Genetic analyses indicate that the Australian endemic scorpion *Urodacus yaschenkoi* (Scorpiones: Urodacidae) is a species complex (#12305)

First submission

Please read the **Important notes** below, and the **Review guidance** on the next page.
When ready **submit online**. The manuscript starts on page 3.

Important notes

Editor and deadline

Jane Hughes / 24 Aug 2016

Files	1 Other file(s) Please visit the overview page to **download and review** the files not included in this review pdf.
Declarations	**One or more DNA sequences were reported.**

Review guidelines

Please in full read before you begin

How to review

When ready **submit your review online**. The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this **pdf** and upload it as part of your review

To finish, enter your editorial recommendation (accept, revise or reject) and submit.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to **PeerJ standard**, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (See **PeerJ policy**).

EXPERIMENTAL DESIGN

- Original primary research within **Scope of the journal**.
- Research question well defined, relevant & meaningful. It is stated how research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- Data is robust, statistically sound, & controlled.
- Conclusion well stated, linked to original research question & limited to supporting results.
- Speculation is welcome, but should be identified as such.

The above is the editorial criteria summary. To view in full visit **https://peerj.com/about/editorial-criteria/**

Genetic analyses indicate that the Australian endemic scorpion *Urodacus yaschenkoi* (Scorpiones: Urodacidae) is a species complex

Karen Luna-Ramirez [1], **Adam D Miller** [2], **Gordana Rašić** Corresp. [3]

[1] Projektgruppe "Bioressourcen", Fraunhofer-Institut für Molekularbiologie und Angewandte Oekologie IME, Gießen, Hessen, Germany
[2] Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Burwood, Victoria, Australia
[3] School of BioSciences, Bio21 Institute, The University of Melbourne, Parkville, Victoria, Australia

Corresponding Author: Gordana Rašić
Email address: gordana.rasic@unimelb.edu.au

Background. The native Australian desert scorpion *Urodacus yaschenkoi*, widely distributed throughout arid zones of the continent, is a promising model organism in biomedical research due to the chemical nature of its venom. Unlike their overseas counterparts, Australian scorpions have received little attention from researchers, and our study provides the first molecular insights into the phylogenetic patterns and history of *U. yaschenkoi*. **Methods.** We employed Bayesian Inference (BI) methods for the phylogenetic reconstructions and divergence dating among lineages, using unique haplotype sequences from two mitochondrial loci (*COXI*, 16*S*) and one nuclear locus (28*S*). We also implemented two DNA taxonomy approaches (GMYC and PTP/dPTP) to evaluate the presence of cryptic species. Linear Discriminant Analysis was used to test whether the linear combination of 21 variables (ratios of morphological measurements) can predict individual's membership to a putative species. **Results**. Multiple lines of evidence suggest that *U. yaschenkoi* is a species complex. High statistical support for the monophyly of several divergent lineages was found both at the mitochondrial loci and at a nuclear locus. The extent of mitochondrial divergence between these lineages exceeds estimates of interspecific divergence reported for other scorpion groups. The GMYC model identified nine entities and the PTP/bPTP approach identified seven, each representing putative species. Ratios of several traits that approximate body shape had a strong predictive power (83–100%) in discriminating two major molecular lineages. A time-calibrated phylogeny dates the early divergence at the onset of continental-wide aridification in late Miocene, with finer-scale phylogeographic patterns emerging during the Pleistocene. These patterns of phylogeograaphic structuring are consistent with the evolutionary history of other arid Australian biota. **Discussion.** Our results indicate that the taxonomic status of *U. yaschenkoi* requires revision. Evidence of a species complex needs to be considered for

future biomedical research, particularly when venom from multiple individuals is pooled for toxinological characterization. The complex evolutionary history of this scorpion highlights the importance of conserving populations from different Australian arid zones in order to preserve patterns of endemism and evolutionary potential.

Title: Genetic analyses indicate that the Australian endemic scorpion *Urodacus yaschenkoi* (Scorpiones: Urodacidae) is a species complex

Authors: Luna-Ramírez K[a], Miller AD[b,c], Rašić G[b*]

[a] Museum Victoria, 11 Nicholson St., Carlton Gardens, Melbourne, VIC 3053, Australia. (Current affiliation: Fraunhofer-Institut für Molekularbiologie und Angewandte Oekologie IME, Projektgruppe "Bioressourcen" Heinrich-Buff-Ring 58/62, Gießen 35392; Germany).

[b]Pest and Environmental Adaptation Research Group, School of BioSciences, The University of Melbourne, Victoria 3010, Australia.

[c]Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Victoria 3280, Australia.

[*]Corresponding author: Gordana Rašić, Pest and Disease Vector Group, School of BioSciences, The University of Melbourne, Victoria 3010, Australia;

phone number: +61 3 9035 5237, fax number: +61 3 8344 2279, Email: gordana.rasic@unimelb.edu.au

Running title: Scorpion *Urodacus yaschenkoi* is a species complex

Abstract

Background. The native Australian desert scorpion *Urodacus yaschenkoi*, widely distributed throughout arid zones of the continent, is a promising model organism in biomedical research due to the chemical nature of its venom. Unlike their overseas counterparts, Australian scorpions have received little attention from researchers, and our study provides the first molecular insights into the phylogenetic patterns and history of *U. yaschenkoi*.

Methods. We employed Bayesian Inference (BI) methods for the phylogenetic reconstructions and divergence dating among lineages, using unique haplotype sequences from two mitochondrial loci (*COXI*, 16*S*) and one nuclear locus (28*S*). We also implemented two DNA taxonomy approaches (GMYC and PTP/dPTP) to evaluate the presence of cryptic species. Linear Discriminant Analysis was used to test whether the linear combination of 21 variables (ratios of morphological measurements) can predict individual's membership to a putative species.

Results. Multiple lines of evidence suggest that *U. yaschenkoi* is a species complex. High statistical support for the monophyly of several divergent lineages was found both at the mitochondrial loci and at a nuclear locus. The extent of mitochondrial divergence between these lineages exceeds estimates of interspecific divergence reported for other scorpion groups. The GMYC model identified nine entities and the PTP/bPTP approach identified seven, each representing putative species. Ratios of several traits that approximate body shape had a strong predictive power (83–100%) in discriminating two major molecular lineages. A time-calibrated phylogeny dates the early divergence at the onset of continental-wide aridification in late Miocene, with finer-scale phylogeographic patterns emerging during the Pleistocene. These

patterns of phylogeograophic structuring are consistent with the evolutionary history of other arid Australian biota.

Discussion. Our results indicate that the taxonomic status of *U. yaschenkoi* requires revision. Evidence of a species complex needs to be considered for future biomedical research, particularly when venom from multiple individuals is pooled for toxinological characterization. The complex evolutionary history of this scorpion highlights the importance of conserving populations from different Australian arid zones in order to preserve patterns of endemism and evolutionary potential.

Keywords: Australian scorpion, cryptic species, phylogeography, COXI , 16*S* and 28*S* rRNA, morphology.

Introduction

Scorpions are among the most ancient arthropods, derived from Silurian ancestors (Dunlop et al., 2008) and are considered 'living fossils' due to a largely unchanged body plan since the Paleozoic period (Cao et al. 2013; Regier et al. 2010). Extant scorpions inhabit a diversity of terrestrial habitats across all continents except Antarctica, with the greatest species diversity found in tropical and subtropical regions of the world (Lourenço, 2001). Their populations can be sensitive to environmental changes due to a low reproductive rate (long generation time, long gestation time, small litter size) and high mortality of immature females (Lourenço and Cuellar, 1995; Fet et al., 1998). Several species have gained threatened status due to over-harvesting for the souvenir and exotic pet trades (CITES, Appendix II, http://www.cites.org/eng/app/appendices.php).

Scorpions have received widespread public attention due to their venomous nature, with the lethality (LD_{50}) of some scorpion venoms considered to be comparable to that of snake venoms (Oukkache et al., 2014; Chippaux and Goyffon, 2008). Despite the high annual incidence of scorpion-induced fatalities in regions of the Americas, Africa and the Middle East (Cupo 2015; Dabo et al. 2011; de Roodt 2014; Dehghani & Fathi 2012), the venom of most scorpion species is considered harmless (Isbister & Bawaskar 2014). To date 30 species, most belonging to the Buthidae family, are considered medically important (Chippaux and Goyffon, 2008), and the complex mixture of molecular constituents in venoms is increasingly recognized as an important source of new therapeutic and insecticidal agents (Possani et al., 2000; Gurevitz et al., 2007; Rodríguez de la Vega et al., 2010). Despite the medical and ecological relevance of scorpions, we have only a limited understanding of their phylogenetic and taxonomic relationships.

Australia has more than 40 described scorpion species organized into four families: Buthidae, Bothriuridae, Urodacidae and Hormuridae (Koch, 1977; Volschenk et al., 2008; Monod and Prendini 2015). The Urodacidae is an Australian endemic family found across the continent except on the south-eastern seaboard. The family was first described by Koch (1977) that under the current classification includes two genera: *Urodacus* and the recently described *Aops* (Volschenk and Prendini, 2008). The genus *Urodacus* contains 20 species (Prendini, 2000; Volschenk and Prendini, 2008), several of which have recently been described based on morphological characters (Volschenk and Prendini, 2008; Volschenk et al., 2012).

