Received: 27 February 2023 | Revised: 24 December 2023 | Accepted: 4 January 2024

DOI: 10.1111/mec.17266

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Unravelling the mystery of endemic versus translocated populations of the endangered Australian lungfish (*Neoceratodus forsteri*)

Roberto Biello[1,2] | Silvia Ghirotto[1] | Daniel J. Schmidt[3] | Silvia Fuselli[1] | David T. Roberts[4] | Tom Espinoza[5] | Jane M. Hughes[3] | Giorgio Bertorelle[1]

[1]Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy

[2]Department of Crop Genetics, John Innes Centre, Norwich, UK

[3]Australian Rivers Institute, Griffith University, Brisbane, Queensland, Australia

[4]Seqwater, Ipswich, Queensland, Australia

[5]Burnett Mary Regional Group, Bargara, Queensland, Australia

Correspondence
Giorgio Bertorelle, Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy.
Email: ggb@unife.it

Funding information
Ministero dell'Università e della Ricerca, Grant/Award Number: 201794ZXTL

Handling Editor: Maren Wellenreuther

[Correction added on 6 January 2025, after first online publication: The copyright line was changed.]

Abstract

The Australian lungfish is a primitive and endangered representative of the subclass Dipnoi. The distribution of this species is limited to south-east Queensland, with some populations considered endemic and others possibly descending from translocations in the late nineteenth century shortly after European discovery. Attempts to resolve the historical distribution of this species have met with conflicting results based on descriptive genetic studies. Understanding if all populations are endemic or some are the result of, or influenced by, translocation events, has implications for conservation management. In this work, we analysed the genetic variation at three types of markers (mtDNA genomes, 11 STRs and 5196 nuclear SNPs) using the approximate Bayesian computation (ABC) algorithm to compare several demographic models. We postulated different contributions of Mary River and Burnett River gene pools into the Brisbane River and North Pine River populations, related to documented translocation events. We ran the analysis for each marker type separately, and we also estimated the posterior probabilities of the models combining the markers. Nuclear SNPs have the highest power to correctly identify the true model among the simulated datasets (where the model was known), but different marker types typically provided similar answers. The most supported demographic model able to explain the real dataset implies that an endemic gene pool is still present in the Brisbane and North Pine Rivers and coexists with the gene pools derived from past documented translocation events. These results support the view that ABC modelling can be useful to reconstruct complex historical translocation events with contemporary implications, and will inform ongoing conservation efforts for the endangered and iconic Australian lungfish.

KEYWORDS
approximate Bayesian computation, Australian lungfish, conservation, demography, translocations

Roberto Biello and Silvia Ghirotto contributed equally to this work.

1 | INTRODUCTION

Approximately 120 years ago, a retired bank manager, Daniel O'Connor, addressed the members of the Royal Society of Queensland in Australia: 'I have the pleasure to inform you that the work of procuring *Ceratodus* and transferring live specimens to new habitats which you entrusted to me is finished' (O'Connor, 1897). The '*Ceratodus*' genus referred to by O'Connor was the Australian lungfish now known as *Neoceratodus fosteri*. It had only just been discovered in the Burnett River in 1870 and was described a few years later by the Brisbane Museum director as 'one of the most famous things in Australia' (Longman, 1925). At this point in time, Australian lungfish had not been recorded or documented in any rivers south of its established range in the Burnett and Mary River catchments. The translocations were thus commissioned in response to a perceived risk the species could face extinction on account of its restricted distribution, and the rarity of juveniles observed in the wild.

The Australian lungfish is still present in the Burnett and Mary catchments and in most of those rivers that received documented translocations, although only abundant in a few (Kind, 2011). The contributions of O'Connor in preventing the perceived extinction of the last of at least 11 lungfish species distributed in the Cretaceous all over Australia (Kemp, 1997) remains unclear, with ongoing conjecture as to the endemic distribution of the species prior to translocations (Kemp & Huynen, 2014). Nevertheless, this IUCN-endangered species is currently threatened by habitat modifications mainly due to water resource development and land use that reduce the quality and availability of important riverine habitats used for reproduction (Brooks et al., 2019; Kemp, 2018). Current conservation management of *N. forsteri* across its range is largely based on minimal intervention strategies of no harvesting, ecological research and permitted hatchery production to satisfy the aquarium trade. No other direct intervention strategies such as restocking or genetic management are currently active. Research has provided insights into improved environmental management actions that can aid *N. forsteri* conservation outcomes including water management principles (Espinoza et al., 2013; Marshall et al., 2015) and specific spawning habitat manipulations; however, these tools are constrained within a broader context of water resource planning at the catchment scale. Conservation efforts and water resource management must find a balance to sustain this threatened species.

