Received: 9 December 2017 | Revised: 17 October 2018 | Accepted: 22 October 2018

DOI: 10.1111/mec.14925

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Genomewide SNP markers breathe new life into phylogeography and species delimitation for the problematic short-necked turtles (Chelidae: *Emydura*) of eastern Australia

Arthur Georges[1] | Bernd Gruber[1] | Greg B. Pauly[2] | Duanne White[1] | Mark Adams[1,3] | Matthew J. Young[1] | Andrzej Kilian[4] | Xiuwen Zhang[1] | H. Bradley Shaffer[5] | Peter J. Unmack[1]

[1]Institute for Applied Ecology, University of Canberra, Canberra, Australian Capital Territory, Australia

[2]Natural History Museum of Los Angeles County, Los Angeles, California

[3]Department of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia

[4]Diversity Arrays Technology Pty Ltd, University of Canberra, Canberra, Australian Capital Territory, Australia

[5]Department of Ecology and Evolutionary Biology, La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, University of California, Los Angeles, California

Correspondence
Arthur Georges, Institute for Applied Ecology, University of Canberra, Canberra, ACT, Australia.
Email: georges@aerg.canberra.edu.au

Funding information
Cooperative Research Centre for Freshwater Ecology; National Science Foundation, Grant/Award Number: 1457832; Collaborative Research Network for Murray-Darling Basin Futures; Australian Research Council, Grant/Award Number: LP140100521

Abstract

Understanding the evolutionary history of diversifying lineages and the delineation of evolutionarily significant units and species remains major challenges for evolutionary biology. Low-cost representational sampling of the genome for single nucleotide polymorphisms shows great potential at the temporal scales that are typically the focus of species delimitation and phylogeography. We apply these markers to a case study of a freshwater turtle, *Emydura macquarii*, whose systematics has so far defied resolution, to bring to light a dynamic system of substantive allopatric lineages diverging on independent evolutionary trajectories, but held back in the process of speciation by low level and episodic exchange of alleles across drainage divides on various timescales. In the context of low-level episodic gene flow, speciation is often reticulate, rather than a bifurcating process. We argue that species delimitation needs to take into account the pattern of ancestry and descent of diverging lineages in allopatry together with the recent and contemporary processes of dispersal and gene flow that retard and obscure that divergence. Underpinned by a strong focus on lineage diagnosability, this combined approach provides a means for addressing the challenges of incompletely isolated populations with uncommon, but recurrent gene flow in studies of species delimitation, a situation likely to be frequently encountered. Taxonomic decisions in cases of allopatry often require subjective judgements. Our strategy, which adds an additional level of objectivity before that subjectivity is applied, reduces the risk of taxonomic inflation that can accompany lineage approaches to species delimitation.

KEYWORDS

ddRAD-seq, hybridization, introgression, landscape genomics, Murray–Darling Basin, species boundaries

1 | INTRODUCTION

Allopatric speciation arises through divergence of populations in isolation and involves the processes of mutation, genetic drift and selection under differential local conditions (Dobzhansky, 1951; Mayr, 1963). Speciation, in its simplest conception, involves the progressive divergence of two lineages in the absence of gene flow until reproductive compatibility is eroded, and ultimately the two lineages become irrevocably committed to distinct evolutionary fates. Early in this process, isolation of two sister lineages may not be absolute, in which case speciation begins with a pattern of reduced introgression in the regions of the genome that are most differentiated between two divergent lineages (Gompert et al., 2012; Harrison & Larson, 2016); it ends when reproductive compatibility becomes compromised, giving permanence to the genetic isolation formerly arising from potentially ephemeral geographic isolation.

This process often does not follow a simple bifurcating path. Speciation can progress over millions of years, only to be subject to an uncertain outcome should the incipient species again come into contact. Contact may accelerate reproductive isolation through selective reinforcement of barrier genes (Harrison & Larson, 2016; Mao, Zhang, & Rossiter, 2016; Wu, 2001) and genes governing other traits that reduce the frequency of maladaptive hybridization (Dobzhansky, 1951; Hoskin, Higgie, McDonald, & Moritz, 2005). Alternatively, contact and progressive genomewide introgression may completely reverse the process of speciation. Speciation may be complicated, and delayed, by periods of strict allopatry punctuated by recurrent opportunities for gene flow (Stuglik & Babik, 2016). The consequences of a protracted process of speciation, and periods of "despeciation" driven by episodic gene flow, clearly have implications for species delimitation and phylogenetic analysis of closely related taxa. Cladogenic processes of divergence under isolation will interact with tokogenic processes when there is limited and possibly episodic gene flow between the diverging populations (Edwards, Potter, Schmitt, Bragg, & Moritz, 2016; Maynard Smith, 1989; Morrison, 2016). Nevertheless, over the last few decades, emphasis in studies of the delineation of species has been on the recovery of a bifurcating pattern of ancestry and descent, with often less attention paid to incorporating historical and contemporary episodes of reticulation through contact and gene flow (Jackson, Carstens, Morales, & O'Meara, 2017). A focus on defining species as substantive lineages also fails to make clear the distinction between divergence accompanied by reproductive isolation as opposed to that arising from geographic isolation alone (discussed by Fujita, Leaché, Burbrink, McGuire, & Moritz, 2012; de Queiroz, 2007; de Queiroz, Howard, & Berlocher, 1998; Sites & Marshall, 2003).

High-throughput parallel sequencing (next-generation sequencing, Metzker, 2010) and low-cost reduced representational sampling of the genome for single nucleotide polymorphisms (SNPs)—DArTSeq, RADSeq and ddRAD (Baird et al., 2008; Jaccoud, Peng, Feinstein, & Kilian, 2001; Kilian et al., 2012; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012; Sansaloni et al., 2011; van Tassell et al., 2008)—have enabled genomics at the level of populations to be considered with phylogenomics in studies of the pattern and process of speciation, species delimitation and phylogeography. Genotyping of geographically comprehensive samples is now possible to deliver better understanding of the historical and contemporary drivers of geographic patterns in genetic diversity at regional scales involving species, subspecies, and other evolutionarily significant units (ESUs).

In this paper, we argue that the first clearly defined step in speciation is the acquisition and retention of diagnostic traits—traits possessed by individuals from one taxon or set of populations, and absent from individuals from another. Thus, a first step in any investigation of species delimitation should be to identify diagnosable taxa (operational taxonomic units, OTUs) from the sampled populations. We acknowledge that rarely is it possible to use diagnosability to pinpoint the precise time, place or circumstance when two or more sister populations become irrevocably committed to different evolutionary paths (Bush, 1993), so the criterion of diagnosability does not distinguish between population structure and species boundaries when applied in allopatry. Diagnosability thus becomes a necessary but not sufficient basis for species delimitation when assessing allopatric taxa. This is especially so as we refine DNA techniques for identifying diagnostic characters at finer and finer scales. Thus, we admit that a level of subjectivity in the delimitation of species in allopatry is unavoidable, even when they are diagnosably distinct.

Given this, our empirical approach to species delimitation involves a series of steps. First, we require comprehensive coverage of populations within the target taxa, both to avoid interpreting sparse sampling of demes on a cline as distinct taxa and to capture evidence of any gene flow across contact zones. Second, we require that putative instances of contemporary gene flow between adjacent populations be identified—in our case, qualitatively by examination of ordination plots and quantitatively using software such as NewHybrids (Anderson & Thompson, 2002). The third step is to identify diagnosable units using analysis of fixed allelic differences, recursively amalgamating populations into operational taxonomic units (OTUs). Only once these OTUs are defined, are phylogenetic techniques applied to identify a pattern of ancestry and descent among the taxonomic units. This represents a substantial departure from common practice, which invokes phylogenetic considerations much earlier in the process of species delimitation (for an alternative approach in a similar context, see Chan et al., 2017). In the fourth and final step, judgement is exercised to decide which of the diagnosable OTUs are significant taxonomically (as species, subspecies, ESUs), which means taking into account the evidence of gene flow between them (or lack of it when opportunity presents, e.g., parapatry), phylogeny and the level of genetic divergence in comparison with other well-defined comparable taxa (Georges & Adams, 1996).

We apply this empirical approach to the delimitation of species in a difficult case study that is likely to reflect frequently encountered circumstances. We focus on genetic variation at a continental scale in a group of freshwater turtles in eastern Australia, hereafter

referred to as the southern *Emydura*. Their range extends almost the full length of the Australian continent from north to south (2,500 km), and they are obligate freshwater organisms with dispersal constrained by the ocean, well-defined drainage divides and dendritic riverine structure. The group as currently recognized is a complex of closely related and morphologically distinct allopatric forms, variously regarded as species, subspecies or distinct morphological lineages (Cann, 1998; Cann & Sadlier, 2017; Cogger, Cameron, & Cogger, 1983; Georges & Thomson, 2010; Turtle Taxonomy Working Group, 2017). These include a melanistic pygmy form from the dune lakes of Fraser Island and the adjacent mainland, an isolated example of gigantism in the inland Lake Eyre Basin, and many other morphological forms (Figure 1). The group exhibits great variability across its range in colour pattern (particularly the presence or absence of a yellow stripe behind the eye) and shell shape. Reliance on coloration and body size to make diagnoses has led to a confused taxonomy at the level of species and subspecies (Georges & Thomson, 2010; Turtle Taxonomy Working Group, 2017). In contrast to the morphological variability used to define these taxa, no fixed allelic differences between any of the presumed species or subspecies were evident in a comprehensive allozyme study (Georges & Adams, 1996); the various morphotypes often shared even rare alleles, which suggests some level of relatively recent gene flow between drainages.

Here, we use SNP and mitochondrial sequence data to evaluate contemporary and historical drivers of genetic structure. We interpret our results in the context of the emerging controversy on the use of lineages (and their resultant clades) to delineate species (Hoelzel, 2016; Sukumaran & Knowles, 2017), and interpret the southern *Emydura* as a contemporary example of episodic speciation and despeciation.

2 | MATERIALS AND METHODS

2.1 | Specimen collection

For the SNP genotyping, 406 individuals were sampled from 57 waterbodies from the coastal drainages of eastern Australia, from the Hunter River in the south (New South Wales) to the Normanby River (Queensland) in the north; the rivers of the Murray–Darling Basin (MDB), including the Paroo drainage, and the Lake Eyre Basin (LEB), and the intervening Bulloo Basin (Figure 2). The populations included from the MDB were representative of much more extensive sampling to be presented elsewhere, chosen after a preliminary analysis

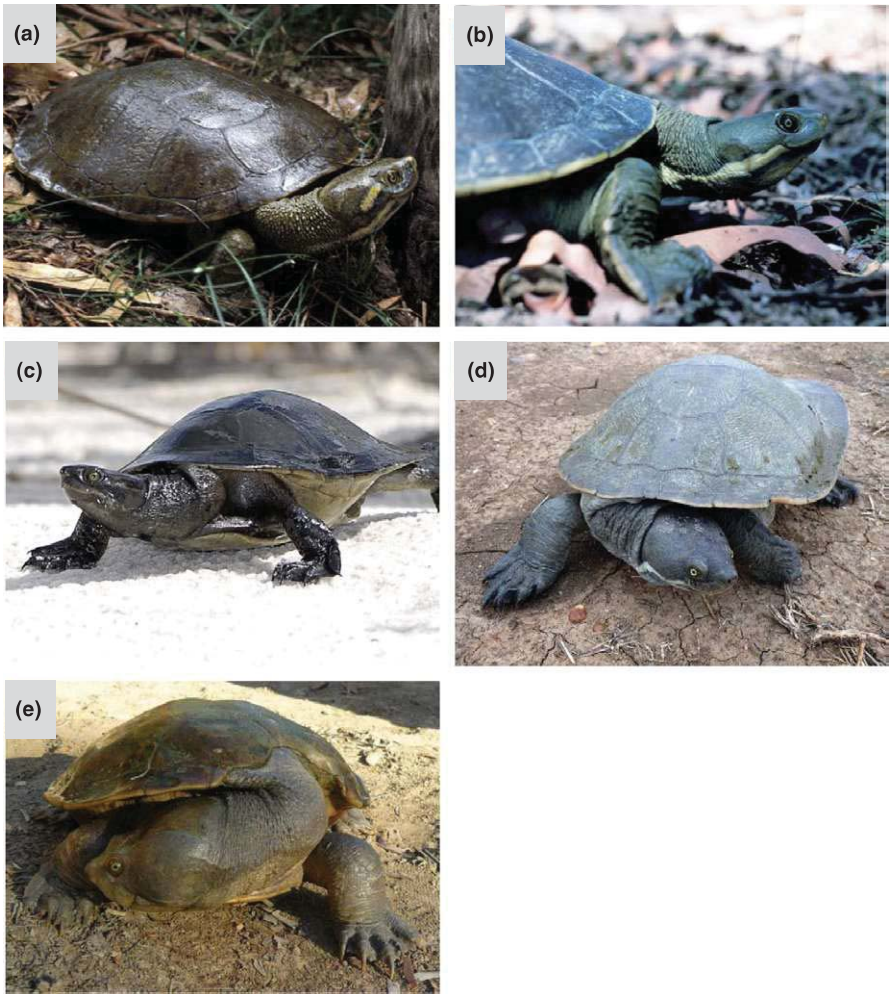

FIGURE 1 Different forms of the southern *Emydura* variously regarded as species or subspecies. (a) *Emydura krefftii* from the E and NE Coast; (b) *E. signata* from the Brisbane River; (c) *E. macquarii nigra* from Fraser Island; (d) *E. macquarii* from the Murray–Darling Basin; and (e) *E. macquarii emmotti* from the Lake Eyre Basin. These taxa have typically been distinguished on coloration, body size and shell shape