Urodacus yaschenkoi (Birula 1903), commonly known as the inland robust scorpion, occupies Australian desert habitats stretching from north-western Victoria through South Australia and across to Western Australia (Fig. 1). The species has emerged as a model organism in toxinology because it produces large volumes of venom compared with other *Urodacus* species (Luna-Ramírez et al., 2013; Luna-Ramírez et al., 2014), and is abundant across a broad geographic area. The biochemical and molecular characterization of venom for biomedical research generally requires large quantities of starting material, therefore venom is usually extracted from multiple individuals and pooled prior to analysis. To satisfy requirements for analytical procedures and potential drug development, it is necessary that pooled samples come from conspecific individuals. Consequently it is imperative that taxonomic status of *U. yaschenkoi* is well supported. This scorpion has had several synonyms throughout its taxonomic history, starting from the original description as *Hemihoplopus yaschenkoi* (Birula 1903), followed by *Urodacus granifrons* (Kraepelin 1916), *U. fossor* (Kraepelin 1916), and *U. kraepelini* (Glauert 1963), and finally by *U. yaschenkoi* (Koch 1977) (https://biodiversity.org.au/afd/taxa/Urodacus_yaschenkoi).

Here we use a combination of molecular and morphological analytical approaches to investigate the taxonomic status of *U. yaschenkoi* across its native range. Specifically, phylogenetic reconstructions were carried out using DNA sequence data from mitochondrial and nuclear loci, and were complemented with the analysis of several morphological characters related to body shape. We also examined the phylogeographic patterns in *U. yaschenkoi* lineages to gain insights into the evolution history of the group and the influence of historical environmental factors on shaping its distribution. The novel results also provide critical information for future toxinological research and conservation of this scorpion.

Materials and Methods

Biological material

Samples of *U. yaschenkoi* were obtained from field and museum collections (Table 1). Live specimens were collected from eight locations (approximately 500 m^2) in the semi-arid and arid regions of Central Australia in December 2010 and October 2011 (Table 1 and Fig. 1). Individuals were collected at night from pitfall traps set in front of their burrows, and those outside their burrows were detected using ultraviolet (UV) lamps that reveal soluble fluorescent components (β-carboniles) in the scorpion exoskeleton (Stachel et al., 1999). Captured scorpions were kept alive and transported to the laboratory for morphological identification according to Koch (1977). Key diagnostic feature that distinguishes *U. yaschenkoi* from other *Urodacus* species is a very small terminal prolateral tarsus unguis. All specimens were handled according to good animal practices defined by the Government of Australia, and all institutions and museums involved approved the animal handling work. Scorpions were anaesthetized by cooling in a refrigerator (4°C) for 5 min before removing ~1 mm^2 of leg muscle tissue, which was stored

123 in 90% ethanol at 4°C or –20°C for subsequent DNA extraction. Additional samples were
124 obtained from collections at the South Australian Museum (SAM) and Western Australian
125 Museum (WAM) containing specimens collected between 2000 and 2010 (Table 1).

126 ***DNA extraction, amplification and sequencing***

127 Total DNA was extracted from the stored muscle tissue using the DNeasy Blood and Tissue Kit
128 (Qiagen, Venlo, Netherlands) following the manufacturer's instructions. Two mitochondrial loci
129 (cytochrome oxidase subunit I, *COXI*; large ribosomal subunit, *16S*) and a single nuclear locus
130 (*28S*) were amplified by PCR with a reaction volume of 20 µl containing 0.5 ng of template
131 DNA, 10 µl of Go Taq Master Mix (Promega, Madison, Wisconsin, USA), 0.5 µl of 10 nM
132 primers and 7 µl of RNase-free water (Qiagen). The primer sequences and PCR amplicon sizes
133 are summarized in Table 2.

134 Primers previously designed for the insect *COXI* gene (Simon et al., 1994; Tanaka et al., 2001)
135 were used to amplify a 630-base pair (bp) fragment from the 3' end of the locus. The
136 amplification conditions comprised an initial denaturing step at 95°C for 5 min followed by 35
137 cycles of denaturing at 94°C for 30 s, annealing at 52°C for 40 s, and extension at 72°C for 45 s,
138 and a final extension phase at 72°C for 5 min. For the mitochondrial *16S* gene, the scorpion-
139 specific primer pairs modified by (Gantenbein et al. 2005) were used to amplify a 425-bp region
140 at the 3' end of the locus. The amplification conditions comprised an initial denaturing step at
141 94°C for 4 min followed by 30 cycles of denaturing at 94°C for 30 s, annealing at 47.5°C for 30
142 s, and extension at 72°C for 30 s, and a final extension phase at 72°C for 7 min. The *COXI* and
143 *16S* gene fragments were also amplified from three specimens keyed out as *Urodacus manicatus*
144 and *U. novaehollandiae* (Table 1). Sequences from these taxa were used as outgroups in
145 downstream phylogenetic reconstruction. Primer pairs R1S and R1AS, and R2S and R2AS,

designed by (Arabi et al., 2012), were used to amplify 1158-bp and 1246-bp fragments of the *28S* locus, respectively. Each set of primers amplifies a different region of the gene, which overlaps by 327 bp, and their sequences were concatenated to form a larger product of 2076 bp. The amplification conditions for both sets of primers comprised an initial denaturing step at 94°C for 4 min, followed by 30 cycles of denaturing at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s, and a final extension phase at 72°C for 7 min.

Museum specimens that were not stored under ideal conditions for preservation failed to yield *COXI* amplicons suitable for direct sequencing. To address this issue, additional PCR primers were designed to amplify smaller fragments for *COXI* locus (Table 2), resulting in amplicons of 150 bp that were used for subsequent analysis. For the SAM specimens, the amplification of the 28*S* nuclear gene failed entirely and these samples were excluded from further analysis. All amplicons were sequenced in both directions using the PCR amplification primers, and carried out on an Applied Biosystems 3130 genetic analyzer by Macrogen Inc. (Seoul, South Korea).

Sequences were aligned and edited in Geneious Pro v6.1 (Biomatters Ltd) using the MUSCLE alignment option with default parameters. All chromatograms were checked for the presence of multiple peaks (which indicate heterozygosity), and authenticity of the *COX1* coding gene was validated by checking for indels and premature stop codons. After this editing process, the alignment of the mitochondrial gene fragments yielded 616-bp and 396-bp products for the *COX1* and *16S* genes respectively, and the final 28*S* alignment was 2076 bp in length. The final dataset contained 68 sequences for each of the mitochondrial genes and 27 sequences for the 28*S* locus (Table 1, [GenBank: KP176717-KP176786]). Shared haplotypes were identified and the uncorrected pairwise genetic distances (%) were calculated using Geneious Pro v6.1 (Biomatters

Ltd). This simple distance measure was implemented to achieve reliable estimates of both intraspecific and interspecific genetic variation.

Phylogenetic analysis and divergence time estimation

Phylogenetic reconstructions and divergence dates among lineages were calculated using unique haplotypes and Bayesian Inference (BI) methods implemented in BEAST v2.1.3 (Bouckaert et al., 2014). We used jModeltest v0.1.1 (Posada, 2008) to select the best-fit model of evolution, based on Akaike Information Criteria (AIC) (Akaike and Company, 1981) for each of the mitochondrial and nuclear genes (GTR + G in each case). Mitochondrial loci were combined for analysis due to their similar modes of evolution (GTR+R), as indicated by the incongruence-length difference (ILD) tests (Farris et al. 1995) implemented in PAUP_4.0b10 (Swofford 2002). In contrast, the nuclear gene (28*S*) was analyzed independently and was not combined with the mitochondrial dataset for phylogenetic genetic analysis due to inconsistencies in taxon sampling (Table 1).

Operators were auto-optimized, and five independent Markov Chain Monte Carlo (MCMC) runs were performed using a Yule (speciation) tree-prior, each running for 5×10^6 generations, sampling every 10,000 states. Log files were examined with Tracer v1.5 (Drummond and Rambaut, 2007) to ensure that runs were sampling from the same posterior distribution, to determine appropriate burn-in, and to ensure that effective sample sizes (ESSs) of parameters of interest were greater than 1000. Tree files of independent runs were then combined using LogCombiner v2.1.3 (Drummond et al., 2012), discarding the first 20% and re-sampling at a lower frequency of 15,000. The maximum clade credibility (MCC) tree was recovered from a sample of 10,000 posterior trees, and branch support was annotated using TreeAnnotator v2.1.3 (Drummond et al., 2012). Each analysis started with a random starting tree and seed with no root

specified. Sequence data from species of the same genus (*U. manicatus* and *U. novaehollandiae*) were used to estimate the root of the mitochondrial gene tree.

The mitochondrial gene tree was time calibrated with divergence times of nodes inferred from 95% highest posterior density (HPD) intervals. Scorpion-specific mutation rates of 0.007 substitutions/site/million years for *COXI* (Gantenbein and Largiadèr, 2003) and 0.005 substitutions/site/million years for 16*S* (Gantenbein et al., 2005) were used to calibrate the tree. These estimates are derived from Buthid scorpions and have been used to estimate divergence times among various scorpion lineages including non-Buthid taxa (Graham et al., 2012; Bryson et al., 2013a,b). Substitution rates were set in BEAUti v1.7.3 (Drummond et al., 2012) using relaxed clock log normal priors. Tracer was then used to obtain parameter estimates for time to most recent common ancestor (tMRCAs) for nodes within the gene tree.

Additional phylogenetic constructions were also performed using a truncated *COXI* alignment to test the influence of missing data on the final tree topology. Because numerous museum collections yielded short *COXI* gene products, we trimmed the alignment to 150-bp to exclude regions of the alignment with high levels of missing data. This exercise demonstrated that the inclusion/exclusion of missing data had little influence on the phylogenetic reconstructions. Consequently, all results presented from this point reflect those from the non-truncated *COX1* alignment.

