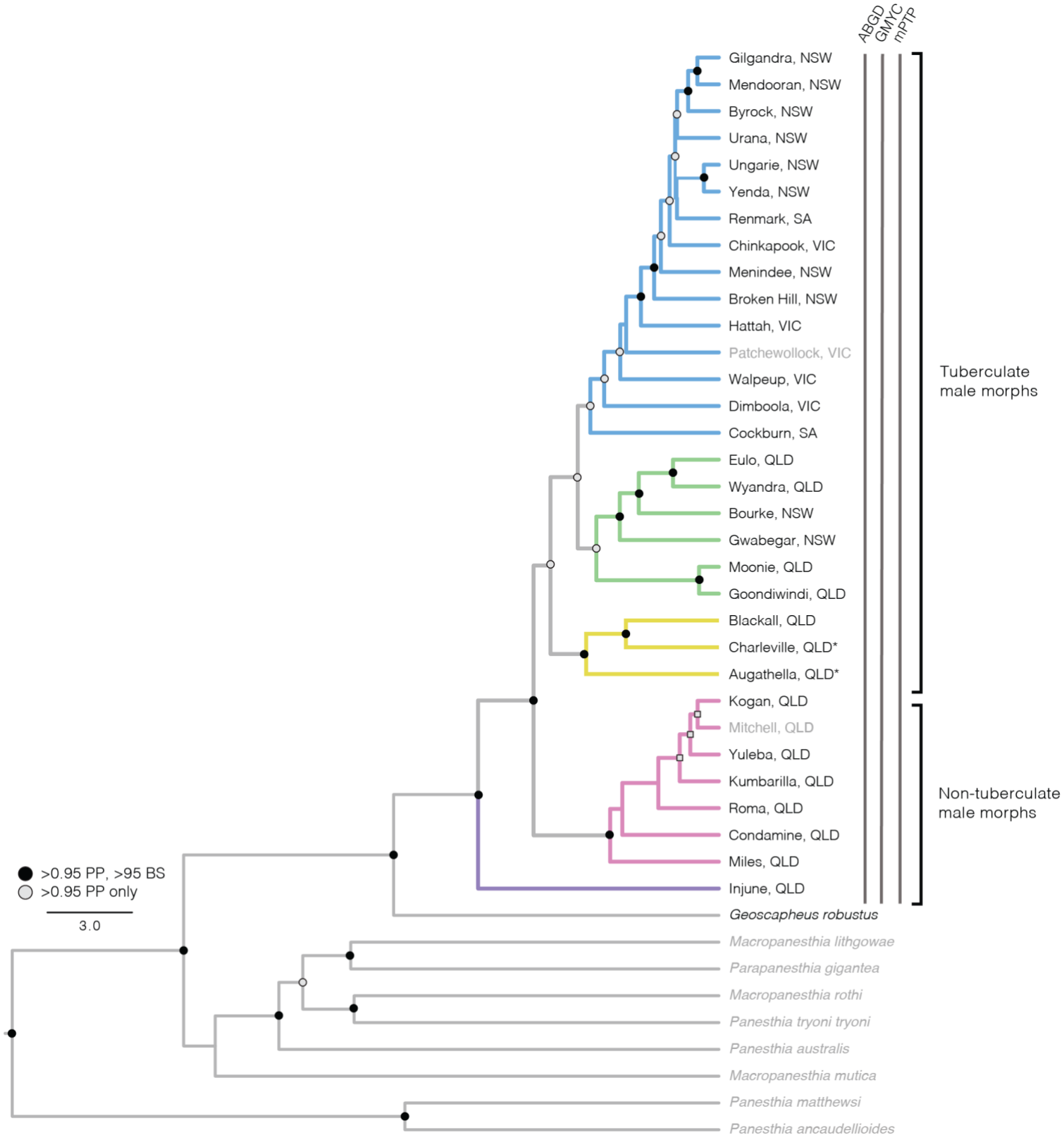
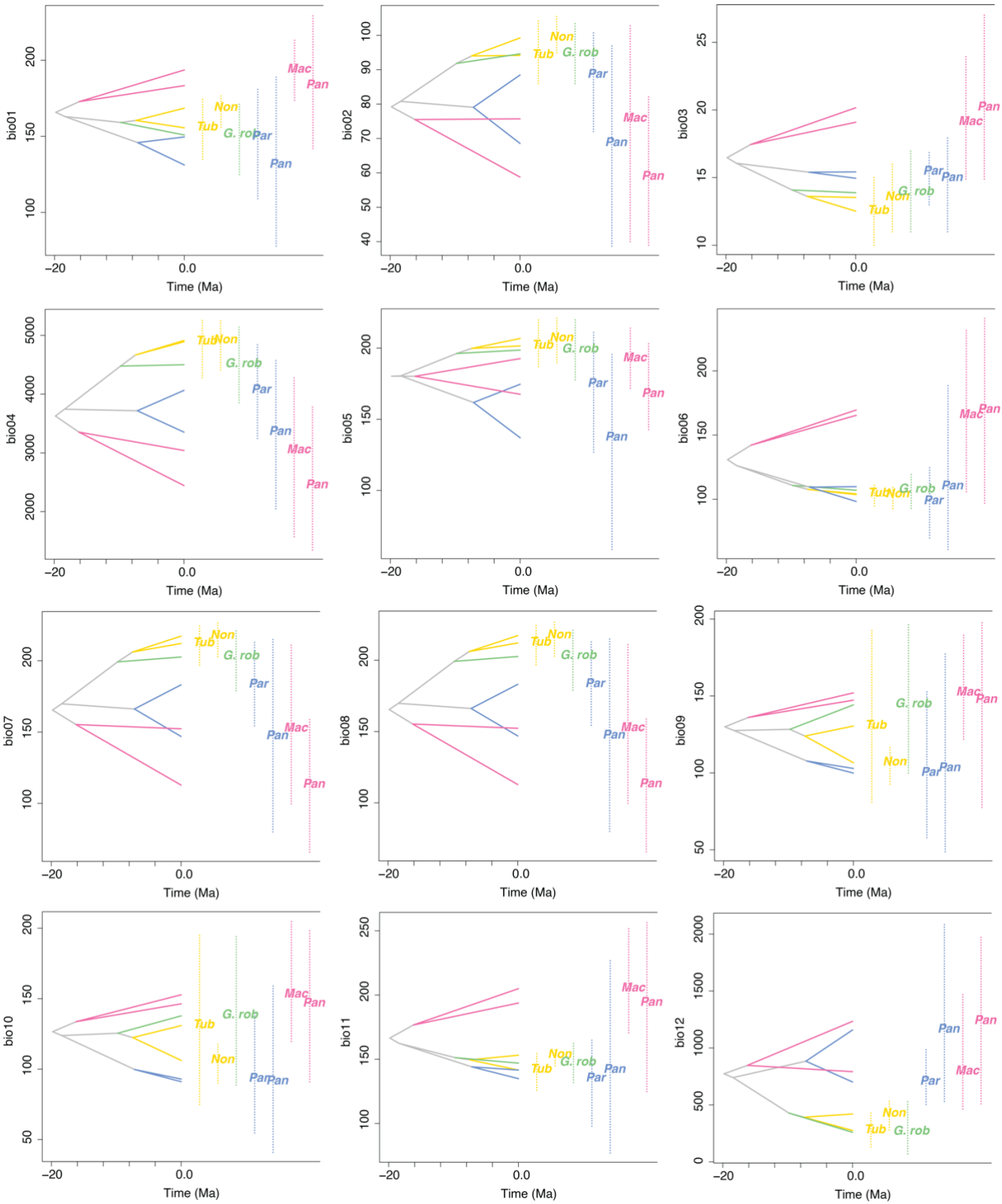
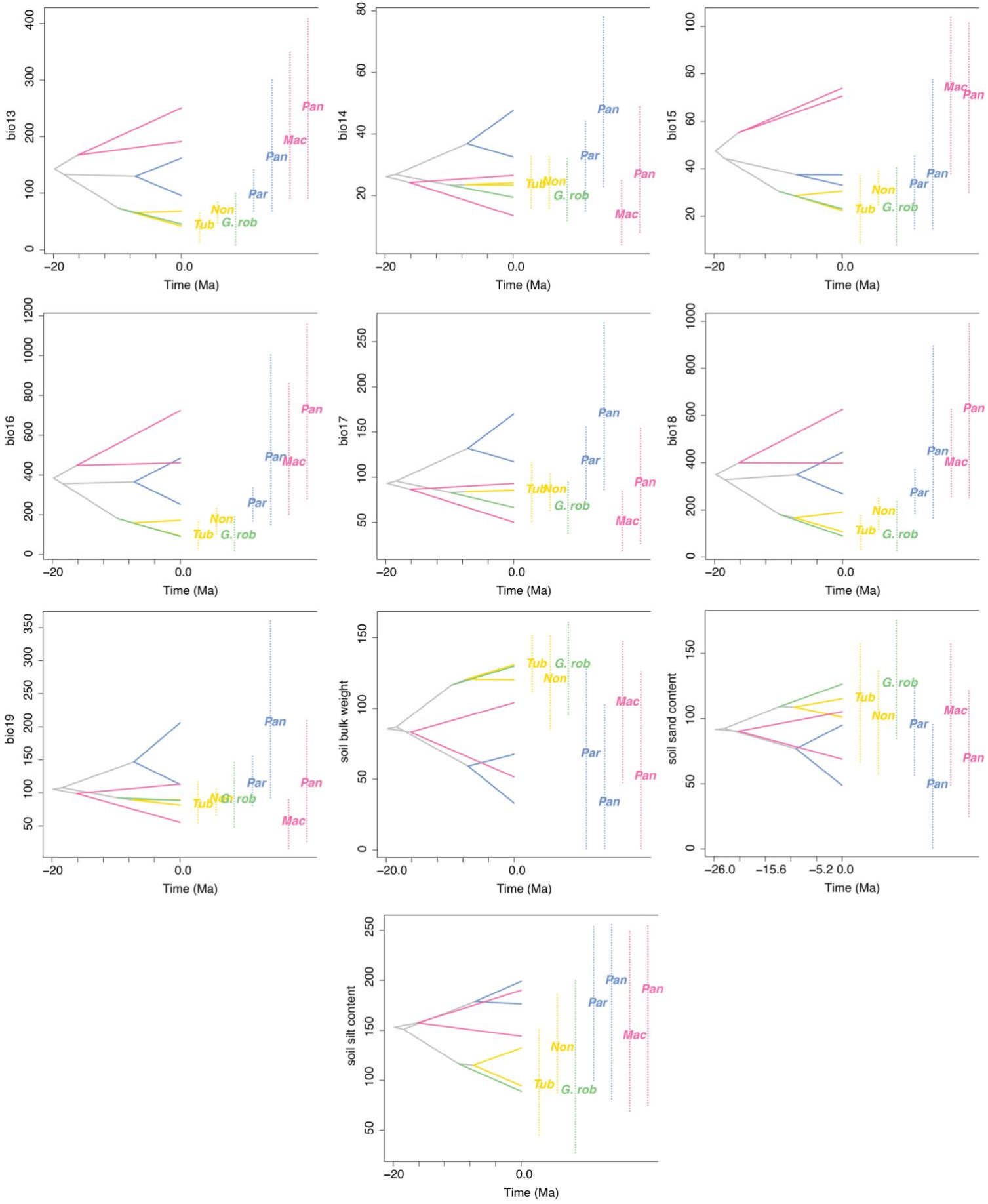


Figure S4: Phylogeny of *Geoscapheus dilatatus* male morphs inferred using BEAST and RAxML with a combined dataset consisting of whole mitochondrial genomes and nuclear markers *18S*, *ITS1*, *5.8S*, *ITS2*, and *28S*. Results of a species delimitation analysis based on this combined dataset are shown to the right of tips. Branches and tip labels are colour-coded according to the scheme shown in Figure 2.



35 **Figure S5:** Ancestral niche reconstructions performed using the *phyloclim* package in R on all 23
36 environmental variables listed in Table S1. Colours and the order of taxa follow the legend used in Figure 4.
37 Y-axis labels refer to environmental variables in Table S1.



37 **Figure S5:** ancestral niche reconstructions continued.