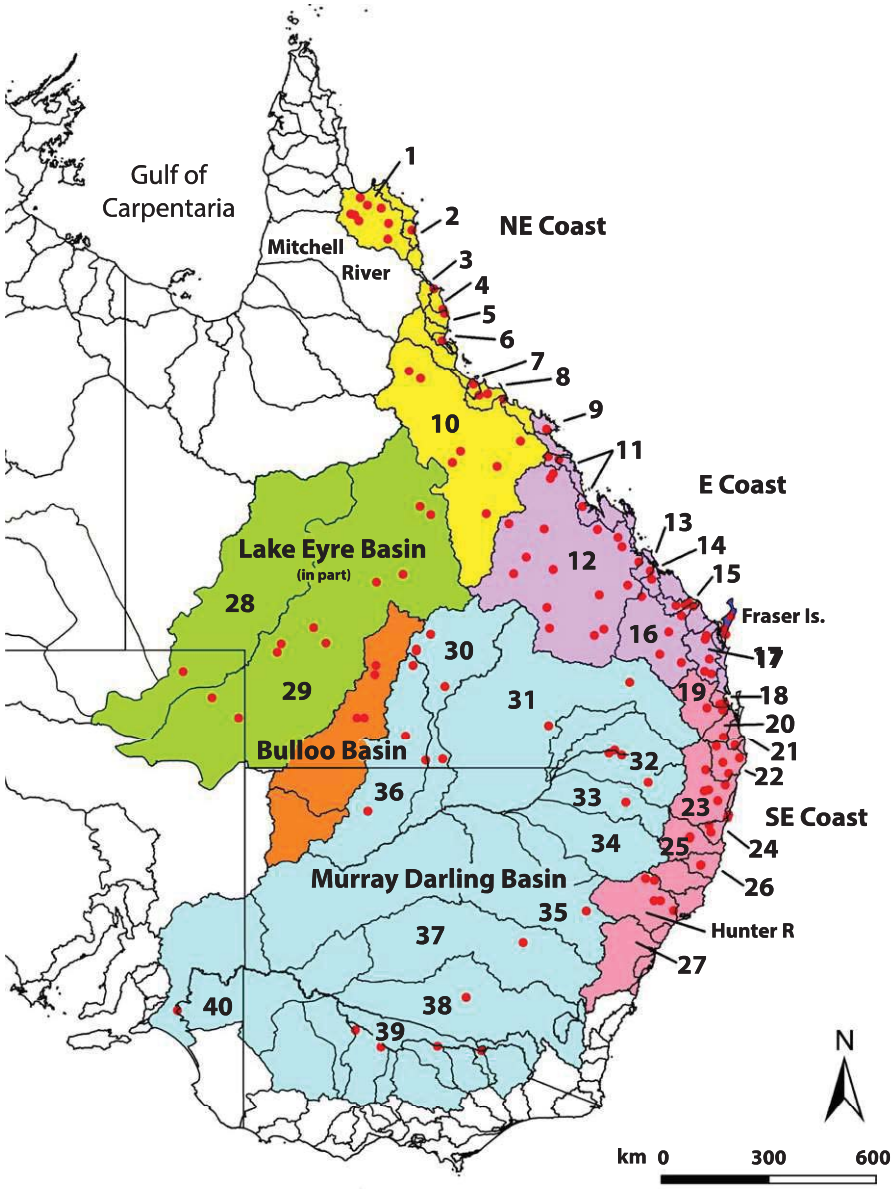

FIGURE 2 A map of eastern mainland Australia showing the distribution of the *Emydura macquarii* complex, the river drainage basins in which it occurs, and the broad regions referred to in the text. NE Coast: 1, Normanby River; 2, Endeavour River; 3, Barron River; 4, Mulgrave–Russell Rivers; 5, Johnstone River; 6, Murray River (Qld); 7, Ross River; 8, Haughton River; 10, Burdekin River; E Coast: 9, Proserpine River; 11, Styx/Pioneer rivers; 12, Fitzroy River; 13, Boyne River; 14, Baffle Creek; 15, Kolan River; 16, Burnett River; 17, Mary River (Qld); SE Coast: 18, Pine River; 19, Brisbane River; 20, Logan–Albert rivers; 21, Tweed River; 22, Richmond River; 23, Clarence River; 24, Bellinger River; 25, Macleay River; 26, Hastings River; 27, Hawkesbury River; LEB: 28, Diamantina River; 29, Cooper Creek; Bulloo: Bulloo River and Lake Bancannia; MDB: 30, Warrego River; 31, Condamine–Culgoa rivers; 32, Border Rivers; 33, Gwydir River; 34, Namoi River; 35, Macquarie–Bogan rivers; 36, Paroo River; 37, Lachlan River; 38, Murrumbidgee River; 39, Murray River (incl. Upper Murray, Kiewa, Ovens, Broken, Goulburn, Campaspe, Loddon, Avoca rivers and the Riverina); 40, Lower Murray. Red dots are collection locations for both SNP and mtDNA samples—see Supporting Information Tables S1 and S2 for exact locations and disaggregation of sample types

indicated that the samples taken from within this single drainage lacked substantive structure. The four individuals from Sandgate Lagoons, Brisbane [66], were removed from the analysis because this urban site is known to have been populated with *Emydura* from the pet trade. This sampling covers the full range of what is currently regarded most broadly as *Emydura macquarii* (Georges & Thomson, 2010; Turtle Taxonomy Working Group, 2017). Our sampling target for the ingroup taxa was 10 individuals per population (not always achieved). Outgroup species consisted of *E. victoriae* (n = 5) from the Victoria River (Northern Territory) and *E. subglobosa worrelli* (n = 5) from the Roper River (NT). A further 128 individuals were collected from 105 waterbodies across the same range for mtDNA sequencing—although these two data sets were not collected for the same specimens, they cover the same range and are virtually all from the same river basins. A full list of specimens and localities is provided in Supporting Information (Tables S1 and S2). Our taxonomy follows that of Georges and Thomson (2010). Nomenclature for drainage basins follows that of the Australian Drainage Divisions and River Basins (Auslig, 2001). We typically refer to these drainages (e.g., Hunter River drainage) and subdrainages (e.g., Gwydir River subdrainage of the Murray–Darling drainage) by abbreviation, for example, Hunter River and Gwydir River, respectively. Fitzroy River, Mary River and Mitchell River refer to the Queensland drainages of Auslig (2001).

Tissue samples comprised a small sliver of skin tissue taken from the trailing edge of the clawless toe of the hind foot in most cases, preserved in 95% ethanol and stored at −20°C, or frozen in liquid nitrogen and stored at −80°C. In some cases, blood was sampled from the jugular vein using a 23-gauge needle and syringe, also preserved in 95% ethanol and held at −20°C. A few samples, drawn from earlier studies (Georges & Adams, 1992, 1996), comprised muscle or blood, snap frozen in liquid nitrogen and stored at −80°C.

2.2 | DNA extraction and sequencing

DNA was extracted by Diversity Arrays Technologies (DArT Pty Ltd, Canberra, Australia) using a NucleoMag 96 Tissue Kit (Macherey-Nagel, Düren, Germany) coupled with NucleoMag SEP (Ref. 744900) to allow automated separation of high-quality DNA on a Freedom Evo robotic liquid handler (TECAN, Männedorf, Switzerland). Tissue was first incubated for four hours (blood) or overnight (skin) with proteinase K, adjusted in concentration depending on the tissue.

Sequencing for SNP genotyping was done using DArTseq™ (DArT Pty Ltd) which uses a combination of complexity reduction using restriction enzymes, implicit fragment size selection and next-generation sequencing (Sansaloni et al., 2011), as described in detail by Kilian et al. (2012). The technique is similar to double-digest restriction-associated DNA sequencing (ddRAD) (Peterson et al., 2012), but has the advantages of accepting lower quantities of DNA, greater tolerance of lower quality DNA and yielding lower allelic dropout rates (Sansaloni et al., 2011). To achieve the most appropriate complexity reduction (the fraction of the genome represented, controlling average read depth and number of polymorphic loci), four combinations of restriction enzymes (PstI enzyme combined with either HpaII, SphI, NspI or MseI) were evaluated and restriction enzyme combination of PstI (recognition sequence 5′-CTGCA|G-3′) and SphI (5′-GCATG|C-3′) was selected.

DNA samples were processed in digestion/ligation reactions, but replacing a single PstI-compatible adaptor of Kilian et al. (2012) with two different adaptors annealed to the two restriction enzyme overhangs. The PstI-compatible adapter included the Illumina flow cell attachment sequence, a sequencing primer sequence, a barcode region of variable length (see Elshire et al., 2011) and the PstI-compatible overhang sequence. The reverse adapter contained flow cell attachment sequence and SphI-compatible overhang sequence. Only fragments generated by the PstI-SphI double digest were effectively amplified in 30 rounds of polymerase chain reaction (PCR). Amplifications consisted of an initial denaturation step of 94°C for 1 min, followed by 30 cycles of PCR with the following temperature profile: denaturation at 94°C for 20 s, annealing at 58°C for 30 s and extension at 72°C for 45 s, with an additional final extension at 72°C for 7 min. After PCR, equimolar amounts of amplification products from each sample were pooled and applied to cBot (Illumina) bridge PCR for sequencing on the Illumina Hiseq 2500. The sequencing (single end) was run for 77 cycles.

For mitochondrial sequencing, DNA was extracted using a salting-out protocol (Sambrook & Russell, 2001) or standard phenol–chloroform procedure (Kocher et al., 1989). Mitochondrial Cytb and surrounding tRNAs were amplified using primers CytbG (Spinks, Shaffer, Iverson, & McCord, 2004) and Thr-8mod (GGTTTACAAGACCAGKGTTT, modified from Spinks et al., 2004). Amplification of the control region used primers EmyTHR (CACCACCCTCCTGAAATACTC) and CR500mod (CCCTGAAAAAACAACCAGAGGCC). PCR products were shipped to Agencourt Biosciences (La Jolla, CA 92039, USA) for sequencing. Sequences were edited and assembled with SEQUENCHER 4.8 (Gene Codes Corporation, Ann Arbor, MI 48108, USA). Sequences from both genes were combined for each individual, and unique haplotypes were identified in COLLAPSE 1.2 (Posada, 2004). Coding sequences were checked for unexpected frameshift errors or stop codons using MEGA 6 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). Autapomorphies were verified by examining original chromatograms.