The impact of O'Connor's translocations requires exploration to better prioritise future conservation efforts. Could these early conservation efforts have resulted in backup populations outside of its endemic distribution, or have these efforts created a new genetically mixed population containing a diverse genetic lineage of endemic and introduced gene pools? Lungfish are not only unique among Australian fauna, but also a crucial biological and evolutionary model for studying and understanding the evolution of tetrapods and adaptations to land (e.g., Meyer et al., 2021; Woltering et al., 2020). As such, understanding and preventing extinction is an important ethical obligation to the world (Antonelli & Perrigo, 2018).

The intention of the original translocation event was to establish new sustainable populations in suitable rivers where the lungfish was not formerly found. Despite many early explorers, naturalists and ichthyologists searching for *Ceratodus* and other unique river fauna within Queensland, there are no historical records describing the occurrence of lungfish in either the Brisbane or North Pine Rivers before the translocations. Diaries of explorers from the 1820s, who extensively explored the Brisbane and surrounding rivers, sourcing fish for consumption and trading fish with the indigenous, made no mention of a fish unlike those more readily recognisable species (Steele, 1972). Of course, absence of evidence is not evidence of absence, especially for a species with a cryptic habit, a restricted geographical range and not previously described by science. The hypothesis that the current Brisbane River and North Pine River populations descended entirely from the O'Connor translocations has been questioned before (Kemp, 1986). Previous conclusions were often drawn in the absence of biological traits about the species, such as fecundity, longevity, survival rates and key drivers for recruitment, all essential to make informed conclusions about potential population dynamics. The presence of lungfish inhabiting the Brisbane River prior to 1896 was postulated recently on the finding of three ancient bone fragments, believed to be analogous to *N. forsteri* bone structures, in a cave located adjacent to the Brisbane River. These bone samples were carbon-dated 3500 years before the present (Kemp & Huynen, 2014), but were unable to be confirmed with DNA as being *N. forsteri*.

The main question explored by this research is whether the translocations that occurred in 1896 were the inception of the present-day lungfish populations in the Brisbane and North Pine Rivers, or whether they already existed in those rivers and were genetically influenced by those translocations (see Figure 1). Fisheries management agencies in Australia consider the Brisbane River and North Pine River populations as introduced (Bishop et al., 2018). The historical translocations described 109 adults being sourced from the Mary River (see Figure 1). Of these, only five and eight individuals were introduced into the Brisbane River and the North Pine River basins, respectively (O'Connor, 1897). All other survivors were released into other rivers across the South East Queensland region and have largely been unsuccessful in establishing viable populations. Anecdotal evidence suggests translocations may have also occurred later, possibly with individuals sourced from the Burnett River, as mentioned in newspapers from the time. Furthermore, recent information suggests that additional illegal translocations of lungfish may still be occurring (Kind, 2011).

Early efforts to use genetic methods to answer this question of endemic vs translocated populations often yielded ambiguous results (Frentiu et al., 2001). Despite improvements in molecular markers, genetic data have provided variable levels of confidence to assess the status, endemic or translocated, of the Brisbane River and North Pine River populations. It is important to briefly summarise the results and the conclusions of these studies to highlight that, in this specific case, but also in general for genetic studies more broadly, different translocation hypotheses are challenging