Species delineation based on molecular data

We implemented two DNA taxonomy approaches to evaluate the presence of cryptic species. First, the general mixed Yule coalescent (GMYC) approach (Pons et al., 2006; Fujisawa and

213 Barraclough, 2013) was applied to an ultrametric tree (produced using BEAST) in R v2.15.3 (R
214 Development Core Team 2008) with the Splits package (http://splits.r-forge.r-project.org). The
215 GMYC model is a process-based approach that detects the threshold in a gene tree at which
216 within-species processes (i.e. coalescence) shift to between-species processes (i.e. speciation and
217 extinction). Second, we combined the Poisson Tree Processes model for species delimitation
218 (PTP) and a Bayesian implementation of PTP (bPTP) to infer putative species boundaries on a
219 given phylogenetic input tree (Zhang et al., 2013). The PTP/bPTP model, unlike the GMYC
220 model, requires a bifurcated phylogenetic tree rather than an ultrametric tree. PTP/dPTP models
221 speciation or branching events in terms of the number of substitutions. The following parameters
222 were used: MCMC, 500,000 generations; thinning, 100; burn-in, 0.1; seed, 123, and assessed
223 convergence in each case to ensure the reliability of the results.

224

225 ***Analysis of demographic history***

226 The demographic history of *U. yaschenkoi* was estimated from the mitochondrial dataset using a
227 Bayesian Markov-Chain Monte Carlo (MCMC) coalescent approach implemented in BEAST
228 v2.1.3 (Bouckaert et al., 2014). The Bayesian skyline plot uses MCMC sampling procedures to
229 estimate a posterior distribution of effective population size through time from a sample of gene
230 sequences, given a previously specified nucleotide substitution model (Drummond et al. 2005).
231 The time dimension of the analyses was calibrated using substitution rates for the respective gene
232 partitions described above. The prior on rate was set to follow a normal distribution allowing for
233 uncertainty around the estimate. Analyses were run using a best-fit model of evolution (GTR+R)
234 and 100 million MCMC generations sampled every 10000 generations and launched from a
235 random starting tree. The analysis was repeated in triplicate and log files were examined using

Tracer ver. 1.5 (Drummond and Rambaut, 2007b) to determine the appropriate burn-in (10% of chain length), to ensure that runs were returning samples from the same distribution, and to ensure that the effective sample sizes for all demographic statistics were greater than 1000. Post-burn-in log and tree files from each independent run were combined using LogCombiner ver. 1.7.3 and the trees in the posterior sample were summarized by Bayesian skyline reconstruction using a stepwise skyline variant.

Analyses of morphological traits

Traits related to body size and shape were assessed in 39 female adult specimens that were keyed out as *U. yaschenkoi* (according to Koch, 1977) collected at 26 locations (Table 1, Fig. 1). Gender was determined by examining the genital opercula of adult scorpions, with males having a small finger-like projection known as the genital papilla. Because our collection contained only three males, the analyses were done only with females.

The following traits were measured under a microscope using an ocular ruler with 1-mm precision: carapace length (CL), metasoma segment V length (MVL), telson length (SL), pedipalp length (PL), chela length (ChL), pecten length (PecL) and pecten width (PecW). Ratios of traits (e.g. CL/MVL, SL/PL etc.) gave in total 21 variables scored in each individual (Supplemental file 4). These variables were treated as predictors in the Linear Discriminant Analysis (LDA) implemented in the R package "MASS" (Venables & Ripley 2002). LDA was used to test whether the linear combination of 21 variables (ratios of morphological measurements) can predict individual's membership to a mitochondrial lineage (putative species). Strong predictive power of morphological variation on the observed molecular divergence would provide additional support for a species complex in *U. yaschenkoi*.

259

260 ## Results

261 We identified 31 unique mitochondrial haplotypes with uncorrected distances between
262 haplotypes ranging from 0.3–7.6% (mean ± standard deviation = 3.0% ± 0.4%) and distances
263 from the outgroup taxa of 8.4–10.2% (mean ± standard deviation = 9.4% ± 1.4%) (Supplemental
264 File 1). A total of 13 nuclear 28*S* haplotypes were identified with uncorrected *p*-distances of 0.1–
265 0.5% (mean ± standard deviation = 0.2% ± 0.1%) (Supplemental File 2). A list of haplotypes for
266 sample locations is provided in Supplemental File 3.

267 ### *Phylogenetic analysis*

268 *Mitochondrial markers*

269 Bayesian inference analysis of the mitochondrial dataset identified several genetically divergent
270 lineages (three major lineages represented as black, red and green clades in Fig. 2), with strong
271 statistical support for their respective monophyl~~ies~~ (posterior probability >0.95). ~~Figure 1 shows~~
272 ~~that the~~ lineages within the black clade are broadly distributed across Victoria, South Australia
273 and Western Australia, whereas the red and green clades are restricted to Western Australia.
274 From this point forward we will refer to the black, red and green clades as the south-central,
275 western and central-western lineages, respectively.

276 The south-central lineage showed significant geographic structure. The most divergent ~~and basal~~
277 sub-lineage (haplotypes 16–17) was ~~found~~ in Western Australia, ~~occurring~~ in sympatry with the
278 central-western lineage (Fig. 1). A second well supported sub-lineage (haplotypes 1–14 and 18-
279 19) was found west of the Central Ranges, through to the Eyre Peninsula in South Australia,
280 while another (haplotypes 20–22, 25 and 28) had a distribution extending from the Central to Mt

Lofty Ranges in South Australia, and across to north-western Victoria. The remaining monophyletic sub-lineage within the south-central grouping (haplotypes 23–24 and 26–27) had a narrow north-south distribution in the central inland and coastal regions of South Australia (Fig. 1).

Mean uncorrected pairwise genetic distances between the three major clades (south-central, western and central-western) ranged from 6.4 to 6.9% (overall mean ± standard deviation = 6.6% ± 0.9%). The mean intra-clade distances ranged from 2.2% ± 0.4% and 0.8% ± 0.2%, respectively (not calculated for the western clade due to only a single recorded haplotype). Mean uncorrected distances between the three major clades and the outgroups ranged from 9.3 to 10.3% (mean ± standard deviation = 9.4% ± 1.4%).

Nuclear marker

Despite low level of variation in the 28*S* dataset, *Bayesian* analysis produced a nuclear gene topology that was largely concordant with the mitochondrial gene tree. Three genetically divergent clades were identified, corresponding to those from the mitochondrial dataset (represented by the black, red and green branch colours in Fig. 3). In each case, strong statistical support for the monophyly of each clade was found (posterior probability >0.95). ~~Due to low level of variation when compared to the mitochondrial dataset,~~ the interrelationships among lineages within each clade in the nuclear gene tree ~~were~~ unresolved, preventing any reliable inferences of phylogeographic patterns.

300

301 ***Divergence dating***

302 Our time calibrated mitochondrial phylogeny suggested that the split between the major *U.*
303 *yaschenkoi* clades (south-central, western and central-western lineages) occurred during the late
304 Miocene/early Pliocene (4–7 MYA) (Fig. 2). Lineage diversifications within the south-central
305 lineages appear to have occurred during the Pliocene and early Pleistocene (1.8–4 MYA), while
306 finer-scale phylogeographic patterning within the sub-lineages arose during the late Pleistocene
307 (<1 MYA). Divergence time estimates should be interpreted with some caution, as the nucleotide
308 substitution rate was derived from a different scorpion family (Buthidae) and there are large
309 errors margins around 95% HPD estimates.

310 ***Molecular-based species delineation***

311 Among the 31 unique mitochondrial haplotypes described above, the GMYC model identified
312 nine entities and the PTP/bPTP approach identified seven, each representing putative species
313 (Table 3). The assignment of haplotypes to putative species groups is shown in Fig. 2, where
314 conspecifics share a common symbol (star, hollow circle, solid circle, triangle, diamond, cross,
315 or square). Species assignments were highly consistent when comparing each of the methods, but
316 we presented the PTP/bPTP results as they are more accurate when the evolutionary distances
317 between lineages are small (Zhang et al., 2013). In summary the south-central, western and
318 central-western clades were recognised as putative species groups, as were the sub-lineages
319 within the south-central ancestral grouping (Fig. 1 and 2).

320 ***Analysis of demographic history***

Bayesian skyline analyses were performed for the south-central ancestral lineage only, as the western and central-western lineages are represented by a limited number of haplotypes. We also performed analyses on two independent datasets consisting of individuals assigned to the two sub-lineages denoted by stars and circles within the south-central ancestral lineage in Fig. 2. This allowed us to determine if potential interspecific influences are biasing estimates of demographic history in the combined south-central lineage dataset. All analyses indicate that *U. yaschenkoi* south-central lineage has remained relatively stable over the last 1 to 2 million years (Fig. 4A-C). Although the plot of mean *N*e indicates gradual growth through time to the present day for the sub-lineage in Fig. 4B, the 95 % highest posterior density intervals are large making it difficult to infer any notable growth. Each figure indicates a decline in *N*e in the last 100,000 years, yet this pattern should be regarded with caution, as it may simply be an artefact of uncharacterised population structure (a common artefact where mixed gene pools are co-analysed; (Heller et al. 2013).

Discriminant power of morphological variation

None of the *U. yaschenkoi* specimens that were characterized at 21 morphological ratio variables were assigned to the western mitochondrial clade, hence the LDA was done on 39 females assigned to the south-central and the central-western clades. Individuals were categorized into four groups (putative species) based on the results of the PTP/bPTP molecular species delineation analysis: 18 females from the group designated with a star, 12 from the circle group, three from the triangle group, and six from the solid square group (Fig. 2). Because our dataset contained four groups, we could find a maximum of three discriminant functions that separate these groups.