2.3 | SNP genotyping

Sequences generated from each lane were processed using proprietary DArT Pty Ltd analytical pipelines. Poor quality sequences were first filtered, applying more stringent selection criteria to the barcode region compared to the rest of the sequence (minimum barcode Phred score 30, pass percentage 75; minimum whole-read Phred score 10, pass percentage 50). In that way, assignment of the sequences to specific samples in the sample disaggregation step was very reliable. Approximately 2,000,000 (+7%) sequences per sample were identified and used in marker calling. These sequences were truncated to 69 bp (including some adaptor sequence where the fragments were shorter than 69 bp) and aggregated into clusters by the DART fast clustering algorithm, taking advantage of the fixed fragment length, with a Hamming distance threshold of 3 bp. The sequences were error-corrected using an algorithm that corrects a low-quality base (Phred score <20) to a corresponding high-quality singleton tag (Phred score >25); where there was more than one distinct high-quality tag, the sequence with the low-quality base was discarded. Identical sequences were then collapsed. These error-corrected sequences were analysed using DART software (DArTsoft14) to output candidate SNP markers. In brief, SNP markers were identified within each cluster by examining parameters calculated for each sequence across all samples—primarily average and variance of sequencing depth, the average counts for each SNP allele and the call rate (proportion of samples for which the marker is scored). Where three sequences survived filtering to this point, the two variants with the highest read depth were selected. The final average read depth per locus was 21×. One third of samples were processed twice from DNA, using independent adaptors, to allelic calls as technical replicates, and scoring consistency (repeatability) was used as the main selection criterion for high-quality/low error rate markers. The DART analysis pipelines have been tested against hundreds of controlled crosses to verify Mendelian behaviour of the resultant SNPs as part of their commercial operations.

The resultant data set contained the SNP genotypes and various associated metadata of which CloneID (unique identity of the sequence tag for a locus), repAvg (proportion of technical replicate assay pairs for which the marker score is identical), avgPIC (polymorphism information content averaged over the reference and alternate SNPs) and SnpPosition (position in the sequence tag at which the defined SNP variant base occurs) are of particular relevance to our analyses.

2.4 | Additional SNP filtering

The SNP data and associated metadata were read into a genlight object ({adegenet}, Jombart, 2008) to facilitate processing with package DARTR (Gruber, Unmack, Berry, & Georges, 2018). Only loci with 100% repeatability (repAvg) were chosen for subsequent analysis. Further filtering was undertaken on the basis of call rate (98% unless otherwise specified). A SNP at a locus in a particular individual may not be called during genotyping because of a mutation at one or both of the restriction enzyme recognition sites. Because of the high read depth (c. 21×), most "missing data" are of this form. Finally, we filtered out secondary SNPs where they occurred in a single sequenced tag, retaining only one SNP at random. Any monomorphic loci arising as a result of the removal of individuals or populations were also deleted. Given the low within-population sample sizes ($n \leq 11$), we did not filter loci for departures from Hardy–Weinberg equilibrium (HWE) or linkage disequilibrium. We regard the data remaining after this additional filtering (c. 16,000 SNP markers) as highly reliable.

2.5 | Visualization

Genetic similarity among individuals and populations was visualized using ordination (principal coordinates analysis or PCoA, Gower, 1966) as implemented in the gl.pcoa and gl.pcoa.plot functions of DARTR. A scree plot of eigenvalues guided the number of informative axes to examine (Cattell, 1966), taken in the context of the average percentage variation explained by the original variables (using the gl.pcoa.scree function in DARTR).

2.6 | Genetic diversity

Observed heterozygosity was used as a measure of relative genetic diversity. Heterozygosity was obtained for each population from allele frequencies using the gl.report.heterozygosity function of DARTR.

2.7 | Fixed difference analysis

A fixed difference between two populations at a locus occurs when the populations share no alleles at that locus. Accumulation of fixed differences between two populations is a robust indication of lack of gene flow. The fixed difference analysis was undertaken using the nuclear SNP data with the gl.collapse.recursive function in DARTR (Gruber et al., 2018). Populations (i.e., field sampling sites) with low sample sizes (≤3) were amalgamated with other populations from the same drainage before analysis—those within each of the Hunter River, Bellinger Coast, Clarence River and Normanby River. Populations showing evidence of contemporary admixture and hybridization or introgression were eliminated from the analysis—populations from the Paroo [108–110] and Bulloo [104,105,107] rivers which lie intermediate to the MDB and LEB, from Lake Dunn [94] close to the border of the LEB and the Burdekin drainage, Hoods Lagoon [32] in the Fitzroy drainage adjacent to the Burdekin drainage, and four individuals from the Barron [10] and Mulgrave-Russell [11] rivers. These cases were analysed separately as described in the section on admixture. Fixed differences were summed over populations taken pairwise, and when two populations had no fixed differences, they were combined and the process repeated until there was no further reduction (Georges & Adams, 1996). The resultant OTUs are, by definition, putatively diagnosable at one or more SNP loci.

While the decision to amalgamate two populations can be made with certainty, the separation of two populations by one or more fixed differences is subject to sampling error. False positives may arise because of the finite sample sizes involved. Simulations (Supporting Information Appendix S1) were used to estimate the expected false positive rate in pairwise comparisons. Putative OTUs with an observed count of fixed differences not significantly different from the expected rate of false positives were further amalgamated to yield a final set of diagnosable OTUs. This was achieved using gl.collapse.pval() in the DARTR package.

This approach differs from the more conventional approach using STRUCTURE (Pritchard, Stephens, & Donnelly, 2000) in the emphasis placed on fixed allelic differences and diagnosability of resultant taxa. STRUCTURE and related analyses define aggregations such that each is in Hardy–Weinberg and linkage equilibrium across all loci. Under these criteria, STRUCTURE will define two populations as different even if they have only modest differences in allele frequencies that would bring them out of equilibrium if combined. Populations that are regarded as distinct by STRUCTURE would not necessarily be diagnosable, a criterion that is central to our method. STRUCTURE also struggles with small sample sizes, whereas our fixed difference approach addresses this issue.

2.8 | Contemporary admixture

Evidence of migration between populations was initially qualitatively assessed by examining the relationships among individuals, identified to population, in PCoA plots. Contemporary admixture was evident as individuals occupying intermediate positions between clearly divergent aggregations of populations. A second approach was to use TREEMIX version 1.13 (Pickrell & Pritchard, 2012) which estimates a bifurcating maximum-likelihood tree for the populations, and then examines the stress between the covariance matrix computed from the tree and the covariance matrix computed from the data. The stress is then reduced by adding putative migration events. Suspected instances of contemporary admixture of individuals were then assessed using NewHybrids (Anderson & Thompson, 2002). NewHybrids is limited to c. 200 loci. We selected a set of 200 loci from loci that were (a) fixed and different or (b) most divergent in terms of allele frequency profiles, when comparing the two nominated parental populations. NewHybrids was run with default parameter settings, the Jeffreys Prior for θ, and 10,000 sweeps. The analysis was replicated 100 times with a random selection of 200 loci, satisfying conditions (a) and (b) above, to confirm that the outcome was not sensitive to the selection of loci used.

2.9 | Phylogeny

We took two approaches to estimate the phylogeny of southern *Emydura*. First, we conducted a Fitch–Margoliash distance analysis, as implemented in PHYLIP 3.695 (Felsenstein, 1989), and applied to the Euclidean distances between OTUs identified in the fixed difference analysis. Bootstrap estimates of node reliability were obtained by randomly resampling loci and recalculating the distance matrix, with 1,000 replicates.

The second approach used SVDquartets analysis (Chifman & Kubatko, 2014, 2015), chosen because of our short reads (<69 bp) and the single variable sites per locus (Chou et al., 2015). Heterozygous SNP positions were represented by standard ambiguity codes (see Felsenstein, 2004:255). SVDquartets takes unlinked multilocus data for subsets of taxa, taken four at a time (quartets), and assigns a score to each of the three possible unrooted topologies for each quartet. The topology with the lowest score is selected as the best supported topology for that quartet. The final set of quartets is combined (Reaz, Bayzid, & Rahman, 2014) to estimate the species tree. We used the implementation of SVDquartets in PAUP* (version 4.0a162; Swofford, 2003) with parameters evalQuartets = random, bootstrap = standard, nreps = 10,000 and ambigs = distribute. Trees were rooted with *E. victoriae* and *E. subglobossa worrelli* (Supporting Information Table S1; Spinks, Georges, & Shaffer, 2015). Instances of historical migration were identified by TREEMIX version 1.13 (Pickrell & Pritchard, 2012).

Phylogenetic analyses of mitochondrial sequences were performed with maximum likelihood (ML) using GARLI 2.0 (Zwickl, 2006). Control region sequences were aligned with the online version of MAFFT 7.046 (Katoh & Standley, 2013) using the G-INS-i algorithm with the scoring matrix for nucleotide sequences set to 1PAM/*K* = 2, a gap opening penalty of 1.53 and an offset value of 0.5. The best partition scheme and models of molecular evolution based on the Akaike information criterion with correction (AICc) in PARTITIONFINDER 2.0 (Lanfear, Calcott, Ho, & Guindon, 2012) using PHYML (Guindon et al., 2010) were TRN+I+G for control region and TVM+G, HKY and GTR for the first through third codon positions for Cytb. We ran GARLI with 10 search replicates with the following default setting values changed: streefname = random; attachmentspertaxon = 130; genthreshfortopoterm = 100,000; scorethreshforterm = 0.05; significanttopochange = 0.00001. For bootstrapping, we ran 1,000 replicates with the previous settings with the following changes: genthreshfortopoterm = 10,000; significanttopochange = 0.01; treerejectionthreshold = 20 as suggested in the GARLI manual to speed up bootstrapping. Trees were rooted with *E. victoriae* and *E. tanybaraga* (Supporting Information Table S2; Spinks et al., 2015).

3 | RESULTS

3.1 | SNP data set

A total of 91,384 polymorphic SNP loci were scored for 402 individuals of the southern *Emydura* from 57 waterbodies of the coastal and inland drainages of eastern Australia, and 10 individuals from two waterbodies for outgroup taxa *E. victoriae* and *E. subglobosa worrelli*. After stringent filtering on repeatability (repAvg = 1.0) and call rate (0.98), the number of SNP loci in the data set dropped to 65,240 and

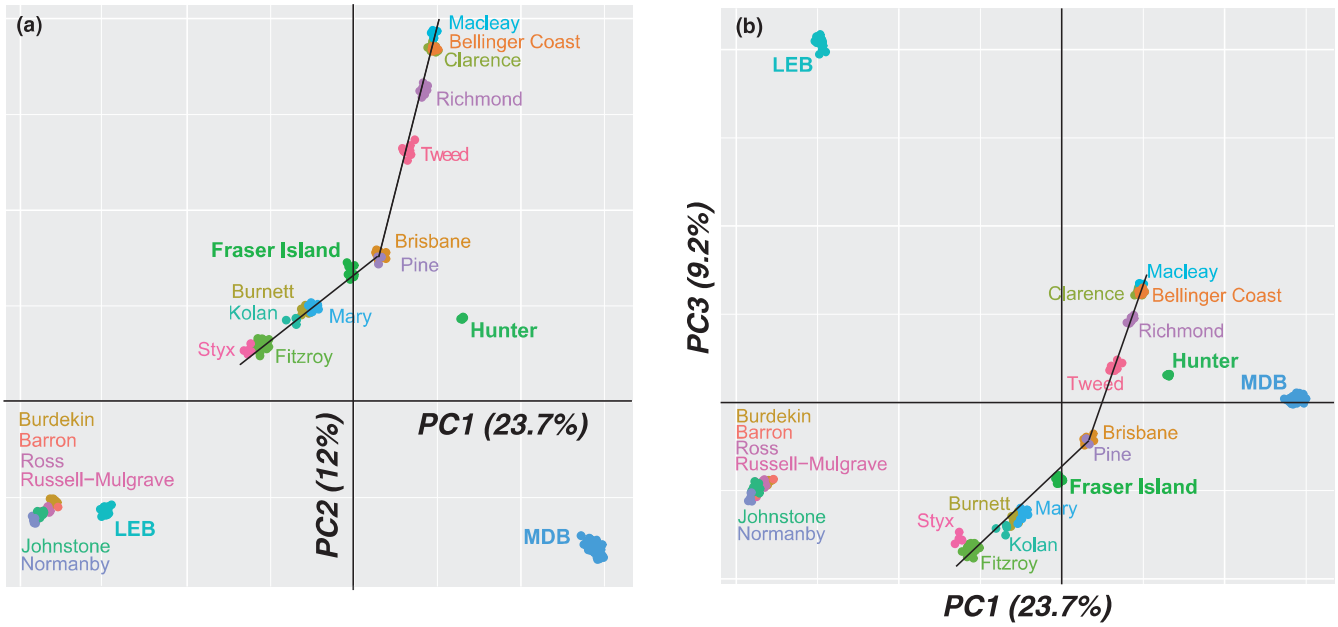

FIGURE 3 Graphical representation of genetic similarity between individuals using principal coordinates analysis. (a) Principal coordinate axis 2 vs. 1; (b) axis 3 vs. 1. A total of 44.9% of variation were explained by the first three ordinated axes. A coastal cline from the Macleay through to the Pine, then on to the Styx is highlighted with a line. Axes not to scale

then 19,096, respectively. A total of 2,837 secondary SNPs were filtered. One specimen, the sole individual from the Albert River (location 69), had a call rate of less than 80% (a threshold set taking into account the presence of the outgroups) and was removed from the data set. Monomorphic loci (n = 18) arising from the removal of the Albert River specimen were deleted. The resultant data set including outgroups is referred to as the full data set, comprising 16,241 polymorphic SNP loci from 56 ingroup populations (n = 5–11, N = 401) and two outgroup species (n = 10).