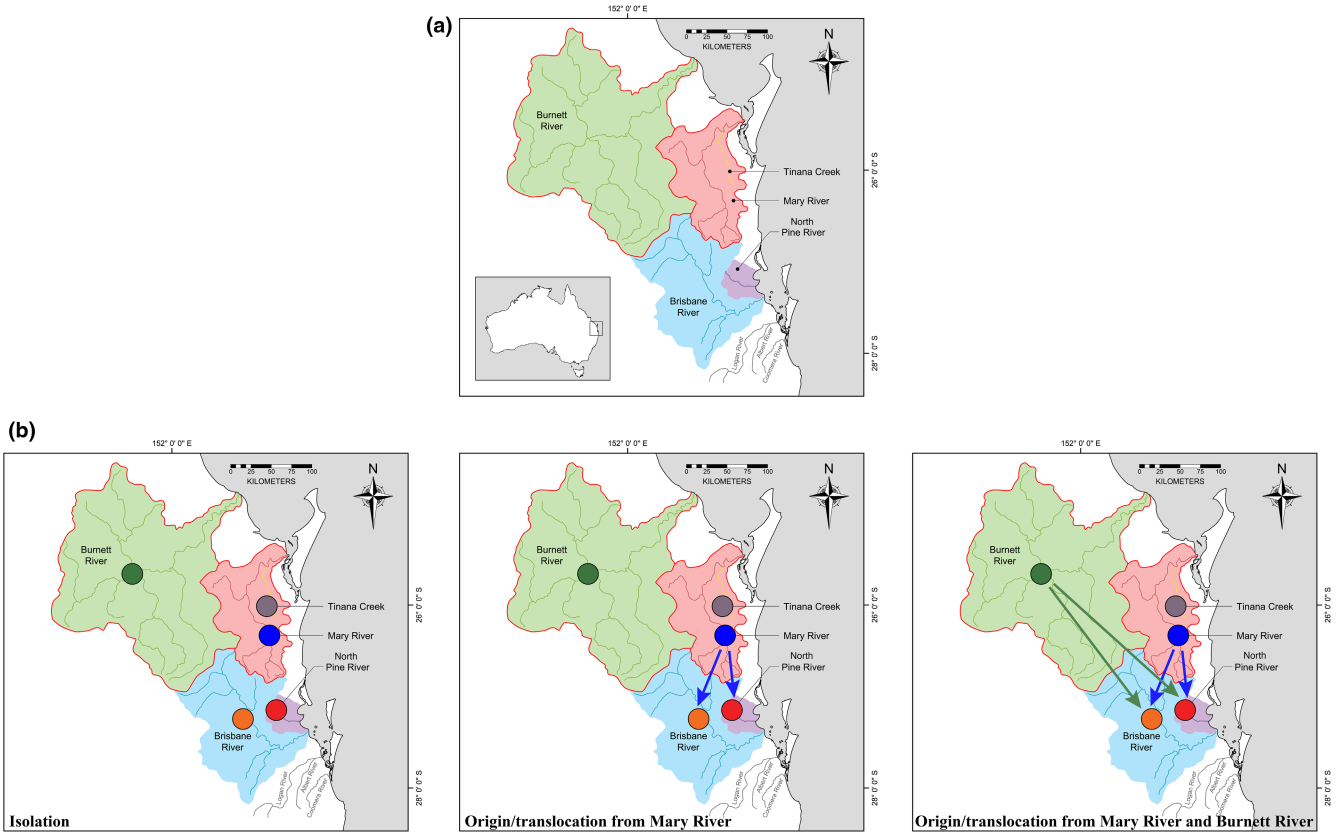


FIGURE 1 (a) Catchment areas of the individuals included in this study. (b) Main demographic models tested with ABC-RF; left panel: populations are evolving in isolation since divergence, corresponding to models 1a to 7a in Figure S1; central panel: Brisbane and North Pine River populations descend completely (models 1b to 7b in Figure S1) or partially (models 1d to 7d in Figure S1) from individuals translocated from Mary River; right panel: Brisbane River and North Pine River populations descend completely (models 1c to 7c in Figure S1) or partially (models 1e to 7e in Figure S1) from individual translocated from Mary River and Burnett River. [Colour figure can be viewed at wileyonlinelibrary.com]

to distinguish using descriptive methods alone, even when multiple genetic markers are available.