The first discriminant function (LD1) achieved 93.7% of the separation, reflecting the morphological distinction of the central-western (green) clade from the south-central (black) clade (Fig. 5). Further separation of the three putative groups within the south-central clade was weak (LD2-3, Fig. 5). We then grouped samples into two putative species (central-western and south-central clade) and tested the accuracy of prediction using 100 jackknife resampling steps. The grouping into two molecular clades based on morphological variation was 100% accurate (33/33) for the south-central clade and 83.3% accurate (5/6) for the central-western clade. Therefore, our results indicate strong predictive power of morphological variation on the observed molecular divergence, and suggest the existence of at least two distinct taxa within *U. yaschenkoi.*

Discussion

Taxonomy of Urodacus yaschenkoi

Multiple lines of evidence indicate that *U. yaschenkoi* is a species complex. High statistical support for the monophyly of three highly divergent lineages at both mitochondrial and nuclear loci was observed. The extent of genetic divergence between the three lineages (6.4–6.9%) exceeds estimates of interspecific divergence previously reported for other scorpion and arthropod groups (Wysocka et al., 2011; Tourinho et al., 2012; Bryson et al., 2014). Collectively, these observations satisfy the requirements for species delineation based on the principles of the phylogenetic species concept (Wheeler, 1999; De Queiroz, 2007). Because genetic 'yardstick' approaches provide crude taxonomic measures and nucleotide substitution rates often vary considerably between taxonomic groups, some caution is needed when considering findings of these analyses alone. Alternative DNA-based species delineation approaches (GMYC and bPTP)

366 also provided significant statistical support for the recognition of the south-central, western and
367 central-western *U. yaschenkoi* clades as distinct species, and potential further cryptic speciation
368 within the south-central clade (Fig. 2). The GMYC method has been criticized for over-splitting
369 species with a pronounced genetic structure (Satler et al., 2013), yet several recent studies have
370 shown that it is highly robust (Fujisawa and Barraclough, 2013; Talavera, Dincă *et al.* 2013).
371 Consistent results between the GMYC and bPTP methods strengthen the molecular evidence in
372 favor of *U. yaschenkoi* being a species complex.

373 We also demonstrated a strong association between the molecular divergence and morphological
374 variation. Namely, ratios of several traits that approximate body shape had a strong predictive
375 power (83-100%) in discriminating two major molecular lineages in *U. yaschenkoi:* the central-
376 western clade and the south-central clade. Until now *U. yaschenkoi* has been described as a
377 species with a wide geographic distribution and is distinguished from other congeneric species
378 by its much smaller terminal prolateral tarsal ungues and by the production of large amounts of
379 venom (Koch, 1977). Based on our results from a limited sample size, morphological variation in
380 this scorpion warrants further investigation. Attention to pigmentation, granulation and
381 burrowing behavior might provide additional diagnostic characters that could be used to describe
382 the *U. yaschenkoi* species complex.

383 ***Phylogeographic history of Urodacus yaschenkoi***

384 Our time-calibrated phylogeny indicates that the south-central, western and central-western *U.*
385 *yaschenkoi* lineages (Fig. 2 and 3) diverged from a common ancestor during the Miocene/early
386 Pliocene (approximately 5-9 MYA). This geological time was marked by a shift to a much drier
387 climate, the significant contraction of rainforests and the expansion of arid habitats (Martin,
388 2006). Time calibrated phylogenies for diverse taxa from Australian arid habitats also indicate

that their deep divergences occurred during the mid-Miocene and the onset of aridification (Byrne et al., 2008). Further diversification within the major ancestral *U. yaschenkoi* lineages appears to have occurred throughout the Pliocene (3-5 MYA), which was a consistently dry period. This is followed by further lineage divergence during the mid and late Pleistocene when the climate was highly dynamic (< 1 MYA), with wetter and drier episodes corresponding to interglacial and glacial cycles (Martin, 2006).

Koch (1977) suggested that the mechanisms underlying the distribution of extant Australian scorpions include a combination of allopatric diversification in refugia, range expansion and overlapping. The refugia concept has also been widely used to describe biogeographic patterning in a range of biota from Southern Australia, specifically relating to mid-Pleistocene climatic oscillations and increased aridity (Byrne, 2008). In *Urodacus,* the fragmentation of temperate forests in southern Australia during arid conditions of the late Cenozoic is likely to have driven the isolation and divergence between *U. novaehollandiae* (south-western species) and *U. manicatus* (south-eastern species) (Koch, 1977). In *U. yaschenkoi,* the contemporary distribution of lineages/putative species suggests some influence of the refugia described in Koch (1977), such as the expansion of the western lineages from the Hamersley refuge area, or the complex history of secondary contact between the south-central lineages expanding from Central Ranges, Eyre Peninsula / Mt Lofty, and Murray Darling River refugia (Fig. 1). However, a more comprehensive sampling is needed to further test these hypotheses.

Implications of findings

Urodacus yaschenkoi is a promising model organism in biomedical research, whose venom is increasingly used in toxinological studies (Luna-Ramírez et al. 2014; Luna-Ramírez et al. 2013). Because large quantities of venom are needed for biochemical and molecular characterization,

412 venom from multiple individuals is commonly pooled for such analyses. If specimens from
413 divergent taxa are used for venom pooling, this would likely result in a spurious identification of
414 bioactive compounds that do not naturally occur in a single taxon. Our study provides the much
415 needed resource for selecting compatible genotypes for future toxinological research.

416 Our findings also have implications for the conservation of. *U. yaschenkoi.* While currently
417 considered as a species of least concern, our study reveals the complex evolutionary history of
418 this scorpion and highlights the importance of conserving populations from different Australian
419 arid zones in order to preserve patterns of endemism and evolutionary potential.

420

421 ## Conclusions

422 Our study provides the first insight into the molecular phylogeny of the endemic Australian
423 scorpion *Urodacus yaschenkoi*. We found strong evidence for deep divergence and complex
424 phylogeographic patterns within this taxon. Concordance between the mitochondrial and nuclear
425 data, along with the morphological variation, all suggest that *U. yaschenkoi* is a species complex.
426 These findings have implications for the future of biomedical research, such as venom
427 characterization for drug and insecticide development. Careful consideration of *U. yaschenkoi*
428 operational taxonomic units will be necessary to achieve sustainable toxinology and conservation
429 of Australian desert biodiversity.

430

Competing interests

Authors declare no competing interests.

Authors' contribution

KLR conceived the study, collected the samples, performed the experimental work, processed the data and contributed to the manuscript writing. GR and ADM analysed the data and led the writing of the manuscript. All authors have read and approved the final manuscript.

Acknowledgments

KLR would like to thank Dr Ken Walker and Dr Joanna Sumner from Museum Victoria for their help and support during the experimental phase of this research project. KLR is grateful to Prof. Mark Harvey at the Western Australian Museum and to Dr. Leslie Chisholm at the South Australian Museum for agreeing to provide scorpion specimens for molecular and morphological characterization. KLR was supported by scholarships from CONACyT and from The Hugh Williamson Foundation, through Museum Victoria.

References

Akaike, H., Company, N.P., 1981. Likelihood of a model and information criteria. J. Econom. 16, 3–14. doi:10.1016/0304-4076(81)90071-3

Arabi, J., Judson, M.L.I., Deharveng, L., Lourenço, W.R., Cruaud, C., Hassanin, A., 2012. Nucleotide composition of CO1 sequences in Chelicerata (arthropoda): Detecting new mitogenomic rearrangements. J. Mol. Evol. 74, 81–95. doi:10.1007/s00239-012-9490-7

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput. Biol. 10. doi:10.1371/journal.pcbi.1003537

Bryson, R.W., Prendini, L., Savary, W.E., Pearman, P.B., 2014. Caves as microrefugia: Pleistocene phylogeography of the troglophilic North American scorpion Pseudouroctonus reddelli. BMC Evol. Biol. 14, 9. doi:10.1186/1471-2148-14-9

Bryson, R.W., Riddle, B.R., Graham, M.R., Smith, B.T., Prendini, L., 2013a. As Old as the Hills: Montane Scorpions in Southwestern North America Reveal Ancient Associations between Biotic Diversification and Landscape History. PLoS One 8. doi:10.1371/journal.pone.0052822

Bryson, R.W., Savary, W.E., Prendini, L., 2013b. Biogeography of scorpions in the Pseudouroctonus minimus complex (Vaejovidae) from south-western North America: Implications of ecological specialization for pre-Quaternary diversification. J. Biogeogr. 40, 1850–1860. doi:10.1111/jbi.12134

Byrne, M., 2008. Evidence for multiple refugia at different time scales during Pleistocene climatic oscillations in southern Australia inferred from phylogeography. Quat. Sci. Rev. 27, 2576–2585. doi:10.1016/j.quascirev.2008.08.032

Byrne, M., Yeates, D.K., Joseph, L., Kearney, M., Bowler, J., Williams, M.A.J., Cooper, S., Donnellan, S.C., Keogh, J.S., Leys, R., Melville, J., Murphy, D.J., Porch, N., Wyrwoll, K.H., 2008. Birth of a biome: Insights into the assembly and maintenance of the Australian arid zone biota. Mol. Ecol. doi:10.1111/j.1365-294X.2008.03899.x

Cao, Z., Yu, Y., Wu, Y., Hao, P., Di, Z., He, Y., Chen, Z., Yang, W., Shen, Z., He, X., Sheng, J., Xu, X., Pan, B., Feng, J., Yang, X., Hong, W., Zhao, W., Li, Z., Huang, K., Li, T., Kong, Y., Liu, H., Jiang, D., Zhang, B., Hu, J., Hu, Y., Wang, B., Dai, J., Yuan, B., Feng, Y., Huang, W., Xing, X., Zhao, G., Li, X., Li, Y., Li, W., 2013. The genome of Mesobuthus martensii reveals a unique adaptation model of arthropods. Nat. Commun. 4, 2602. doi:10.1038/ncomms3602

Chippaux, J.P., Goyffon, M., 2008. Epidemiology of scorpionism: A global appraisal. Acta Trop. doi:10.1016/j.actatropica.2008.05.021

484 De Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886.
485 doi:10.1080/10635150701701083

486 Drummond, A., Rambaut, A., 2007. Tracer v1. 5. Available from http//beast.bio.ed.ac.uk/Tracer.

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

489 Dunlop, J.A., Tetlie, O.E., Prendini, L., 2008. Reinterpretation of the Silurian scorpion
490 Proscorpius osborni (Whitfield): Integrating data from Palaeozoic and recent scorpions.
491 Palaeontology 51, 303–320. doi:10.1111/j.1475-4983.2007.00749.x

492 Fet, V., Polis, G.A., Sissom, W.D., 1998. Life in sandy deserts: the scorpion model. J. Arid
493 Environ. doi:10.1006/jare.1997.0386

494 Fet, V., Soleglad, M.E., 2005. Contributions to Scorpion Systematics. I. On Recent Changes in
495 High-Level Taxonomy. Euscorpius. Occas. Publ. Scorpiology 1–13.