For the ingroup taxa only, 77,998 SNP loci were polymorphic for the 401 individuals, of which 53,970 had a repeatability of 1.0. Filtering on call rate (0.98) reduced the number of loci to 18,141. The removal of 1,960 secondary SNPs at a locus yielded what we refer to as the ingroup data set, comprising 16,181 polymorphic SNP loci from 56 populations (n = 401).

3.2 | Qualitative analysis

Preliminary analysis of the data with PCoA applied to the ingroup data set revealed clear evidence of contemporary admixture in waterbodies close to the boundaries of key drainages: between the MDB and the LEB via the Paroo River (MDB) [location labels 108–110 of Supporting Information Table S1] and the intervening Bulloo River [104,105,107]; between the LEB at Lake Dunn [94] and the Burdekin River (NE Coast); and between the coastal Fitzroy River at Hoods Lagoon [32] (E Coast) and the Burdekin River (NE Coast). These admixed populations were removed and considered in a separate analysis (see below), and the PCoA repeated on the remaining 15,183 loci scored for 335 individuals from 48 populations (Figure 3).

The most obvious feature of the PCoA plot is the clear representation of a coastal geographic series of populations extending from the Macleay River in the south (SE Coast) to the Styx River in the north (E Coast), with geographic drainage adjacency maintained in the genetic representation. This isolation-by-distance pattern strongly suggests some level of recent or contemporary allelic exchange between adjacent drainages. There was a significant discontinuity in the composition of SNP loci contributing to the observed genetic distances between the southeast populations and the populations to the north of the Brisbane–Pine rivers cluster, presumably reflecting some impediment to gene flow greater than between other adjacent drainages in the series. There is also a distinct gap, representing absent or exceptionally low gene flow, between the central coastal populations of the Fitzroy/Styx drainages (E Coast) and the populations in the Burdekin River and rivers north to the Normanby (NE Coast). The coastal Hunter River population is drawn out from this coastal series towards the inland populations of the MDB. When the first three dimensions of the ordination are considered (Figure 3), the MDB and the LEB form two distinct clusters, with the LEB now distinct from the NE Coast populations.

The northeast series ranging from the Burdekin to the Normanby shows evidence of gene flow across the Great Dividing Range in the vicinity of the Barron River and the Mulgrave–Russell basins (Supporting Information Figure S1). Four individuals are drawn out from the linear series of populations of the NE Coast by variation at loci not contributing to variation among the remaining individuals of the NE Coast. The population contributing this suite of novel alleles has not been sampled in this study, but is presumably the unsampled *Emydura* species found in the Mitchell River to the west.

3.3 | Genetic diversity

Average heterozygosity was 0.038 ± 0.0019 (0.007–0.060, n = 48) after taking out outlier populations. The Hunter River was a significant outlier, low in heterozygosity (0.007–0.008, $X^2 = 4.76$, $p < 0.05$), and, reflecting the above-mentioned introgression, the Barron River was a significant outlier, high in heterozygosity (0.089, $X^2 = 11.42$, $p < 0.001$).

3.4 | Fixed difference analysis

Both of the outgroup taxa, *E. victoriae* and *E. subglobosa*, emerged as diagnosable in the fixed difference analysis applied to the full data set, differing significantly ($p < 0.0001$) from each other by 1,272 fixed differences (7% of loci). The southern *Emydura* OTUs differed from the northern Australian species by between 2,469 and 4,024 fixed differences (13%–22% of loci), consistent with the treatment of the northern species as outgroup taxa (Table 1).

Five diagnosable OTUs emerged from the fixed difference analysis (Table 2 and Supporting Information Table S4). Populations from the MDB and LEB, both of which are west of the Great Dividing Range, each emerged as diagnosable, and there was a single diagnosable OTU taking in coastal populations from the Macleay River in the south to the Styx River in the north, including the Fraser Island populations (Figure 2). Populations from the Burdekin River in the south to the Normanby River in the north formed a fourth diagnosable OTU, with the genetically distinctive population in the Hunter River identified as the fifth OTU. This is in broad agreement with the structure evident in the PCoA plot (Figure 3).

3.5 | Contemporary admixture

In the NewHybrids analysis, populations from the Bulloo River and Paroo River showed evidence of admixture of SNP alleles from the MDB (of which, the Paroo is intermittently a part) and the LEB (Table 3, Figure 4). Populations from these drainages were selected from the full data set, and additional filtering on call rate applied (0.995) to yield 8,572 polymorphic loci from 191 individuals from 22 populations. Eulbertie Waterhole [101] was selected as the parental population for the LEB; Lake Forbes [Lachlan River, 125] and Goondiwindi Weir [Border Rivers, 121] were selected as representing the parental genotypes for the MDB as they were geographically distant from the drainages in which admixture was occurring. All individuals from the MDB (excluding the Paroo) were assigned to the expected parental class, and all

of the individuals from the LEB were assigned to their expected parental class (Table 3, except three individuals from Lake Dunn—see below). Individuals with genotypes consistent with LEB genotypes were present in the Bulloo River (n = 3), and individuals with genotypes consistent with the MDB genotypes were present in the Paroo River (n = 20). In addition, in the Bulloo River there was an F1 hybrid, four F2 crosses and eleven backcrosses to the LEB. Two individuals from Toonborough (Paroo) showed indications of admixture (Figure 4) but at a lower level that was not scored by NewHybrids. The Paroo River contained an F1 hybrid, an F2 individual and eight backcrosses to the MDB genotype (excluding Paroo). The three exceptions from Lake Dunn in the LEB that were not assigned to the LEB initially scored in NewHybrids as backcrosses between an F1 LEBxMDB crossed individual and the LEB, but on closer examination, two of these were F1 hybrids between the LEB and the adjacent Burdekin River and one a backcross to the LEB. When represented in a principal coordinates plot (Figure 4), the presence of gene flow across the Paroo and Bulloo rivers and between the LEB and the Burdekin is quite evident and broadly consistent with the results of the NewHybrids analysis.

Finally, there was evidence in the NewHybrids analysis of admixture of individuals from the Fitzroy River (six locations) and Burdekin River (one location) via Hoods Lagoon [32] in the Fitzroy River, close to the boundary of the two drainages. Of the six individuals in this lagoon, there was an F1 hybrid, two F2 individuals, one backcross to the Burdekin and two backcrosses to the Fitzroy (Supporting Information Table S5).

Analysis using TREEMIX (Supporting Information Figure S2) supported and complemented the results evident in the PCoA, identifying suspected migration events between the Paroo River of the MDB [108–110] and the Bulloo River [104,105,107] in both directions, from the coastal Burdekin River [19] (NE Coast) to the Fitzroy River at Hoods Lagoon [32] (E Coast), and from the Brisbane [67]/Pine rivers [65] to the Tweed River [70]. There was evidence of migration between the Barron River [10] and the stem lineages of the outgroup taxon *E. victoriae*, supporting the source of introgression predicted from the PCoA (Supporting Information Figure S1). TREEMIX did not identify migration events between the Burdekin River and

TABLE 2 Matrix of Euclidean genetic distances (above diagonal) and fixed genetic differences (below diagonal) between the final set of operational taxonomic units to arise from a fixed difference analysis applied to the ingroup data set

	MDB	LEB	Hunter	EC&SEC	NEC
MDB	0	31.0	33.8	26.3	40.4
LEB	385	0	36.1	30.5	42.5
Hunter	471	1304	0	22.4	32.6
EC&SEC	3	165	311	0	31.4
NEC	155	275	908	22	0

Notes. Comparisons were based on an average of 12,730 loci after filtering for call rate >98%. All fixed differences were significant at $p < 0.0001$.
EC: East Coast; Hunter: Hunter River, NSW; LEB: Lake Eyre Basin; MDB: Murray–Darling Basin; NEC: Northeast Coast; Normanby: Normanby River; SEC: Southeast Coast.

TABLE 1 Counts of loci with fixed differences and Euclidean genetic distances between outgroup taxa (*Emydura sublogobosa worrelli* and *E. victoriae*) and the operational taxonomic units that arose from a fixed difference analysis applied to the full data set

	MDB	LEB	Hunter	EC&SEC	Fraser Is. [Nth]	Fraser Is. [Sth]	NEC	Normanby	*E. subglobosa*	*E. victoriae*
Fixed differences										
E. subglobosa	2,971	3,638	4,024	2,509	3,358	3,483	3,280	3,627	0	1,272
E. victoriae	2,915	3,594	3,970	2,469	3,303	3,437	3,250	3,600	1,272	0
Eucl. distances										
E. subglobosa	67.8	70.2	71.2	65.1	68.6	69.2	69.0	70.2	0.0	47.8
E. victoriae	67.3	69.8	71.0	64.7	68.2	68.8	68.8	70.1	47.8	0.0

Notes. Comparisons were based on an average of 19,746 loci after filtering for call rate >98%. All nonzero fixed differences were significant ($p < 0.0001$).
EC: East Coast; Hunter: Hunter River, NSW; LEB: Lake Eyre Basin; MDB: Murray–Darling Basin; NEC: Northeast Coast; Normanby: Normanby River; SEC: Southeast Coast.

TABLE 3 Analysis of contemporary admixture between the Lake Eyre Basin (LEB) and the Murray–Darling Basin (MDB) via the intervening the Bulloo and Paroo rivers