Frentiu et al. (2001) supported the translocation hypothesis based on one individual from the Brisbane River sharing a rare mtDNA haplotype, not found in the Burnett River, with two individuals sampled in the Mary River. But the most common of the eight haplotypes found in that study was shared among the three rivers. In addition, rare haplotypes are often missed in small-size samples, and only 13 or 14 individuals from each river were analysed by Frentiu et al. (2001). Lissone (2003) and Lissone et al. (2004) later found in a sample of 28 (Mary River and Burnett River) and 29 (Brisbane River) individuals that one polymorphic fingerprinting band was private in the Brisbane River, and even if the other eight bands were shared with other rivers, concluded that translocation could be excluded.

Most recently, genetic variation was analysed with modern markers, allowing a finer comparison among genetic pools across the *N. forsteri* range. Eleven microsatellites identified four major genetic pools corresponding to the Mary River and Brisbane River, Burnett River, North Pine River and Tinana Creek, the latter being a tributary of the Mary River (Hughes et al., 2015). The similarity between Mary River and Brisbane River populations was considered a signature supporting the translocation origin of the Brisbane River population, and the distinctiveness of the North Pine River population was attributed to a founder effect resulting from the small number of individuals reported to have been translocated into this system. However, genetic bottleneck signals expected from a recent origin of these two populations were not found, and very similar levels of diversity in terms of heterozygosity and allelic richness were found across all sites. Complete mitochondrial genomes from 71 lungfish (Bishop et al., 2018) found 173 segregating sites in 16,573 base pairs, and 38 out of 41 haplotypes were private in each of the five rivers analysed in the microsatellite study. The three shared haplotypes, however, were shared among the Mary River and North Pine River (two haplotypes) and the Mary River and Brisbane River (one haplotype). These results were interpreted as evidence of the translocation origin of the North Pine River and Brisbane River populations, with the many private haplotypes resulting from new mutations after the translocation and/or the small sample size, but the possibility of an endemic origin was not excluded (Bishop et al., 2018). Finally, a panel of 15,201 SNPs supported a clear genetic divergence between Mary River, Burnett River and Brisbane River populations, with the latter showing slightly higher affinity with the Mary River compared with the Burnett River and a slightly lower genetic variation compared with both (Schmidt et al., 2018). A

possible impact of the O'Connor Mary-to-Brisbane translocations in explaining these findings was speculated.

Here, we used three partially published datasets from mitochondrial genomes (Bishop et al., 2018), 11 microsatellites (Hughes et al., 2015) and 5196 SNPs (Schmidt et al., 2018; unpublished data) to model the possible origin scenarios for the current Brisbane River and North Pine River populations. We applied a model-based approach of random forest approximate Bayesian computation, both on single marker types and in a combined marker dataset, and we also analysed the different powers of these markers to reconstruct simulated translocation histories. Our results provide evidence that the Brisbane River and North Pine River lungfish populations have a genomic signature that indicates each of these rivers contains an endemic component, which should be considered in future conservation management. We also provide information on the power of the ABC approach and of different markers to address questions on past translocation events.

2 | MATERIALS AND METHODS

2.1 | The genetic dataset

The genetic data we analysed refer to five lungfish populations located in Queensland (Australia) (Figure 1a: Burnett River [BU], Tinana Creek [TI], Mary River [MA], Brisbane River [BR] and North Pine River [NP]). We analysed three different genetic markers, namely the mitochondrial genome (mtDNA; sample sizes: 71; Bishop et al., 2018), 11 microsatellite loci (STRs; sample sizes: 135; Hughes et al., 2015) and 5196 SNPs obtained from RAD sequencing (sample sizes: 100; Schmidt et al., 2018; unpublished data). Samples sizes for the SNPs dataset are as follows: 20 individuals each from BU, MA and BR (randomly sampled from those published in Schmidt et al., 2018) plus 20 individuals each from NP and TI (unpublished data). The new data have been generated following the same procedure reported in Schmidt et al. (2018). STACKS version 1.48 (Catchen et al., 2013) was used for de novo assembly and genotype calling of loci. Low-quality reads (phred score <10 within a 13 bp window) were removed with *process_radtags*. The assembly parameters, optimised using the Rochette and Catchen (2017) protocol, included a minimum stack depth of three reads (-m, 3), allowance of up to two mismatches between stacks (-M, 2), and a maximum of two mismatches between catalogue loci (-n, 2). This set of parameters was used to build an initial loci catalogue. Genotype corrections were performed with *rxstacks*, and a refined catalogue was generated with *cstacks* and *sstacks*. The *populations* program was applied for genotype call filtering. The criteria for filtering included: (i) all loci must be biallelic, (ii) each locus must be scored in ≥80% of individuals in the entire sample of 92 individuals, (iii) all loci are represented by a single SNP position (i.e., -write_single_snp flag) and (iv) any locus showing significant deviation from expected Hardy–Weinberg proportions in >1 river population sample was excluded from further analyses. The genetic structure inferred from the SNP dataset (which includes also original data not analysed before) was visualised using a discriminant analysis of principal components (DAPC; Jombart et al., 2010), implemented in the R package *adegenet* (Jombart, 2008). Neighbour-joining population trees based on F_{ST} distances at the STRs and the SNPs markers, reconstructed using the *ape* R package (Paradis & Schliep, 2019), are also used to visualise the relationships among groups. For the mtDNA phylogenetic tree, see Bishop et al. (2018).