496 Fox, J., 2005. The R Commander: A Basic-Statistics Graphical User Interface to R. J. Stat.
497 Softw. 14, 1–42. doi:10.1234/12345678

498 Fujisawa, T., Barraclough, T.G., 2013. Delimiting species using single-locus data and the
499 generalized mixed yule coalescent approach: A revised method and evaluation on simulated
500 data sets. Syst. Biol. 62, 707–724. doi:10.1093/sysbio/syt033

501 Gantenbein, B., Fet, V., Gantenbein-Ritter, I.A., Balloux, F., 2005. Evidence for recombination
502 in scorpion mitochondrial DNA (Scorpiones: Buthidae). Proc. Biol. Sci. 272, 697–704.
503 doi:10.1098/rspb.2004.3017

504 Gantenbein, B., Largiadèr, C.R., 2003. The phylogeographic importance of the Strait of Gibraltar
505 as a gene flow barrier in terrestrial arthropods: A case study with the scorpion Buthus
506 occitanus as model organism. Mol. Phylogenet. Evol. 28, 119–130. doi:10.1016/S1055-
507 7903(03)00031-9

508 Graham, M.M.R., Oláh-Hemmings, V., Fet, V., 2012. Phylogeography of co-distributed dune
509 scorpions identifies the Amu Darya River as a long-standing component of Central Asian
510 biogeography: (Scorpiones: Buthidae. Zool. Middle East 55, 95–110.
511 doi:10.1080/09397140.2012.10648924

512 Gurevitz, M., Karbat, I., Cohen, L., Ilan, N., Kahn, R., Turkov, M., Stankiewicz, M., Stühmer,
513 W., Dong, K., Gordon, D., 2007. The insecticidal potential of scorpion β-toxins. Toxicon.
514 doi:10.1016/j.toxicon.2006.11.015

515 Hillis, D.M., Dixon, M.T., 1991. Ribosomal DNA: Molecular evolution and phylogenetic
516 inference. Q. Rev. Biol. 66, 411–446. doi:10.1086/417338

Koch, L.E., 1977. The taxonomy, geographic distribution and evolutionary radiation of Australo-Papuan scorpions. Rec. West. Aust. Museum 5, 79.

Lourenço, W.R., 2001. The scorpion families and their geographical distribution. J. Venom. Anim. Toxins. doi:10.1590/S0104-79302001000100002

Lourenço, W.R., Cuellar, O., 1995. Scorpions, scorpionism, life history strategies and parthenogenesis. J. Venom. Anim. Toxins. doi:10.1590/S0104-79301995000200002

Luna-Ramírez, K., Bartok, A., Restano-Cassulini, R., Quintero-Hernández, V., Coronas, F.I. V, Christensen, J., Wright, C.E., Panyi, G., Possani, L.D., 2014. Structure, molecular modeling, and function of the novel potassium channel blocker urotoxin isolated from the venom of the Australian scorpion Urodacus yaschenkoi. Mol. Pharmacol. 86, 28–41. doi:10.1124/mol.113.090183

Luna-Ramírez, K., Quintero-Hernández, V., Vargas-Jaimes, L., Batista, C.V.F., Winkel, K.D., Possani, L.D., 2013. Characterization of the venom from the Australian scorpion Urodacus yaschenkoi: Molecular mass analysis of components, cDNA sequences and peptides with antimicrobial activity. Toxicon 63, 44–54. doi:10.1016/j.toxicon.2012.11.017

Martin, H.A., 2006. Cenozoic climatic change and the development of the arid vegetation in Australia. J. Arid Environ. 66, 533–563. doi:10.1016/j.jaridenv.2006.01.009

Oukkache, N., El Jaoudi, R., Ghalim, N., Chgoury, F., Bouhaouala, B., El Mdaghri, N., Sabatier, J.M., 2014. Evaluation of the lethal potency of scorpion and snake venoms and comparison between intraperitoneal and intravenous injection routes. Toxins (Basel). 6, 1873–1881. doi:10.3390/toxins6061873

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D., Vogler, A.P., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst. Biol. 55, 595–609. doi:10.1080/10635150600852011

Posada, D., 2008. jModelTest: Phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256. doi:10.1093/molbev/msn083

Possani, L.D., Merino, E., Corona, M., Bolivar, F., Becerril, B., 2000. Peptides and genes coding for scorpion toxins that affect ion-channels. Biochimie. doi:10.1016/S0300-9084(00)01167-6

Prendini, L., 2000. Phylogeny and Classification of the Superfamily Scorpionoidea Latreille 1802 (Chelicerata , Scorpiones): An Exemplar Approach. Society 78, 1–78. doi:10.1006/clad.1999.0127

550 Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J.W.,
551 Cunningham, C.W., 2010. Arthropod relationships revealed by phylogenomic analysis of
552 nuclear protein-coding sequences. Nature 463, 1079–1083. doi:10.1038/nature08742

553 Rodríguez de la Vega, R.C., Schwartz, E.F., Possani, L.D., 2010. Mining on scorpion venom
554 biodiversity. Toxicon. doi:10.1016/j.toxicon.2009.11.010

555 Satler, J.D., Carstens, B.C., Hedin, M., 2013. Multilocus species delimitation in a complex of
556 morphologically conserved trapdoor spiders (mygalomorphae, antrodiaetidae, Aliatypus).
557 Syst. Biol. 62, 805–823. doi:10.1093/sysbio/syt041

558 Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook., P., 1994. Evolution, weighting,
559 and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved
560 PCR primers. Ann. Entomol. Soc. Am. 87, 651–701.

561 Stachel, S.J., Stockwell, S.A., Van Vranken, D.L., 1999. The fluorescence of scorpions and
562 cataractogenesis. Chem. Biol. 6, 531–539. doi:10.1016/S1074-5521(99)80085-4

563 Talavera, G., Dincă, V., Vila, R., 2013. Factors affecting species delimitations with the GMYC
564 model: Insights from a butterfly survey. Methods Ecol. Evol. 4, 1101–1110.
565 doi:10.1111/2041-210X.12107

566 Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics
567 Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.
568 doi:10.1093/molbev/msm092

569 Tanaka, H., Roubik, D.W., Kato, M., Liew, F., Gunsalam, G., 2001. Phylogenetic position of
570 Apis nuluensis of northern Borneo and phylogeography of A. cerana as inferred from
571 mitochondrial DNA sequences. Insectes Soc. doi:10.1007/PL00001744

572 Tourinho, J.L., Sole-Cava, A.M., Lazoski, C., 2012. Cryptic species within the commercially
573 most important lobster in the tropical Atlantic, the spiny lobster Panulirus argus. Mar. Biol.
574 159, 1897–1906. doi:http://dx.doi.org/10.1007/s00227-012-1977-7

575 Volschenk, E.S., Harvey, M.S., Prendini, L., 2012. A new species of Urodacus (Scorpiones:
576 Urodacidae) from Western Australia. Am. Museum Novitiates 1–18. doi:10.1206/3748.2

577 Volschenk, E.S., Mattoni, C.I., Prendini, L., 2008. Comparative anatomy of the mesosomal
578 organs of scorpions (Chelicerata, Scorpiones), with implications for the phylogeny of the
579 order. Zool. J. Linn. Soc. doi:10.1111/j.1096-3642.2008.00426.x

580 Volschenk, E.S., Prendini, L., 2008. Aops oncodactylus, gen. et sp. nov., the first troglobitic
581 urodacid (Urodacidae:Scorpiones), with a re-assessment of cavernicolous, troglobitic and
582 troglomorphic scorpions. Invertebr. Syst. 22, 235–257. doi:10.1071/IS06054

583 Wheeler, Q.D., 1999. Why the phylogenetic species concept?-Elementary. J. Nematol. 31, 134–
584 141.

585 Wysocka, A., Krzysztofiak, L., Krzysztofiak, A., Zołnierkiewicz, O., Ojdowska, E., Sell, J.,
586 2011. Low genetic diversity in Polish populations of sibling ant species: Lasius niger (L.)
587 and Lasius platythorax Seifert (Hymenoptera, Formicidae). Insectes Soc. 58, 191–195.
588 doi:10.1007/s00040-010-0135-9

589 Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A., 2013. A general species delimitation method
590 with applications to phylogenetic placements. Bioinformatics 29, 2869–2876.
591 doi:10.1093/bioinformatics/btt499

592

593 **Web references**:

594 Department of the Environment, Water, Heritage and the Arts (12 February 2010). "Species
595 Urodacus yaschenkoi (Birula, 1903)".*Australian Biological Resources Study: Australian*
596 *Faunal Directory*. Commonwealth of Australia. Retrieved 20 July 2015.