Population	LEB	F1	F2	F1xLEB	F1xMDB	MDB	*N*
101_LEBCoopEulb	**0.9999**	**0.0000**	**0.0001**	**0.0000**	**0.0000**	**0.0000**	**9**
96_LEBCoopAvin	0.9999	0.0000	0.0001	0.0000	0.0000	0.0000	10
102_LEBCoopCull	0.9999	0.0000	0.0001	0.0000	0.0000	0.0000	10
94_LEBCoopDunn	0.9999	0.0000	0.0001	0.0000	0.0000	0.0000	8
105_BULBullQuil	0.9999	0.0000	0.0001	0.0000	0.0000	0.0000	3
105_BULBullQuil	0.0001	0.0000	0.0002	0.9997	0.0000	0.0000	3
107_BULBullGumb	0.9999	0.0000	0.0001	0.0000	0.0000	0.0000	1
107_BULBullGumb	0.0001	0.0000	0.0002	0.9997	0.0000	0.0000	3
104_BULBullComo	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	1
104_BULBullComo	0.0000	0.0000	0.9867	0.0133	0.0000	0.0000	3
104_BULBullComo	0.0000	0.0000	0.9603	0.0397	0.0000	0.0000	1
104_BULBullComo	0.0001	0.0000	0.3077	0.6922	0.0000	0.0000	1
104_BULBullComo	0.0002	0.0000	0.0002	0.9996	0.0000	0.0000	5
108_MDBParoBoot	0.0000	0.0000	0.0001	0.0000	0.0000	0.9999	9
109_MDBParoEulo	0.0000	0.0000	0.0001	0.0000	0.0000	0.9999	4
109_MDBParoEulo	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	1
109_MDBParoEulo	0.0000	0.0000	0.0001	0.0000	0.9998	0.0000	5
110_MDBParoToon	0.0000	0.0000	0.0001	0.0000	0.0000	0.9999	7
110_MDBParoToon	0.0000	0.0000	0.0988	0.0000	0.9012	0.0000	1
110_MDBParoToon	0.0000	0.0000	0.9999	0.0001	0.0000	0.0000	1
110_MDBParoToon	0.0000	0.0001	0.0001	0.0000	0.9998	0.0000	2
111_MDBWarrLang	0.0000	0.0000	0.0001	0.0000	0.0000	0.9999	1
112_MDBWarrAmba	0.0000	0.0000	0.0001	0.0000	0.0000	0.9999	2
113_MDBWarrDart	0.0000	0.0000	0.0001	0.0000	0.0000	0.9999	10
114_MDBWarrSanf	0.0000	0.0000	0.0001	0.0000	0.0000	0.9999	10
115_MDBWarrBiny	0.0000	0.0000	0.0001	0.0000	0.0000	0.9999	10
117_MDBCondArch	0.0000	0.0000	0.0001	0.0000	0.0000	0.9999	10
123_MDBGwydBing	0.0000	0.0000	0.0001	0.0000	0.0000	0.9999	9
124_MDBMacqCudg	0.0000	0.0000	0.0001	0.0000	0.0000	0.9999	10
127_MDBMurrAlbu	0.0000	0.0000	0.0001	0.0000	0.0000	0.9999	10
131_MDBMurrMBri	0.0000	0.0000	0.0001	0.0000	0.0000	0.9999	10
125_MDBLachForb	**0.0000**	**0.0000**	**0.0001**	**0.0000**	**0.0000**	**0.9999**	**10**
121_MDBBordGoon	**0.0000**	**0.0000**	**0.0001**	**0.0000**	**0.0000**	**0.9999**	**10**

Notes. Nominated parental populations for the LEB and MDB are shown in bold. Posterior probabilities (NewHybrids, Anderson & Thompson, 2002) assigning individuals to one of the parental populations, F1 hybrids, F2 hybrids and the corresponding backcrosses are shown as shaded values. *N* = sample sizes for each classification. Three individuals from Lake Dunn [94] that represent admixture with the Burdekin River, Qld, are omitted.

Lake Dunn in the LEB, despite evidence for this in the PCoA, and one individual with an F1 hybrid profile and two putative backcrosses to the LEB among the 11 individuals sampled from Lake Dunn (LEB).

The mitochondrial sequences also provided evidence of gene flow across drainage divides. Two haplotypes are of relevance to the admixture in the intervening drainages between the MDB and the LEB (Figure 6). One is present in the Bulloo River, Paroo River (MDB) and Warrego (MDB) River, with a minor variant found in the Condamine River and Border Rivers (MDB; Supporting Information Table S4). The second is found in the Diamantina River (LEB), Barcoo River (LEB), Cooper Creek (LEB) and Bulloo River (Supporting Information Table S4). This is evidence of introgression of the mitochondrial haplotype into the Bulloo River from both the MDB and LEB.

3.6 | Phylogeny

The phylogeny, based on Euclidean distance (Table 2) calculated from the SNPs for the five OTUs arising from the fixed difference analyses (i.e., excluding the admixed populations), was very robust, with 99.8%–100% bootstrap support for all nodes but one (Figure 5b).

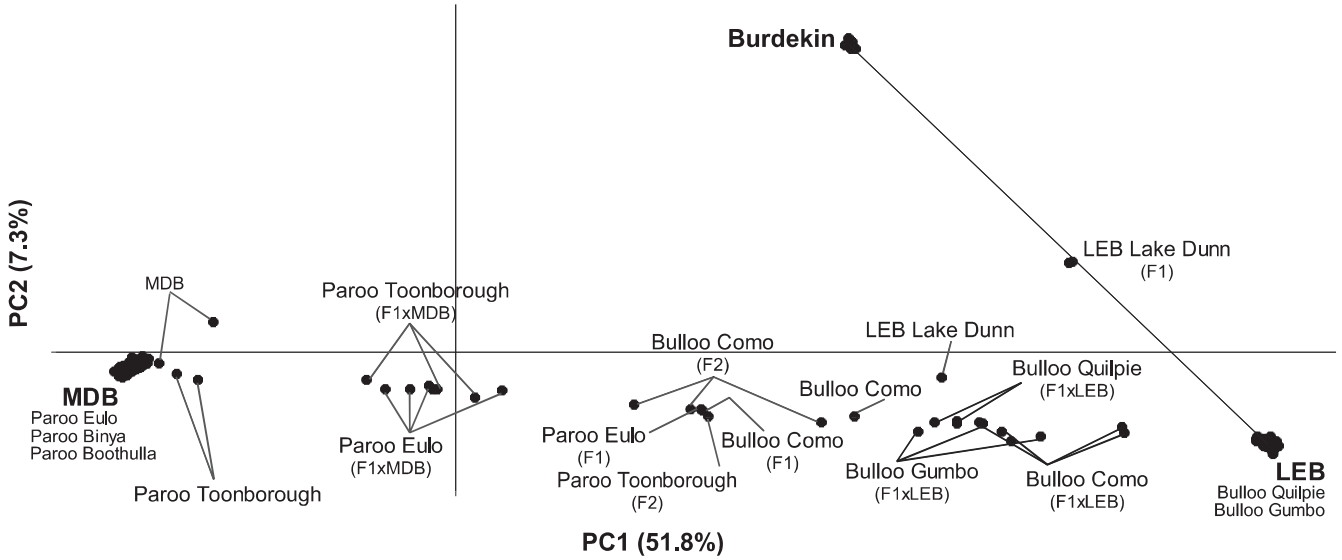

FIGURE 4 A principal coordinates analysis plot of individuals from the Murray–Darling Basin (MDB), Lake Eyre Basin (LEB), Bulloo River and Paroo River, showing their assignments based on posterior probabilities generated by NewHybrids (Anderson & Thompson, 2002). The individuals labelled Toonborough (Paroo River) were not identified as admixed individuals by NewHybrids but appear to show some level of introgression with the MDB. The individual labelled Como (Bulloo River) was identified by NewHybrids, but its assignment on the basis of posterior probabilities was ambiguous (Table 2). Evidence of hybridization between the Burdekin River and LEB genotypes in Lake Dunn of the LEB is also shown

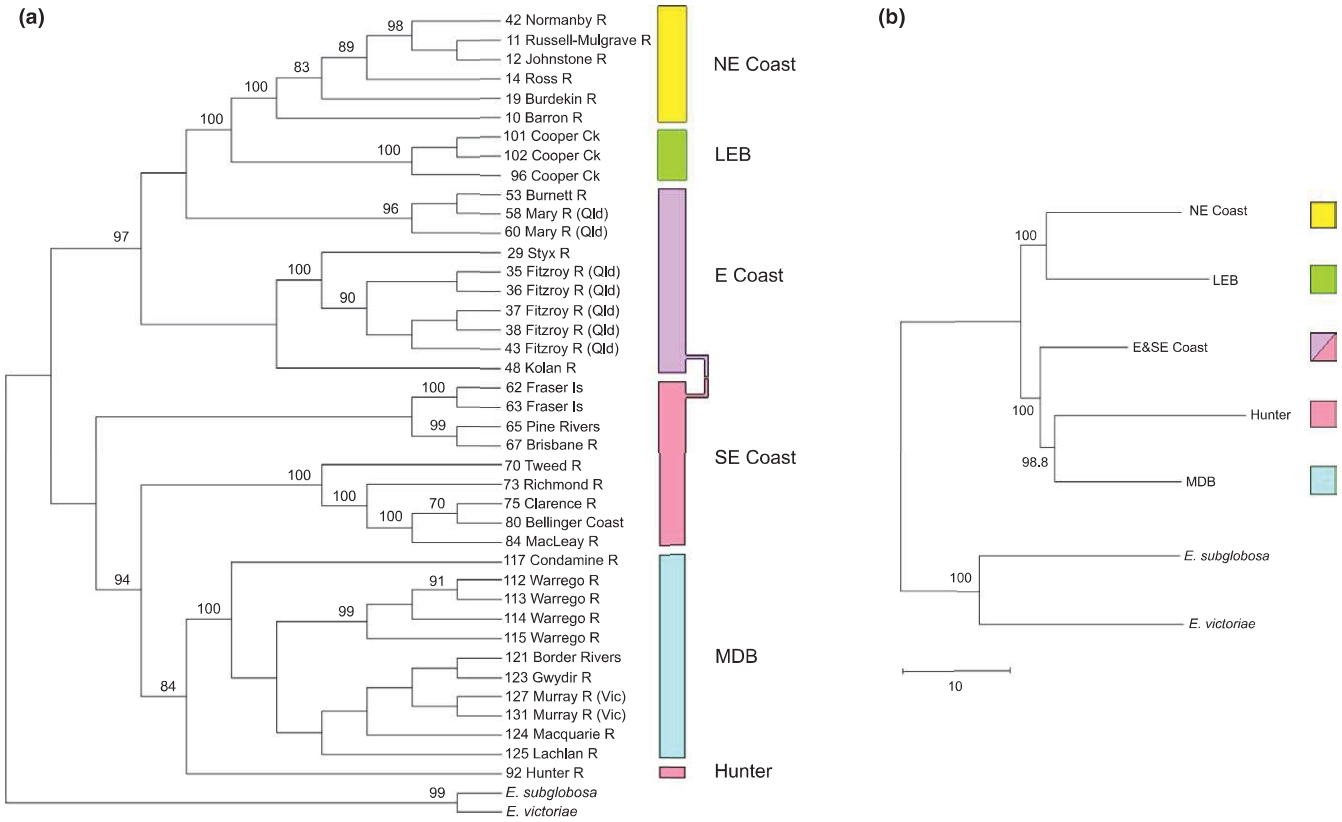

FIGURE 5 Phylogenies for the southern *Emydura* based on (a) SVD Quartets applied to SNP genotypes with ambiguity codes substituted for heterozygous sites; and (b) Fitch–Margoliash maximum likelihood applied to distances between the diagnosable units identified by the fixed difference analysis

The NE Coast OTU is sister to the LEB OTU with 100% bootstrap support. The Hunter River OTU is sister to the MDB OTU also with 99.8% bootstrap support, and SE Coast sister to that MDB/Hunter clade, with 100% bootstrap support.