2.2 | The models

Different hypotheses regarding the translocation histories of lungfish populations were tested using a model-based approach. Datasets of variation of the same size of those available were generated by simulation under different evolutionary and translocation scenarios, and the models were scored based on the resemblance between simulated and real data (see below for details). Different models assumed different topologies representing the relationships among populations and different plausible translocation histories. We analysed seven population topologies (labelled Models 1–7), each of them assuming five different translocation histories (labelled from *-a* to *-e*) (Figure S1).

2.3 | Population topologies

Simulations require the definition of the history of the sampled populations. Even if the identification of the best model to describe the relationships between the populations is not of interest in this study, the results on the translocation history would be biased by forcing the populations into a fixed and possibly wrong topological model. We therefore hypothesised seven different relationships (Figure S1) based on different logical criteria and let the data identify the best topology that can be used to study the translocation scenarios.

The topology in Models 1- should be considered as a sort of topological null hypothesis, with similar divergence between all pairs of endemic populations. Models 2- assume a tighter relationship between Tinana Creek and Mary River since they belong to the same catchment. Models 3- add to Models 2- a tighter relationship among rivers having closer estuaries, considering the possibility of gene flow within Northern and Southern catchments during glaciations. Models 4- and 5- are similar to Models 3-, with a longer (or much longer) isolation of Tinana Creek population, which was shown to be highly differentiated at the genetic level (Hughes et al., 2015). Models 6- are designed as Models 4-, but Mary River population is more closely related to the Southern group where Mary River individuals were introduced. Finally, Models 7- are similar to Models 6-, but with longer isolation of North Pine River population.

2.4 | Translocation scenarios

Models *-a* refer to the null hypothesis for the translocations, that is, these models exclude recent translocation events and

therefore assume that both Brisbane River and North Pine River populations have an ancient origin (as in Figure 1b, left panel; see also Figure S1). Models -*b* and -*c* assume that Brisbane River and North Pine River populations originated only in recent times, from a translocation event from Mary River (models -*b*; see Figure 1b, central panel, and Figure S1) or from both Mary and Burnett rivers (models -*c*; see Figure 1b, right panel, and Figure S1). Under models -*d* and -*e*, the Brisbane River and North Pine River populations are both the result of an admixture event between an endemic component and a more recent introgression due to translocations from Mary River (model -*d*; Figure 1b, central panel, and Figure S1) or from Mary River and Burnett River (models -*e*; Figure 1b, right panel, and Figure S1).