597 CITES Appendix II, http://www.cites.org/eng/app/appendices.php; accessed on Sep 10, 2014.

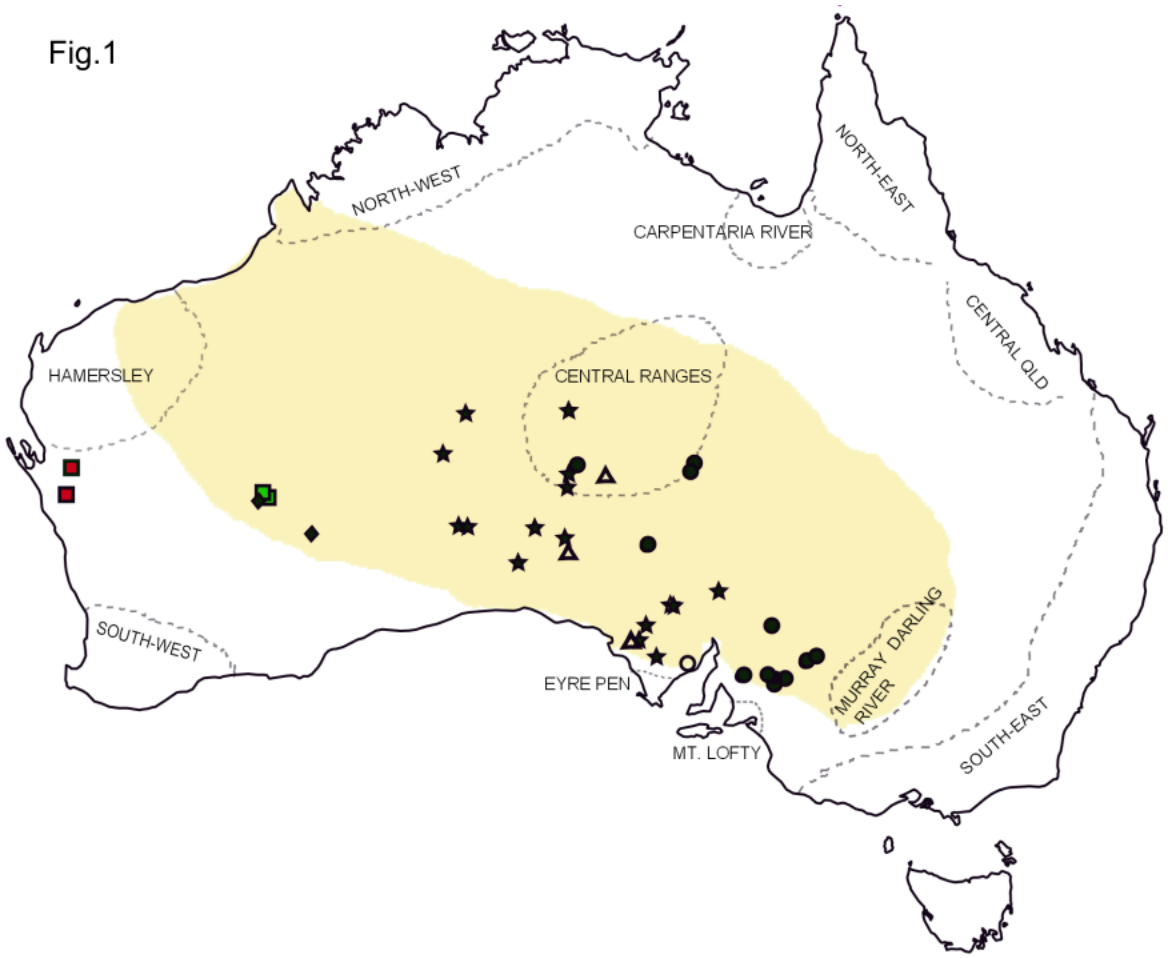
598 Medscape, http://emedicine.medscape.com/article/168230-overview; accessed on Aug 20, 2014.

599 Image of *Urodacus yaschenkoi* downloaded from:
600 http://images.ala.org.au/image/details?imageId=8ead32f0-1589-461a-988f-08c35b6a3e46

601 **Figure 1. Map depicting the sampling locations across the natural distribution of *Urodacus***
602 ***yaschenkoi* .** Samples are coded based on the major mitochondrial lineage (red, green, black) and
603 putative species (star, triangle, circle etc.). The extant distribution (yellow shading) and major
604 Australian refugia (dashed line) are adopted from (Koch 1977).