The SVDquartets tree applied to populations (sampling sites) without evidence of admixture generated a tree with strong bootstrap support across many nodes (Figure 5a). Each of the NE Coast, LEB and MDB OTUs identified in the fixed difference analysis emerged as strongly supported clades (100% bootstrap support), and the Hunter River OTU was supported as a lineage sister to the MDB (84% bootstrap support). The relationships among populations of SE Coast and E Coast were less well resolved. The southern rivers of the SE Coast (Macleay, Bellinger Coast, Clarence, Richmond and Tweed) formed a well-supported NSW clade (sites 70, 73, 75, 80, 84; 100% bootstrap support) to the exclusion of the Brisbane/Pine rivers and Fraser Island populations to the north from the same OTU. The Brisbane River and Fraser Island populations were not resolved within a polytomy including the MDB and NSW clade. The Fitzroy/Styx rivers (100% bootstrap support) and the Mary/Burnett rivers (90%) each formed well-supported clades within the E Coast OTU

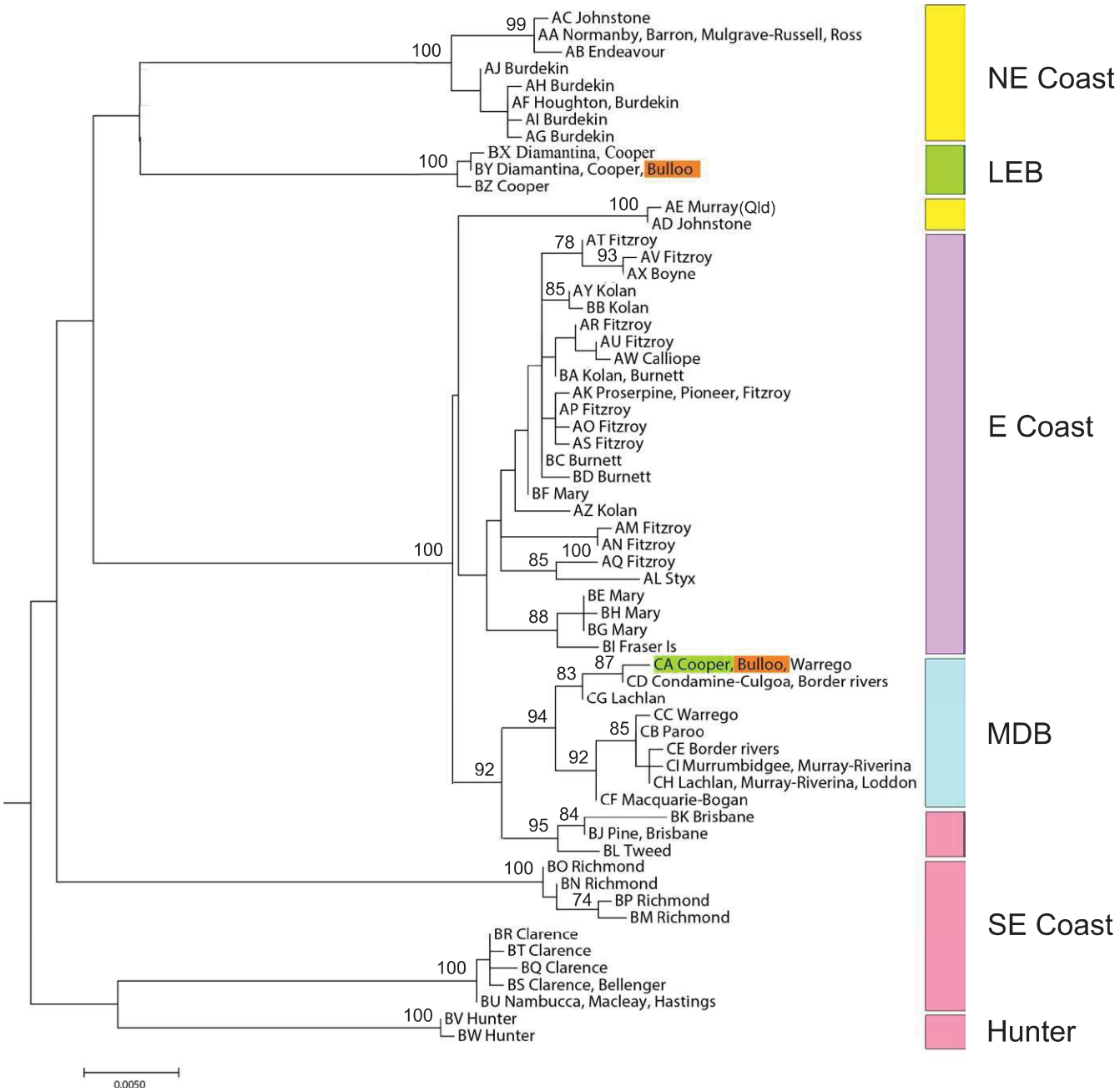

FIGURE 6 A mitochondrial phylogeny for the southern *Emydura* based on control region and *Cytb* sequence. The phylogeny was generated using maximum likelihood in GARLI 2.0 (Zwickl, 2006). Specimen label abbreviations: Haplotype, followed by River(s). Refer to Supporting Information Tables S1 and S3 for locality data

identified in the fixed difference analysis, but the relationship of the Mary/Burnett clade was unresolved within a polytomy involving the Fitzroy/Styx clade and the NE Coast/LEB clade.

The mtDNA phylogeny drew from 456 bp of aligned control region sequence and 1,050 bp of cytb sequence. Excluding outgroups, 1,315 of the 1,506 bp were constant, 24 variable characters were parsimony uninformative, and 167 characters were parsimony informative. ML recovered one tree with an ln score of −4492.8346 (Figure 6). A number of clades were well-supported, including most of those from the nuclear DNA analyses. Disregarding a few recalcitrant haplotypes (discussed above under admixture), the NE Coast (100%), LEB (100%), MDB (94%) and Hunter River haplotypes (100%) formed well-supported clades. As with the nuclear phylogeny, population of the SE Coast and E Coast OTUs was poorly resolved. A southern NSW clade comprised the Macleay, Bellinger Coast and Clarence rivers (100%), but to the exclusion of a Richmond River clade (100%) to the immediate north. Tweed River haplotypes, whose population was part of a clade within the SE Coast in the nuclear phylogeny, formed a well-supported clade (95%) with the Brisbane and Pine rivers haplotypes. The haplotypes from the E Coast OTU formed a clade, but without bootstrap support. The deeper relationships among these groups were generally poorly resolved with the exception of a strong E Coast–MDB sister group relationship that was not recovered in the nuclear data.

4 | DISCUSSION

In this paper, we have taken the position that the delimitation of species necessarily requires assessment of where a putative taxon lies on the trajectory of speciation and that speciation itself can involve an interplay between dichotomous lineage formation (cladogenesis) and reticulating processes of recurrent gene flow (tokogenesis) that retard and obscure lineage divergence. With comprehensive geographic sampling, we used an approach that allows more explicit studies of gene flow at the boundaries of putative species, to yield more nuanced, and ultimately more reasonable, conclusions on taxonomic status. We believe our approach will be of wide applicability, informing decisions on species delimitation that are more conservative than defining closely related "species" on typological grounds (Cann & Sadlier, 2017) or simply naming substantive lineages in a mitochondrial (Figure 6) or nuclear phylogeny (Figure 5). By adding a step of defining diagnosable OTUs before invoking lineage considerations, our approach may place a check on taxonomic inflation that has plagued the southern *Emydura* and other systems (Chan et al., 2017; Frankham et al., 2012; Garnett & Christidis, 2017; Isaac, Mallet, & Mace, 2004; Sukumaran & Knowles, 2017).

4.1 | Species and speciation

We embarked on this study anticipating that we would clarify species boundaries within southern *Emydura* and elucidate the historical relationships among them. After all, past morphological studies, albeit with inadequate quantitation, have identified multiple species in this complex, including *E. macquarii* restricted to the MDB, *E. signata* in the Brisbane River and adjacent SE Coast drainages, *E. nigra* from Fraser Island, *E. krefftii* from the Mary River (E Coast) to the Normanby River (NE Coast) and *E. emmotti* from the LEB (Cann, 1998; Cann & Sadlier, 2017; Cogger et al., 1983; Goode, 1967) plus numerous named subspecies (Cann, 1998; Cann & Sadlier, 2017; McCord, Cann, & Joseph-Ouni, 2003) (Figure 1). Evidence from allozymes contrasted with this view. Many of the above-mentioned species shared alleles at all loci, even rare alleles (Georges & Adams, 1996). The allozyme data provided evidence that the southern *Emydura* is a single polytypic species comprising a series of reasonably well-defined but closely related ESUs (Cogger, 2014; Georges & Thomson, 2010; Todd, Blair, & Jerry, 2014). We anticipated that the large numbers of nuclear SNP markers and comprehensive spatial sampling of sequenced individuals would resolve species boundaries in the southern *Emydura*, and so resolve the disparity arising from the morphological and allozyme studies. Instead, we have uncovered a dynamic system of lineages progressing on a trajectory towards complete speciation but held back by episodic exchange of alleles across drainage divides on various timescales (Table 4). Our analysis and interpretation support the proposition that *E. macquarii* is a polytypic species complex comprising a set of allopatric lineages and incipient species at various, relatively shallow levels of molecular and morphological divergence. Individuals from parapatric populations but representing different lineages can interbreed and do so when and where their ranges episodically come into contact. We interpret these as a "syngameon of semispecies" (*sensu* Grant, 1981:74) rather than as fully speciated.

Several lineages, and therefore candidate species, exist in the southern *Emydura* (evident following Steps 1–3 in our empirical approach to species delimitation). If one were to apply the subjective judgement necessary in cases of allopatry (our Step 4) to devise a defensible classification for the complex below the level of species then, in the interests of utility, an arrangement consistent with both the nuclear and mitochondrial evidence would recognize five taxa—(a) LEB; (b) MDB; (c) Hunter River; (d) SE and E Coast rivers from the Macleay River in the south to the Styx River in the north (including Fraser Island, recognized as two separate clades of unresolved affinities in the mtDNA tree); and (e) NE Coast rivers from the Burdekin River in the south to the Normanby River in the north (Figures 2, 5 and 6). Each of these five OTUs is supported by the qualitative PCoA, the fixed difference analysis and high bootstrap support for their integrity as clades in the independent phylogenetic analyses of nuclear and mitochondrial data. One could argue that the SE Coast and E Coast taxon could be further split on the basis of the discontinuity that occurs in the Brisbane-Pine rivers region, as is evident in the PCoA (Figure 3), and the nuclear sequence phylogenetic analysis. However, in applying the fourth step of our species delimitation strategy, the admittedly subjective judgement to devise a defensible classification for the complex, we regard these five or six as ESUs

TABLE 4 Portals between drainages that potentially explain evidence for dispersal of individuals across drainage divides and the patterns of contemporary episodic gene flow/hybridization and introgression we observed, and historical hybridization and introgression inferred

Portal name	Drainages	Landform	Type	Latitude/longitude	Elevation (m asl)	Divide width (km)	Height above drainage (m)	Last inundated	Reference
Biboohra	Mitchell (Qld)–Barron	Alluvial plain	Parallel channel	16.9S, 145.4E	400	3	20	Active for last ~2 Ma	Whitehead and Nelson (2014)
Whippo-Wippa	Bulloo–LEB	Alluvial plain, ephemeral chain of ponds	Anticline divide	28.2S, 142.5E	125	3	<10	Holocene	Ingram and Senior (1970)
Bulloo terminus	Bulloo–LEB	Dunefield	Tectonically defeated drainage	30.4S, 141.8E	110	15	20	Pre-Pleistocene	Dury (1973)
Bindegolly	Bulloo–Paroo	Dunefield with interdune lakes	Fault and monocline divide	27.8S, 144.2E	150	15	5	Holocene	Senior (1971) and Timms (2006)
Cuttaburra Creek	MDB (Warrego)–Paroo	Distributary channel	Alluvial fan	29.3S, 145.1E	130	0	0	Contemporary	Balcombe et al. (2006)
Jericho	LEB–Burdekin	Alluvial plain	Laterite uplands	23.5S, 146.2E	370	10	10	Holocene	Vine and Doutch (1972)
Ulan	MDB–Hunter	Erosional ridge	Bedrock uplands	32.3S, 149.7E	490	2	40	Neogene (volcanic derangement) or early Pleistocene (climatic)	Tomkins and Hesse (2004) and Wellman and McDougall (1974)
Rugby	Fitzroy–Burdekin	Alluvial plain, gilgai depressions	Bedrock uplands	21.9S, 147.8E	250	10	10	Holocene	Malone, Corbett, and Jensen (1964)

Notes. Reference to Holocene indicates that, while there is no clear evidence of standing water on the divide from satellite records, the landscape between defined channels on either side of the portal is flat enough to potentially permit standing water to exist on the drainage divide during exceptional rain events.

or incipient rather than full species because of (a) the exceptionally low levels of divergence among them compared to that between the southern *Emydura* species complex and *E. victoriae* and *E. subglobosa* to the north; (b) evidence of low-level episodic gene flow among them via adjacent boundary populations (e.g., the Bulloo and Paroo populations connecting the LEB and MDB via infrequently inundated gateways or portals—Table 4); and (c) their often parapatric rather than overlapping distributions, which suggests that, as taxa, they are unable to establish noninterbreeding populations in sympatry despite opportunity. As with many other organismal groups, we cannot say whether these ESUs are destined to become full species with the passage of time. This will depend on future climates—a protracted climatic regime of low precipitation and river flow that eliminates episodic gene flow may ultimately see the taxa speciate; a wetter climatic regime or human-induced dispersal may see continuation of the cycle of speciation and despeciation or perhaps homogenization. Our view is that the complex of allopatric and parapatric ESUs and incipient species we have identified is not sufficient to challenge the view prevailing in some quarters (Cogger, 2014; Georges & Thomson, 2010) of a single polytypic species, *E. macquarii*, referred to in the vernacular as the southern *Emydura*.