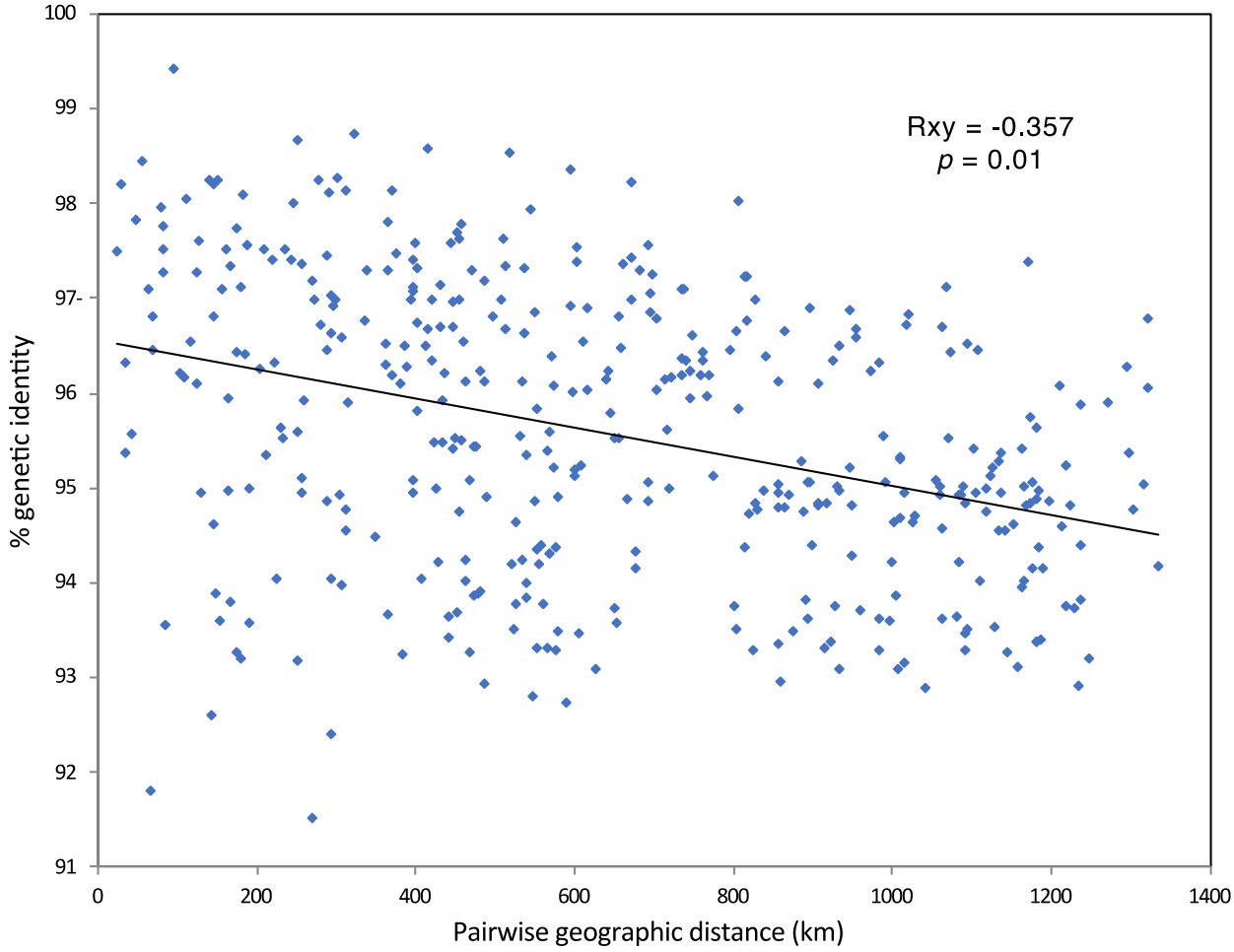


Figure S6: Mantel test results from GenAlEx comparing the % genetic identity between *G. dilatatus* samples and their geographic distance from one another. Under a scenario of isolation-by-distance, pairwise genetic distance would be expected to decrease as genetic identity increases, as shown here.