2.5 | Simulation details and summary statistics

All the parameters of the models are listed in Table S1 together with their associated prior distributions. Historical evidence was used to define the prior distributions of the parameters related to the translocation events. For the parameters related to topologies (divergence times and population sizes), large uniform (or loguniform) distributions were used. We assumed that endemic populations have constant population size to avoid over-parametrisation of the models. For each model, we drew 50,000 combinations of demographic parameters from prior distributions and generated a simulated dataset using the *fastsimcoal2* backward (coalescent) simulator (Excoffier et al., 2013). We separately generated mtDNA, STR and SNP data, considering the different effective population sizes at nuclear and mitochondrial markers. We used a log uniform prior for the mitochondrial mutation rate, between 0.011 and 0.125 mutations per nucleotide per million years, as in Burridge et al. (2008). When simulating STR loci, we considered a generalised stepwise mutation model (GSM; Estoup et al., 2002), with average mutation rate across loci having a normal prior distribution with mean equal to 0.0005 and variance equal to 0.0002. At each STR locus, the mutation rate had a gamma distribution with a shape parameter equal to 2, as in Marino et al. (2013) (Table S1). The simulated data have been summarised by different summary statistics. For the mtDNA, we calculated the number of haplotypes, the haplotype diversity, the total and private number of segregating sites, the average number of pairwise differences for each population and Tajima's *D*, the global and pairwise F_{ST} and the mean number of differences between pairs of populations. For the STRs, we calculated the mean and standard deviation over loci of the number of alleles, of the heterozygosity, of the modified Garza-Williamson index (average ratio between the number of alleles and the allelic range) for each population, and the global and pairwise F_{ST}. For the SNPs, we calculated the heterozygosity, F_{ST}, and three categories of segregating sites: per population, private and fixed for opposite sites. We calculated the summary statistics using *arlsumstat* (Excoffier & Lischer, 2010) and the three categories of segregating sites through a home-made R script.

2.6 | Strategy of model comparison through ABC-RF

The model comparison was performed through an ABC random forest (ABC-RF) approach (Pudlo et al., 2016). The ABC-RF, compared with the classical ABC algorithm (Beaumont et al., 2002) works well also with a reduced number of simulations per model and an increased number of summary statistics (Blum & François, 2010), thus allowing the study of several complex models. With ABC-RF, the classifier is constructed from simulations from the prior distribution for all the models that must be compared (called reference table) via a machine-learning RF algorithm. Once the classifier is constructed and applied to the observed data, the posterior probability of the resulting model can be approximated through another RF that regresses the selection error over the statistics used to summarise the data. Reliable estimates of the posterior probabilities can be achieved with just a few thousand simulations, and the informative statistics are systematically extracted from the pool used to summarise the data (Pudlo et al., 2016). The ABC-RF model selection estimates have been obtained using the function *abcrf* from the R package *abcrf* and employing a forest of 500 classification trees, a number suggested to provide the best trade-off between computational efficiency and statistical precision (Pudlo et al., 2016).

The demographic models detailed in the previous paragraph were compared exploiting a hierarchical approach. We initially identified, within each of the seven topologies tested, the translocation dynamic showing the better fit with the data, thus performing seven independent model selection experiments (e.g., comparing 1*a* to 1*e* and then comparing 2*a* to 2*e*). For Models 1 and 2 we performed a comparison among five models, while the specificity of the other models required a comparison among 8 (Model 5) or 11 (Models 3, 4, 6, 7) models. This difference is due to the different order of the branching events that require specific tests. Models were compared four times: separately considering the three types of genetic markers (mitochondrial genomes, STRs and SNPs), and combining them in the same reference table. After this first run of comparisons, we explicitly compared the best scenarios emerging from each topology for each genetic marker considered, and for the whole dataset comprising all the genetic markers together, to identify the model best representing the observed variation, both for the topology and the translocation dynamic.

The confusion matrices and the out-of-bag classification error (CE) were computed in each comparison. The confusion matrix is a square matrix reporting in different lines (one for each of the simulated scenarios) the fraction of simulated datasets supporting the scenarios indicated in the columns. The classification error is the sum of the cells for each line in the confusion matrix, when the cell corresponding to the simulated model is excluded. It measures the proportion of simulated data where the true demographic history is not correctly identified by the classification algorithm. CE is also informative about the respective inferential power of the different genetic markers used. To identify which marker or combination of markers may lead to the best classification algorithm, we also

evaluated the Prior Error Rate, which is the average value of the misclassification errors (CE) computed for each scenario (thus assuming an identical prior probability for each model).

The variation generated by the different demographic models, and the position of the observed data, were visualised through a linear discriminant analysis (LDA). LDA creates new uncorrelated variables, based on linear combinations of predictors (here the summary statistics), that maximise the differences between groups (here the models). The new variables (with decreasing discriminatory power) can be visualised in pairs.