We recognize that some researchers might disagree and regard our five (or more) ESUs as full species, subspecies or a combination of species, subspecies and ESUs. Where it deemed necessary, as a practical convenience rather than a reflection of biological reality, one could assign to species, the MDB OTU (*E. macquarii* (Gray 1830)), the LEB OTU (*E. emmotti* Cann et al., 2003), the Hunter River OTU (*E. gunabarra* Cann, 1998) and a single coastal species *E. krefftii* Gray, 1871 with subspecies for the NE Coast (no available name), E Coast (*E. krefftii krefftii*, as defined here), SE Coast (*E. krefftii signata* Ahl 1932) and Fraser Island OTUs (*E. krefftii nigra* McCord et al., 2003). Whatever operational decision is taken, it will be controversial, as the decision involves considerable subjectivity in imposing a classification on what is essentially a dynamic, sometimes reticulate, and incomplete evolutionary process (Kindler et al., 2017).

4.2 | Mechanisms of genetic diversification

The genetic structure evident in the southern *Emydura* is driven by a number of potential factors. The Australian Great Dividing Range separates the eastern coastal rivers from those flowing to the west—the Murray–Darling and Lake Eyre basins—though for a substantial part of its length, the Great Dividing Range is a "cartographic myth" (Ollier, 1979); effectively, the range is nothing more than a drainage divide comprising relatively low hills that even a freshwater turtle might traverse (e.g., via the Biboohra and Jericho portals of Table 4). In the coastal drainages, the most important factors appear related to drainage adjacency; whereby, resistance to dispersal between this linear array of drainages is afforded by their drainage divides (Todd et al., 2014; Unmack, 2001). A history of recurrent interconnection or persistent isolation during sea level change (+1 to −135 m in the Pleistocene) depends on the extent of the adjacent continental shelf (Unmack, Hammer, Adams, Johnson, & Dowling, 2013). There is evidence in the fixed difference analysis of a genetic discontinuity associated with the tongue of aridity extending from the inland to the coast in the Fitzroy–Burdekin region, that is, in the vicinity of the so-called Burdekin gap (Todd et al., 2014; Unmack, 2001; Wong, Keogh, & McGlashen, 2004). There is also evidence in allele frequencies of a genetic discontinuity between the Brisbane-Pine and the Mary rivers. An exceptionally narrow coastal plain between the northernmost drainage, the Normanby and those drainages to the south affords resistance to dispersal that is reflected in the genetic signature of Supporting Information Figure S1. For the inland drainages (west of the Great Dividing Range), avenues for historical and contemporary dispersal between coastal rivers and those of the MDB across the Great Dividing Range are important (e.g., the Ulan portal, Table 4), as are avenues for dispersal between the MDB and LEB via the intervening Paroo River and Bulloo River (via several portals and the intermittent Cuttacutta Creek, Table 4). In many of these cases, exchange of individuals is likely to be episodic, a consequence of the highly variable and unpredictable Austral climate. During wetter periods on timescales varying from the Pleistocene to the present, rivers with exceptionally low elevation across their drainage divides will interconnect (Table 4).

The Hunter taxon warrants special mention, being was the most distinctive OTU in the southern *Emydura* to emerge from the fixed difference analysis. These populations did not fall into the coastal series of populations in the PCoA (from the Macleay River to the Styx River), but rather fell between those and the MDB populations. The Hunter River OTU was sister to the MDB OTU in the phylogenetic analysis of SNP data, but had unresolved affinities in the mtDNA analysis. The Hunter River populations also had the lowest heterozygosity, suggesting a founder effect or protracted bottleneck. While there is no evidence of contemporary dispersal between the Hunter River and the MDB, the data suggest some level of historical exchange of alleles. The Hunter River contains at least two fish species found also in the MDB but not in coastal rivers immediately north or south of the Hunter River (Crowley & Ivantsoff, 1990; Unmack, 2001), indicating that there has been movement of fishes between these basins, presumably via similar avenues (Table 4) used by freshwater turtles.

5 | CONCLUDING REMARKS

Understanding the evolutionary history of diversifying lineages, spatially and temporally, remains a major challenge in evolutionary biology. Vicariance plays an important role in speciation, but seldom involves absolute isolation. Divergence through isolation by distance and imperfect barriers to dispersal, as observed in our study, is likely to be typical of many "speciation events" (Jackson et al., 2017). Most importantly, attempts at species delineation involving allopatric lineages are best conducted under explicit

considerations of the possible reversal of speciation and reticulation through episodic dispersal across those barriers (Edwards et al., 2016; Hoelzel, 2016).

Our results support Sukumaran and Knowles (2017), who argued that (a) speciation is a (potentially protracted) process, not an event, (b) this complicates analyses that take a phylogenetic lineage approach to species delimitation, and (c) methods are needed that accommodate this complexity in deciding which lineages are species as opposed to structure within species. We elaborate on these concerns. Phylogenetic approaches recover a bifurcating pattern of ancestry and descent and do not accommodate protracted processes of episodic exchange of alleles and resultant partial or complete reticulation in ancestry and descent that we have observed and inferred for the southern *Emydura*. The phylogeny for the diagnosable taxa (Figure 5b) assumes that they have been on independent evolutionary trajectories since their inception in that it assumes there is an underlying bifurcating history to uncover. In our context, using phylogeny for decisions on species delimitation is likely to be highly misleading—speciation is often not a bifurcating process (Morrison, 2016). There have been numerous approaches to deal with these issues (see Table 2 of Singhal, Hoskin, Couper, Potter, & Moritz, 2018), and we need a more rigorous framework to assess the taxonomic status of significant lineages uncovered by increasingly sophisticated molecular data and analyses (Singhal et al., 2018; Sukumaran & Knowles, 2017). Such a framework will need to incorporate data that considers contemporary, possibly episodic, gene flow as well as lineage divergence and historical patterns of ancestry and descent among populations.

One of the few downsides of the age of molecular phylogenetics is that our ability to generate simple gene and lineage bifurcating phylogenies from ever-increasingly complex genetic and genomic data sets has tended to reduce our focus on one of the most fundamental concepts in species delimitation, that of diagnosability. Our strategy requires that unequivocally diagnosable OTUs are identified early in the analysis of species boundaries. As such, it differs from the conventional lineage approach (de Queiroz et al., 1998) in that we place our threshold at the time and place where fixed allelic differences have evolved, rather than immediately at the point of initial lineage divergence, which is always difficult to clearly identify given the often observed gene tree/species tree disparity. We thus differ in placing greater emphasis on diagnosability (fixed differences) rather than on divergence per se (which draws also from allele frequency differences). In addition, our strategy adds an additional level of objectivity to species delimitation. We believe that this strategy makes a potentially important contribution to any rigorous framework for judgements on species boundaries.

ACKNOWLEDGEMENTS

We would like to thank those who contributed samples to the University of Canberra Wildlife Tissue Collection, and in particular, Colin Limpus, Tony Tucker, Enzo Guarino, Duncan Limpus, Scott Thomson and Kate Hodges. We also thank Jason Carling for explaining the filters in the DART analysis pipelines, David Berman for discussions on the approach to analysis and Bruno Ferronato for comments on the manuscript. The study was conducted with funding from the Australian Collaborative Research Network for Murray–Darling Basin Futures (AG, BG, PJU), the CRC for Freshwater Ecology (AG), the Australian Research Council (DP-A19600367: AG, HBS; LP140100521: PJU, BG, DW, AG and AK) and the US National Science Foundation (DEB 1457832: HBS).

DATA ACCESSIBILITY

All mtDNA sequences used in this study are deposited in GenBank, Accession Nos. MH730305–MH730434, and the sequence alignment is deposited in Dryad, https://doi.org/10.5061/dryad.t2c668d. The SNP data and r scripts that were used in the analysis are also deposited in Dryad under the same doi.

CONFLICT OF INTEREST

Author Andzrej Kilian declares a potential conflict of interest in that he is the Director of Diversity Arrays Technology Pty Ltd, the company that generated the SNP data on a cost recovery basis. The remaining authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

A.G. and H.B.S. conceived the project and the sampling regime and obtained initial funding from the Australian Research Council to undertake the work. A.G. and P.J.U. subsequently recast the project scope and design and obtained funding from the CRN for Murray-Darling Futures and the ARC Linkage Program (LP140100521). B.G. contributed expertise in spatial analysis and R, and D.W. contributed to the geomorphologic interpretations. A.K. guided the analyses to generate the SNP data, M.Y. and X.Z. undertook the laboratory work for the SNP analysis, and G.P. undertook the laboratory work for the mitochondrial genome sequencing. M.Y., P.U., H.B.S. and A.G. undertook the field work. A.G. and P.J.U. led the writing of the manuscript to which all authors contributed.

ORCID

Arthur Georges https://orcid.org/0000-0003-2428-0361
Bernd Gruber https://orcid.org/0000-0003-0078-8179
Peter J. Unmack https://orcid.org/0000-0003-1175-1152

REFERENCES

Anderson, E. C., & Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, *160*, 1217–1229.

Auslig (2001). *Australian drainage divisions and river basins*. Canberra, ACT: Commonwealth of Australia. https://doi.org/10.1186/1471-2164-16-S10-S2

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, *3*(10), e3376. https://doi.org/10.1371/journal.pone.0003376

Balcombe, S. R., Arthington, A. H., Foster, N. D., Thoms, M. C., Wilson, G. G., & Bunn, S. E. (2006). Fish assemblages of an Australian dryland river: Abundance, assemblage structure and recruitment patterns in the Warrego River, Murray-Darling Basin. *Marine and Freshwater Research*, *57*, 619–633. https://doi.org/10.1071/MF06025

Bush, G. L. (1993). A reaffirmation of Santa Rosalia, or why are there so many kinds of small animals? In D. Edwards & D. R. Rees (Eds.), *Evolutionary patterns and process* (pp. 229–249). New York, NY: Academic Press.

Cann, J. (1998). *Australian freshwater turtles*. Singapore, Singapore: Beaumont Publishing.

Cann, J., & Sadlier, R. (2017). *Freshwater turtles of Australia*. Melbourne, Vic.: CSIRO Publishing.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Chan, K. O., Alexander, A. M., Grismer, L. L., Su, Y.-C., Grismer, J. L., Quah, E. S. H., & Brown, R. M. (2017). Species delimitation with gene flow: A methodological comparison and population genomics approach to elucidate cryptic species boundaries in Malaysian Torrent Frogs. *Molecular Ecology*, *26*, 5435–5450. https://doi.org/10.1111/mec.14296

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, *30*, 3317–3324. https://doi.org/10.1093/bioinformatics/btu530

Chifman, J., & Kubatko, L. (2015). Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *Journal of Theoretical Biology*, *374*, 35–47. https://doi.org/10.1016/j.jtbi.2015.03.006

Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S., & Warnow, T. (2015). A comparative study of SVDquartets and other coalescent-based species tree estimation methods. *BMC Genomics*, *16*(Suppl 10), S2, 1–11.

Cogger, H. G. (2014). *Reptiles and amphibians of Australia* (7th ed.). Canberra, ACT: CSIRO Publishing.

Cogger, H. G., Cameron, E. E., & Cogger, H. M. (1983). *Zoological catalogue of Australia Vol I: Amphibia and reptilia*. Canberra, ACT: Australian Government Publishing Service.

Crowley, L. E. L. M., & Ivantsoff, W. (1990). A review of species previously identified as *Craterocephalus eyresii* (Pisces: Atherinidae). *Proceedings of the Linnean Society of NSW*, *112*, 87–103.

Dobzhansky, T. (1951). *Genetics and the origin of species* (3rd ed.). New York, NY: Columbia University Press.

deQueiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*, 879–886. https://doi.org/10.1080/10635150701701083

de Queiroz, K., Howard, D. I., & Berlocher, S. H. (1998). The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations. In D. J. Howard & S. H. Berlocher (Eds.), *Endless forms: Species and speciation* (pp. 57–75). New York, NY: Oxford University Press.