605

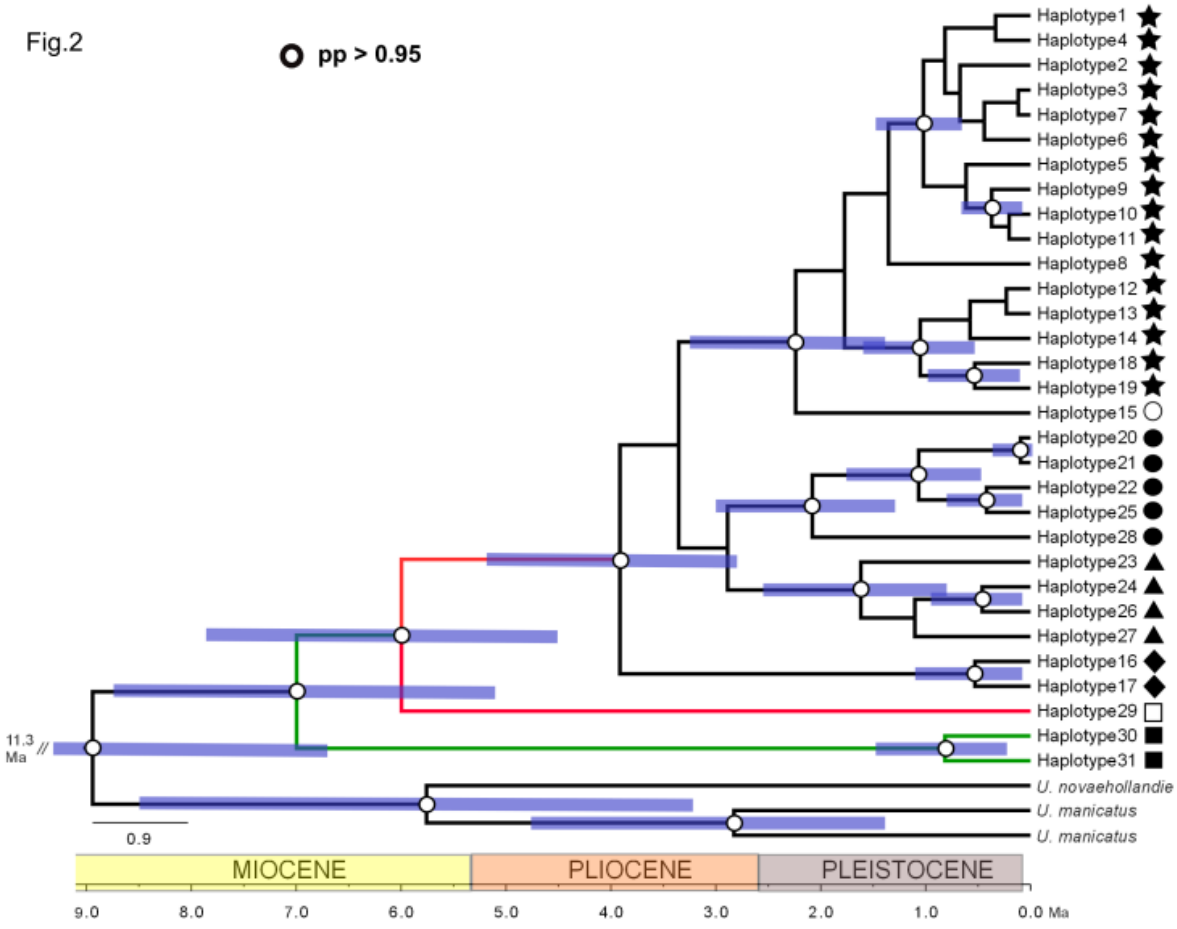
606



607

608

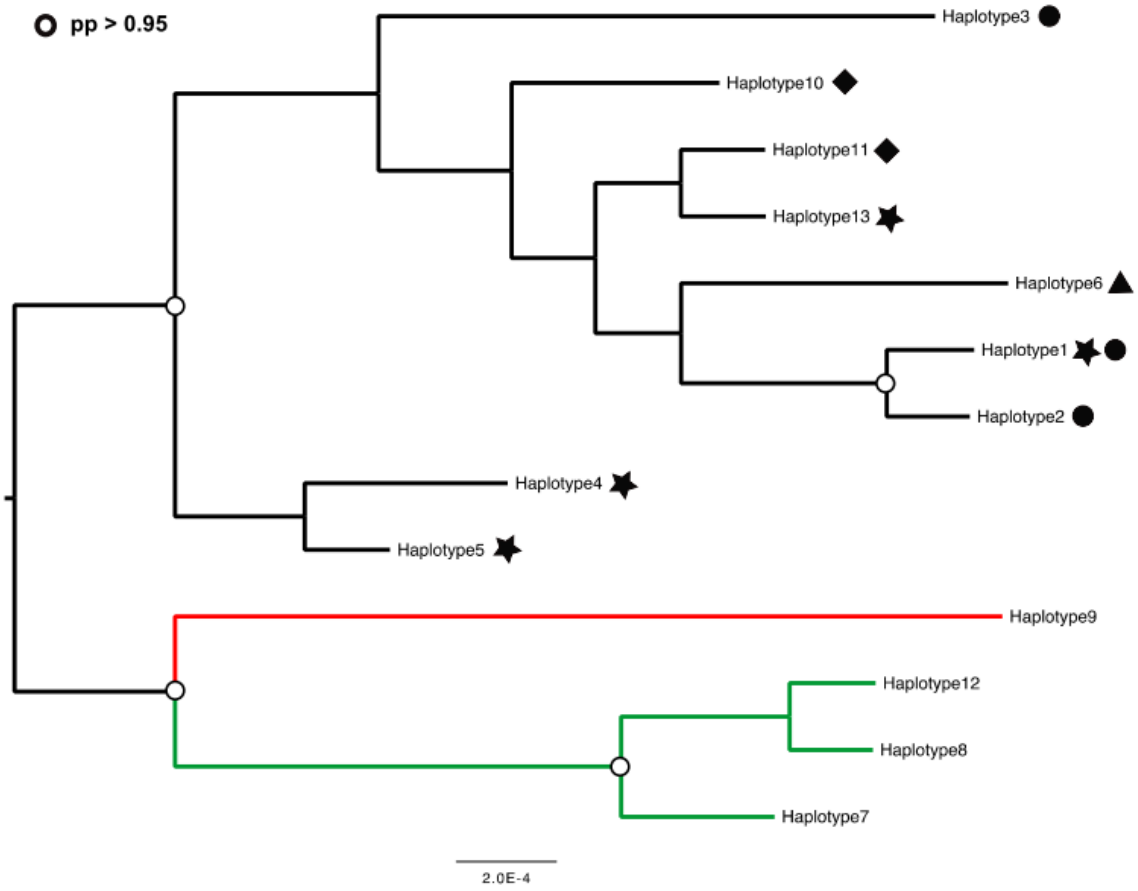
609 **Figure 2. Bayesian Inference phylogeny based on the concatenated sequences from two**
610 **mitochondrial genes (COXI and 16S).** Nodes supported by Bayesian posterior probabilities
611 greater than or equal to 0.95 are indicated by black circles with white fill. Blue shaded bars
612 indicate the 95% highest posterior density (HPD) for node ages (scale bars represents time in
613 millions of years from the present day). The assignment of haplotypes to putative species groups
614 identified by PTP/bPTP is depicted with symbols (star, hollow circle, solid circle, triangle,
615 diamond, cross, square).
616
617



618 **Figure 3. Bayesian Inference phylogeny based on nuclear 28SA data**. Nodes supported by
619 Bayesian posterior probabilities greater than or equal to 0.95 are indicated by black circles with
620 white fill. Three major ancestral lineages are coded as green, red and black (corresponding to
621 those identified in the mitochondrial phylogeny).

622

623 Fig.3

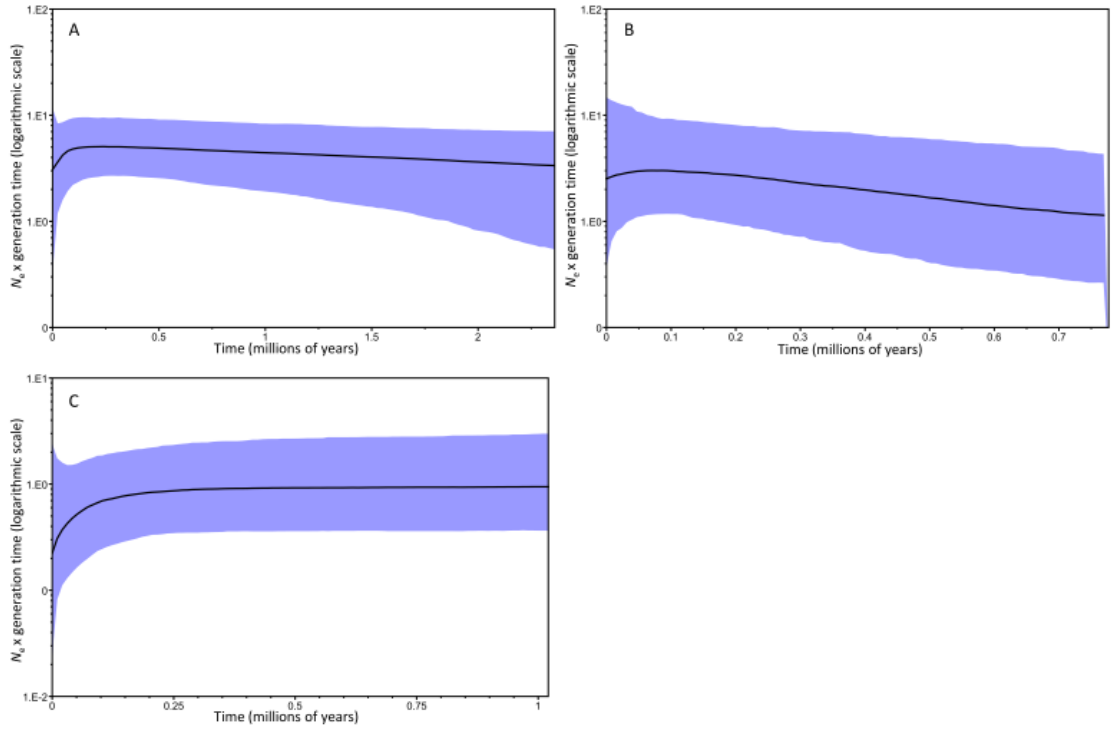


624

625 **Figure 4**. **Demographic history of *U. yaschenkoi* determined by Bayesian skyline analysis of**
626 **mitochondrial data**. Reported are the plots of *Ne* (effective population size multiplied per
627 generation time) against time in years from present. Mean *Ne* is shown as a solid black line and
628 shaded blue areas enclose the 95 % highest posterior density interval. Figure 'A' represents the
629 analysis of all haplotypes assigned to the south-central ancestral lineage, while figures 'B' and
630 'C' represent the analyses of haplotypes assigned to the 'star' and 'circle' sub-lineages within the
631 south-central ancestral lineage.

632

633 Fig.4

634 

635 **Figure 5. Linear Discriminant Analysis with ratios of morphological traits in *U. yaschenkoi*.**

636 Separation of 39 adult females assigned to the four putative species (designated with the star,

637 circle, triangle and square symbols) based on a linear combination of 21 variables. 93.7%

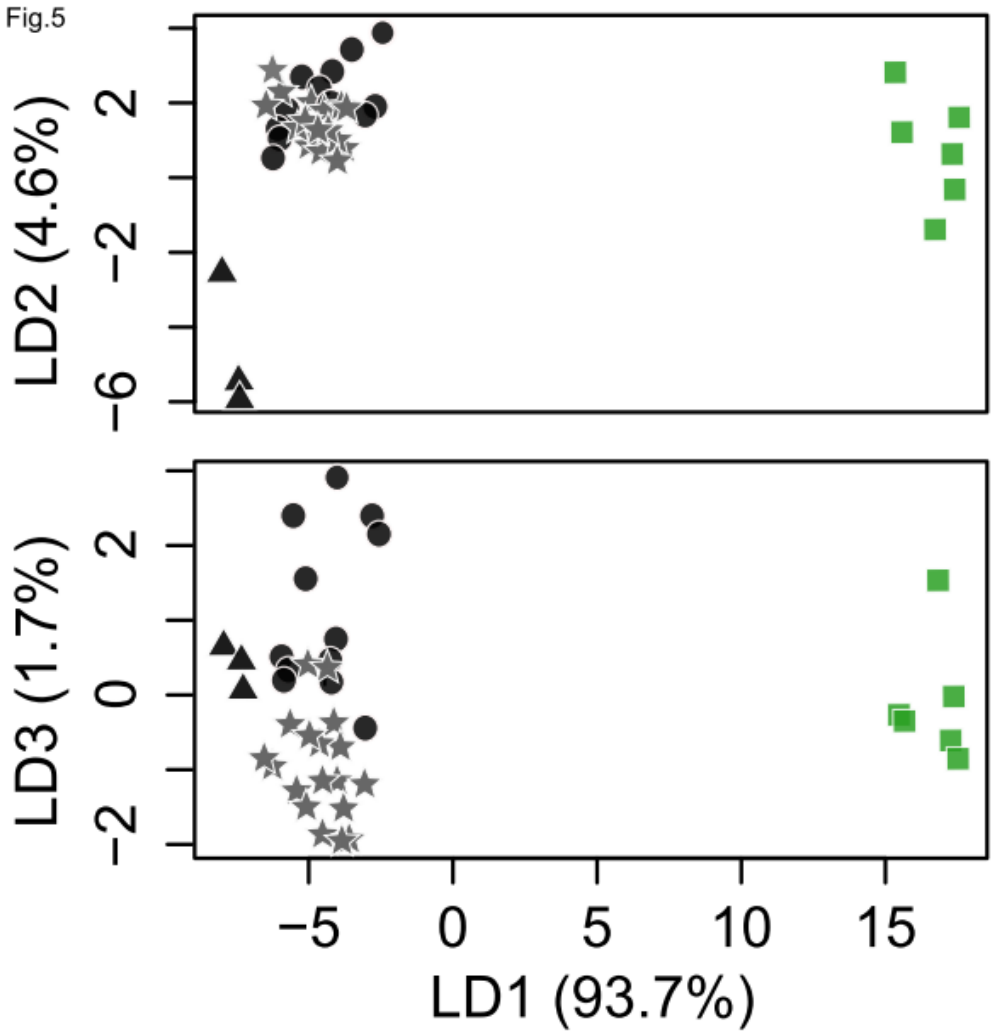
638 separation between individuals belonging to the western-central (green) clade and all other

639 individuals from the south-central (black) clade was achieved by the first discriminant function

640 (LD1).