Dury, G. H. (1973). Paleohydrologic implications of some pluvial lakes in northwestern New South Wales, Australia. *Geological Society of America Bulletin*, *84*, 3663. https://doi.org/10.1130/0016-7606(1973)84<3663:PIOSPL>2.0.CO;2

Edwards, S. V., Potter, S., Schmitt, C. J., Bragg, J. G., & Moritz, C. (2016). Reticulation, divergence, and the phylogeography–phylogenetics continuum. *Proceedings of the National Academy of Sciences*, *113*, 8025–8032. https://doi.org/10.1073/pnas.1601066113

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*, e19379. https://doi.org/10.1371/journal.pone.0019379

Felsenstein, J. (1989). PHYLIP – Phylogeny inference package (version 3.2). *Cladistics*, *5*, 164–166.

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

Frankham, R., Ballou, J. D., Dudash, M. R., Eldridge, M. D. B., Fenster, C. B., Lacy, R. C., … Ryder, O. A. (2012). Implications of different species concepts for conserving biodiversity. *Biological Conservation*, *153*, 25–31. https://doi.org/10.1016/j.biocon.2012.04.034

Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A., & Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution*, *27*, 480–488. https://doi.org/10.1016/j.tree.2012.04.012

Garnett, S. T., & Christidis, L. (2017). Taxonomy anarchy hampers conservation. The classification of complex organisms is in chaos. *Nature*, *546*, 26–27.

Georges, A., & Adams, M. (1992). A phylogeny for Australian chelid turtles based on allozyme electrophoresis. *Australian Journal of Zoology*, *40*, 453–476. https://doi.org/10.1071/ZO9920453

Georges, A., & Adams, M. (1996). Electrophoretic delineation of species boundaries within the short-necked freshwater turtles of Australia (Testudines: Chelidae). *Zoological Journal of the Linnean Society*, *118*, 241–260. https://doi.org/10.1111/j.1096-3642.1996.tb01266.x

Georges, A., & Thomson, S. (2010). Diversity of Australasian freshwater turtles, with an annotated synonymy and keys to species. *Zootaxa*, *2496*, 1–37.

Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L., & Buerkle, C. A. (2012). Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species: Genomics of speciation. *Evolution*, *66*, 2167–2181. https://doi.org/10.1111/j.1558-5646.2012.01587.x

Goode, J. (1967). *Freshwater tortoises of Australia and New Guinea (in the family Chelidae)*. Melbourne, Vic.: Lansdowne Press.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, *53*, 325–338. https://doi.org/10.1093/biomet/53.3-4.325

Grant, V. (1981). *Plant speciation*. New York, NY: Columbia University Press.

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartR: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, *18*, 691–699. https://doi.org/10.1111/1755-0998.12745

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, *59*, 307–321. https://doi.org/10.1093/sysbio/syq010

Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, *25*, 2454–2466. https://doi.org/10.1111/mec.13582

Hoelzel, A. R. (2016). The road to speciation runs both ways. *Science*, *354*, 414–415. https://doi.org/10.1126/science.aaj2007

Hoskin, C. J., Higgie, M., McDonald, K. R., & Moritz, C. (2005). Reinforcement drives rapid allopatric speciation. *Nature*, *437*, 1353–1356. https://doi.org/10.1038/nature04004

Ingram, J. A., & Senior, D. A. (1970). *Bulloo 1:250,000 map sheet, Australia*. Canberra, ACT: Bureau of Mineral Resources.

Isaac, N., Mallet, J., & Mace, G. M. (2004). Taxonomic inflation: Its influence on macroecology and conservation. *Trends in Ecology and Evolution*, *19*, 464–469. https://doi.org/10.1016/j.tree.2004.06.004

Jaccoud, D., Peng, K., Feinstein, D., & Kilian, A. (2001). Diversity arrays: A solid state technology for sequence information independent genotyping. *Nucleic Acids Research*, *29*, e25–e25. https://doi.org/10.1093/nar/29.4.e25

Jackson, N. D., Carstens, B. C., Morales, A. E., & O'Meara, B. C. (2017). Species delimitation with gene flow. *Systematic Biology*, *66*, 799–812.

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*, 772–780. https://doi.org/10.1093/molbev/mst010

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., … Uszynski, G. (2012). Diversity arrays technology: A generic genome profiling technology on open platforms. *Methods in Molecular Biology*, *888*, 67–89. https://doi.org/10.1007/978-1-61779-870-2

Kindler, C., Chèvre, M., Ursenbacher, S., Böhme, W., Hille, A., Jablonski, D., … Fritz, U. (2017). Hybridization patterns in two contact zones of grass snakes reveal a new Central European snake species. *Scientific Reports*, *7*, 7378. https://doi.org/10.1038/s41598-017-07847-9

Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Pääbo, S., Villablanca, F. X., & Wilson, A. C. (1989). Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences*, *86*, 6196–6200. https://doi.org/10.1073/pnas.86.16.6196

Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*, 1695–1701. https://doi.org/10.1093/molbev/mss020

Malone, E. J., Corbett, D., & Jensen, A. R. (1964). *Geology of the Mount Coolon 1:250,000 sheet area*. Canberra, ACT: Bureau of Mineral Resources.

Mao, X., Zhang, S., & Rossiter, S. J. (2016). Differential introgression suggests candidate beneficial and barrier loci between two parapatric subspecies of Pearson′s horseshoe bat *Rhinolophus pearsoni*. *Current Zoology*, *62*, 405–412. https://doi.org/10.1093/cz/zow017

Maynard Smith, J. (1989). Trees, bundles or nets. *Trends in Ecology and Evolution*, *4*, 302–304. https://doi.org/10.1016/0169-5347(89)90026-8

Mayr, E. (1963). *Animal species and evolution*. Cambridge, MA: Belknap Press. https://doi.org/10.4159/harvard.9780674865327

McCord, W. P., Cann, J., & Joseph-Ouni, M. (2003). A taxonomic assessment of *Emydura* (Testudines: Chelidae) with descriptions of new subspecies from Queensland, Australia. *Reptilia*, *27*, 59–63 [60].

Metzker, M. L. (2010). Sequencing technologies – The next generation. *Nature Reviews Genetics*, *11*, 31–46. https://doi.org/10.1038/nrg2626

Morrison, D. A. (2016). Genealogies: Pedigrees and phylogenies are reticulating networks not just divergent trees. *Evolutionary Biology*, *43*, 456–473. https://doi.org/10.1007/s11692-016-9376-5

Ollier, C. D. (1979). Evolutionary geomorphology of Australia and Papua: New Guinea. *Transactions of the Institute of British Geographers*, *4*, 516–539. https://doi.org/10.2307/622212

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*, e1002967. https://doi.org/10.1371/journal.pgen.1002967

Posada, D. (2004). Collapse ver. 1.2. A tool for collapsing sequences to haplotypes. Retrieved from http://darwin.uvigo.es.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.

Reaz, R., Bayzid, M. S., & Rahman, M. S. (2014). Accurate phylogenetic tree reconstruction from quartets: A heuristic approach. *PLoS ONE*, *9*(8), e104008. https://doi.org/10.1371/journal.pone.0104008

Sambrook, J., & Russell, D. W. (2001). *Molecular cloning: A laboratory manual* (3rd ed.). New York, NY: Cold Spring Harbor Laboratory.

Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., & Kilian, A. (2011). Diversity Arrays Technology (DArT) and next-generation sequencing combined: Genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. *BMC Proceedings*, *5*(Suppl 7), P54. https://doi.org/10.1186/1753-6561-5-S7-P54

Senior, B. R. (1971). *Eulo, Qld. 1: 250 000 geological series explanatory notes. Sheet SH/55-1*. Canberra, ACT: Bureau of Mineral Resources.

Singhal, S., Hoskin, C. J., Couper, P., Potter, S., & Moritz, C. (2018). A framework for resolving cryptic species: A case study from the lizards of the Australian wet tropics. *Systematic Biology*, *67*, 1061–1075.

Sites, J. W., & Marshall, J. C. (2003). Delimiting species: A Renaissance issue in systematic biology. *Trends in Ecology and Evolution*, *18*, 462–470. https://doi.org/10.1016/S0169-5347(03)00184-8

Spinks, P. Q., Georges, A., & Shaffer, H. B. (2015). Phylogenetic uncertainty and taxonomic re-revisions: An example from the Australian short-necked Turtles (Testudines: Chelidae). *Copeia*, *103*, 536–540. https://doi.org/10.1643/CH-14-147

Spinks, P. Q., Shaffer, H. B., Iverson, J. B., & McCord, W. P. (2004). Phylogenetic hypotheses for the turtle family Geoemydidae. *Molecular Phylogenetics and Evolution*, *32*, 164–182. https://doi.org/10.1016/j.ympev.2003.12.015

Stuglik, M. T., & Babik, W. (2016). Genomic heterogeneity of historical gene flow between two species of newts inferred from transcriptome data. *Ecology and Evolution*, *6*, 4513–4525. https://doi.org/10.1002/ece3.2152

Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences*, *114*, 1607–1612. https://doi.org/10.1073/pnas.1607921114

Swofford, D. L. (2003). *PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4*. Sunderland, MA: Sinauer Associates.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*, 2725–2729. https://doi.org/10.1093/molbev/mst197

Timms, B. V. (2006). The geomorphology and hydrology of saline lakes of the middle Paroo, arid-zone Australia. *Proceedings of the Linnean Society of New South Wales*, *127*, 154–174.

Todd, E. V., Blair, D., & Jerry, D. R. (2014). Influence of drainage divides versus arid corridors on genetic structure and demography of a widespread freshwater turtle, *Emydura macquarii krefftii*, from Australia. *Ecology and Evolution*, *4*(5), 606–622. https://doi.org/10.1002/ece3.968

Tomkins, K. M., & Hesse, P. P. (2004). Evidence of Late Cenozoic uplift and climate change in the stratigraphy of the Macquarie River valley, New South Wales. *Australian Journal of Earth Sciences*, *51*, 273–290. https://doi.org/10.1111/j.1440-0952.2004.01057.x

Turtle Taxonomy Working Group (2017). Turtles of the world: Annotated checklist and atlas of taxonomy, synonymy, distribution, and conservation status (8th Edition). *Conservation Monographs*, *7*, 1–292.

Unmack, P. J. (2001). Biogeography of Australian freshwater fishes. *Journal of Biogeography*, *28*, 1053–1089. https://doi.org/10.1046/j.1365-2699.2001.00615.x

Unmack, P. J., Hammer, M. P., Adams, M., Johnson, J. B., & Dowling, T. E. (2013). The role of continental shelf width in determining freshwater phylogeographic patterns in south-eastern Australian pygmy

perches (Teleostei: Percichthyidae). *Molecular Ecology*, *22*, 1683–1699. https://doi.org/10.1111/mec.12204

van Tassell, C. P., Smith, T. P. L., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., … Sonstegard, T. S. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods*, *5*, 247–252. https://doi.org/10.1038/nmeth.1185

Vine, R., & Doutch, H. (1972). *Galilee SF 55–10 map sheet explanatory notes. Australia 1: 250,000 geological series*. Canberra, ACT: Bureau of Mineral Resources.

Wellman, P., & McDougall, I. (1974). Cainozoic igneous activity in eastern Australia. *Tectonophysics*, *23*, 49–65. https://doi.org/10.1016/0040-1951(74)90110-3

Whitehead, P. W., & Nelson, P. N. (2014). Displacement of the Great Divide in north Queensland associated with Neogene lava flows. *Australian Journal of Earth Sciences*, *61*, 803–809. https://doi.org/10.1080/08120099.2014.927788

Wong, B. B. M., Keogh, J. S., & McGlashen, D. J. (2004). Current and historical patterns of drainage connectivity in eastern Australia inferred from population genetic structuring in a widespread freshwater fish *Pseudomugil signifer* (Pseudomugilidae). *Molecular Ecology*, *13*, 391–401. https://doi.org/10.1046/j.1365-294X.2003.02085.x

Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*, 851–865. https://doi.org/10.1046/j.1420-9101.2001.00335.x

Zwickl, D. J. (2006). Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion (PhD Thesis). University of Texas at Austin, Texas, USA.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Georges A, Gruber B, Pauly GB, et al. Genomewide SNP markers breathe new life into phylogeography and species delimitation for the problematic short-necked turtles (Chelidae: *Emydura*) of eastern Australia. *Mol Ecol*. 2018;00:1–19. https://doi.org/10.1111/mec.14925