641

642



643 **Table 1. *Urodacus yaschenkoi* specimen location and analyses made**. List of *Urodacus*
644 *yaschenkoi* collected from the field as live specimens (Field) or obtained from the Australian
645 museum collections (South Australian Museum - SA, Western Australian Museum - WA).
646 Geographic position (lat/log) and the geographic region details are reported for each sample. List
647 of molecular loci (*COXI*, 16*S*, 28*S*) scored in each individual. Seven morphological characters
648 were scored in 39 specimens (☞).

Sample	Source	Latitude	Longitude	Geographic Region	Markers	Morpho	Museum ID/Reg.No.
BKA11	Field	-33.2283	141.3011	NSW	COXI, 16S, 28S		NA
BKA12	Field	-33.2283	141.3011	NSW	COXI, 16S, 28S		NA
BKB10	Field	-33.2199	141.3089	NSW	COXI, 16S, 28S		NA
BKB12	Field	-33.2242	141.3061	NSW	COXI, 16S, 28S		NA
BKB13	Field	-33.2283	141.3011	NSW	COXI, 16S, 28S		NA
MARR1	Field	-26.3400	133.2000	SA	COXI, 16S, 28S		NA
MARR2	Field	-26.3400	133.2000	SA	COXI, 16S, 28S		NA
PIM1	Field	-31.2509	136.5089	SA	COXI, 16S, 28S		NA
PIM2	Field	-31.2509	136.5089	SA	COXI, 16S, 28S		NA
PIM5	Field	-31.2509	136.5089	SA	COXI, 16S, 28S		NA
PIM6	Field	-31.2509	136.5089	SA	COXI, 16S, 28S		NA
PIM7	Field	-31.2509	136.5089	SA	COXI, 16S, 28S		NA
PIM8	Field	-31.2509	136.5089	SA	COXI, 16S, 28S		NA
POP1	Field	-33.0710	141.6372	NSW	COXI, 16S, 28S		NA
POP4	Field	-33.0710	141.6372	NSW	COXI, 16S, 28S		NA
POP5	Field	-33.0710	141.6372	NSW	COXI, 16S, 28S		NA
SAM1397	SAM	-30.7667	138.1767	SA	COXI, 16S	☞	NS1397

SAM1399	SAM	-27.1192	132.8300	SA	COXI, 16S	☞	NS1399
SAM1400	SAM	-27.1191	132.8300	SA	COXI, 16S	☞	NS1400
SAM1403	SAM	-26.6453	132.8858	SA	COXI, 16S	☞	NS1403
SAM1406	SAM	-31.2878	136.5831	SA	COXI, 16S	☞	NS1406
SAM1412	SAM	-26.2747	137.3269	SA	COXI, 16S	☞	NS1412
SAM1415	SAM	-33.8555	140.5361	SA	COXI, 16S	☞	NS1415
SAM1416	SAM	-34.0583	140.1500	SA	COXI, 16S	☞	NS1416
SAM1606	SAM	-26.6922	134.1722	SA	COXI, 16S		NS1606
SAM1607	SAM	-26.5767	137.1933	SA	COXI, 16S		NS1607
SAM1812	SAM	-33.3267	137.0931	SA	COXI, 16S	☞	NS1812
SAM1823	SAM	-33.7511	140.2747	SA	COXI, 16S	☞	NS1823
SAM1824	SAM	-33.7725	140.2117	SA	COXI, 16S	☞	NS1824
SAM1825	SAM	-33.7230	140.1238	SA	COXI, 16S	☞	NS1825
SAM1831	SAM	-33.7183	139.9300	SA	COXI, 16S	☞	NS1831
SAM1834	SAM	-33.7236	139.0438	SA	COXI, 16S	☞	NS1834
SAM1835	SAM	-33.7236	139.0438	SA	COXI, 16S	☞	NS1835
SAM1837	SAM	-33.7400	139.0816	SA	COXI, 16S	☞	NS1837
SAM1917	SAM	-32.6244	135.0322	SA	COXI, 16S	☞	NS1917
SAM1939	SAM	-33.1233	136.0214	SA	COXI, 16S	☞	NS1939
SAM2038	SAM	-33.1167	136.0000	SA	COXI, 16S	☞	NS2038
SAM2053	SAM	-24.4036	132.8886	NT	COXI, 16S	☞	NS2053
SAM2054	SAM	-28.4627	129.0102	SA	COXI, 16S	☞	NS2054
SAM2055	SAM	-28.4627	129.0102	SA	COXI, 16S	☞	NS2055
SAM2056	SAM	-28.4627	129.0102	SA	COXI, 16S	☞	NS2056
SAM2060	SAM	-28.4977	129.3205	SA	COXI, 16S	☞	NS2060

SAM2061	SAM	-28.4977	129.3205	SA	COXI, 16S	☞	NS2061
SAM2062	SAM	-24.5060	129.2619	NT	COXI, 16S	☞	NS2062
SAM2067	SAM	-32.0033	135.6558	SA	COXI, 16S		NS2067
SAM2070	SAM	-28.8969	132.7575	SA	COXI, 16S	☞	NS2070
SAM2071	SAM	-28.8969	132.7575	SA	COXI, 16S	☞	NS2071
SAM2073	SAM	-28.5319	131.6903	SA	COXI, 16S		NS2073
SAM2076	SAM	-29.7706	131.1081	SA	COXI, 16S		NS2076
SAM2120	SAM	-31.9972	140.0644	SA	COXI, 16S	☞	NS2120
SAM2125	SAM	-29.1286	135.6997	SA	COXI, 16S	☞	NS2125
SAM2126	SAM	-29.1286	135.6997	SA	COXI, 16S		NS2126
SAM2133	SAM	-32.4947	135.3644	SA	COXI, 16S	☞	NS2133
SAM2140	SAM	-29.4053	132.8556	SA	COXI, 16S	☞	NS2140
WAM20	WAM	-27.4867	122.3119	WA	COXI, 16S, 28S	☞	85020
WAM31	WAM	-27.4867	122.3119	WA	COXI, 16S, 28S	☞	85031
WAM32	WAM	-27.4867	122.3119	WA	COXI, 16S, 28S	☞	85032
WAM36	WAM	-27.3893	115.1847	WA	COXI, 16S, 28S		78236
WAM37	WAM	-27.6145	121.9947	WA	COXI, 16S, 28S		112637
WAM38	WAM	-26.4408	115.3661	WA	COXI, 16S, 28S		78238
WAM46	WAM	-28.7333	123.8667	WA	COXI, 16S, 28S		80246
WAM55	WAM	-27.4867	122.3119	WA	COXI, 16S, 28S	☞	83855
WAM56	WAM	-27.4867	122.3119	WA	COXI, 16S, 28S	☞	83856
WAM75	WAM	-27.4867	122.3119	WA	COXI, 16S, 28S	☞	83875
WAM88	WAM	-25.9307	128.4526	WA	COXI, 16S, 28S		95988
Um1814	SAM	-33.1997	138.2189	SA	COXI, 16S		NS0001814
Um2714	SAM	-33.1997	138.2189	SA	COXI, 16S		NS0002714

Un2112	SAM	-31.6597	129.1083	SA	COXI, 16S	NS0002112

NSW: New South Wales; SA: South Australia; WA: Western Australia; NT: Northern Territory
NA: Not applicable.

653 **Table 2. List of primer sequences and corresponding amplicons sizes for the three**
654 ***Urodacus yaschenkoi* loci** (*COXI*, 16*S* rRNA, 28*S* rRNA).

Marker	Primer	Primer sequence	Size (bp)	Reference
COXI	F C1-J-2183	5'-CAACATTTATTTTGATTTTTTGG - 3'	550-630	(Simon et al., 1994)
	R COXIKG-R2	5'- GATATTAATCCTAAAAAATGTTGAGG-3'		(Tanaka et al., 2001)
COXI	Nested F	5'-AGGAACCTTTTGGGGCTTT-3'	150	
COXI	Nested R	5'-AGGAACCTTTTGGGGCTTT-3'		
16S	F 16SF	5'- AACAAAACCCACAGCTCACA- 3'	422	(Gantenbein et al., 2005)
	R 16SR	5'- GTGCAAAGGTAGCATAATCA- 3'		
28S	R1	F R1S (5'-ACCCGCTGAATTTAAGCAT-3'),	1158	(Arabi et al., 2012)
		R R1AS (5'- GCTATCCTGAGGGAAACTTC-3')		
	R2	F R2S (5'-CGACCCGTCTTGAAACACGGA-3'),	1246	
		R R2AS (5'-CACCTTGGAGACCTGCTGCGGAT-3')		

655

656

657

658

659

660

661

662 **Table 3. Species delineation analyses in *Urodacus yaschenkoi* based on 31 unique**
663 **mitochondrial haplotypes**.

Analysis type	*# Entities*	*Statistics*
GMYC	9	Likelihood null model: 32.7519; likelihood best model: 33.36569; likelihood ratio: 1.2255; P-value, 0.0001, confidence interval: 1-10
PTP/bPTP (ML and BL)	7	Acceptance rate: 0.50975; merge: 49942; split: 50058

664

Supplemental Information

The data sets supporting the results of this article are included within the article and its additional files in Supplemental_Files_1-4.xlsx

Supplemental File 1. Pairwise uncorrected *p*-distance between 31 unique *U. yaschenkoi* haplotypes and three outgroup haplotypes (*U. novaehollandiae* and two *U. manicatus*). Haplotypes were generated from the concatenated partial sequences of COXI and 16S loci.

Supplemental File 2. Pairwise uncorrected *p*-distance between 13 unique *U. yaschenkoi* haplotypes generated from the partial 28S sequence.

Supplemental File 3. List of haplotype numbers assigned to the *U. yaschenkoi* samples.

Supplemental File 4. Measures (in mm) of seven morphological traits in 39 *U. yaschenkoi* adult females.