2.7 | Parameter estimation

The parameter values of the most supported model were estimated after increasing the number of simulations to 500,000 and by maximising the fit between observed and simulated data. One per cent of the simulations closest to the observed dataset was retained for the estimation, after a *logtan* transformation of the parameters (Hamilton et al., 2005) and using the local linear regression approach (Beaumont et al., 2002). Model comparison under the random forest approach is efficient also with a few tens of thousands of simulations (Pudlo et al., 2016), but parameter estimation under the local linear regression approach requires more simulated datasets (Beaumont et al., 2002). We estimated parameters considering all the genetic markers together in the same analysis. As suggested by Wegmann et al. (2009), we linear-transformed the vector of summary statistics into partial least square (PLS) components through the *findPLS.r* script within the ABCtoolbox package (Wegmann et al., 2010). This analysis is able to extract from the set of statistics computed in the three panels of markers the components best explaining the parameter variation. Time in generations was converted in years assuming three to four lungfish generations in 100 years (Schmidt et al., 2018).

2.8 | Additional simulations to test for robustness of parameter estimates

We performed an additional set of simulations, using the *fastsimcoal2* simulator (Excoffier et al., 2013), to analyse in more detail the accuracy of our procedure in estimating low levels of translocation from different sources. To do this we generated 100 pseudo-observed datasets (pods) under known conditions: a fixed topology (model *1e*), the effective population sizes and the divergence time fixed to the mode values previously estimated for the selected model, and all the four translocation parameters fixed to a proportion of 0.01. We performed 100 estimations of the proportions of translocated individuals through the same procedure used to analyse real data and assessed the quality of the estimates by computing four indices: the bias, the root mean square error, the Factor 2 and the 95% coverage. The bias was calculated as $\frac{1}{n}\sum_{i=1}^{n}\frac{\theta_i-\theta}{\theta}$, where θ_i is the estimator of the parameter θ (true value), and n is the number of pods used (100 in our case). Because bias is relative, a value of 1 corresponds to a bias equal to 100% of the true value. The root mean square error (RMSE) is calculated as $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\theta_i-\theta)^2}$. Factor 2 corresponds to the proportion of the estimated median values within the interval defined by the 50% and the 200% of the true parameter value. Finally, the 95% coverage is the proportion of times that the known value lies within the 95% credible interval of the estimates.

3 | RESULTS

3.1 | Genetic variation in the real dataset

The DAPC analysis based on 100 individuals typed at the SNP panel yielded results consistent with previous studies based on other markers (Bishop et al., 2018; Hughes et al., 2015), providing further evidence of distinct genomic pools across the five basins (Figure 2). The population trees showed a partial consensus with the DAPC plot, with Mary River and Brisbane River more closely related than other populations, Tinana Creek and Burnett River genetically divergent (but clustering together, as suggested by the *x*-axis in the DAPC analysis), and the North Pine River more isolated than inferred by the DAPC plot (Figure S2). All trees consistently showed five distinct groups across the five basins (Figure S2; see Bishop et al., 2018). The summary statistics computed in the empirical data are reported in Tables S2–S4.

3.2 | Comparison of models

Tables S5–S11 report the detailed results for the first hierarchical step, that is, the comparisons between models that assume the same topology. For example, in Table S5, we compared all models with the same 'null' topology (equal differences between all pairs of endemic populations, Models 1-), but with different translocation scenarios (models 1a to 1e). The comparison within the same topological hypothesis required more than five alternatives (-*a* to -*e*) in some cases, since different branching orders required independent treatment in the simulations. For example, for model 3a, the split between Brisbane River and North Pine River could be simulated as the first, the intermediate, or the most recent in the topology (models 3a_1, 3a_2 and 3a_3 in Table S7, respectively).

In general, the models showed a good identifiability, with a classification error (CE, measuring the fraction of data simulated with a certain model not classified as originating from that model) in many cases below 10% and in some cases lower than 0.5% (Figure 3; Tables S5–S10; Figures S3 and S4). The highest CE values were observed in comparisons accounting for similar models (e.g., within Models 4, 6 or 7), whose specific topology definition made it essential to test 11 different models to explore the different demographic histories proposed. Some differences in CE levels also emerged among markers, with mtDNA and STRs showing higher classification errors than SNPs (Figure 3; Figures S3 and S4). Random subsamples of SNP loci showed that CEs remain stable moving from 4000 to 500 but increased

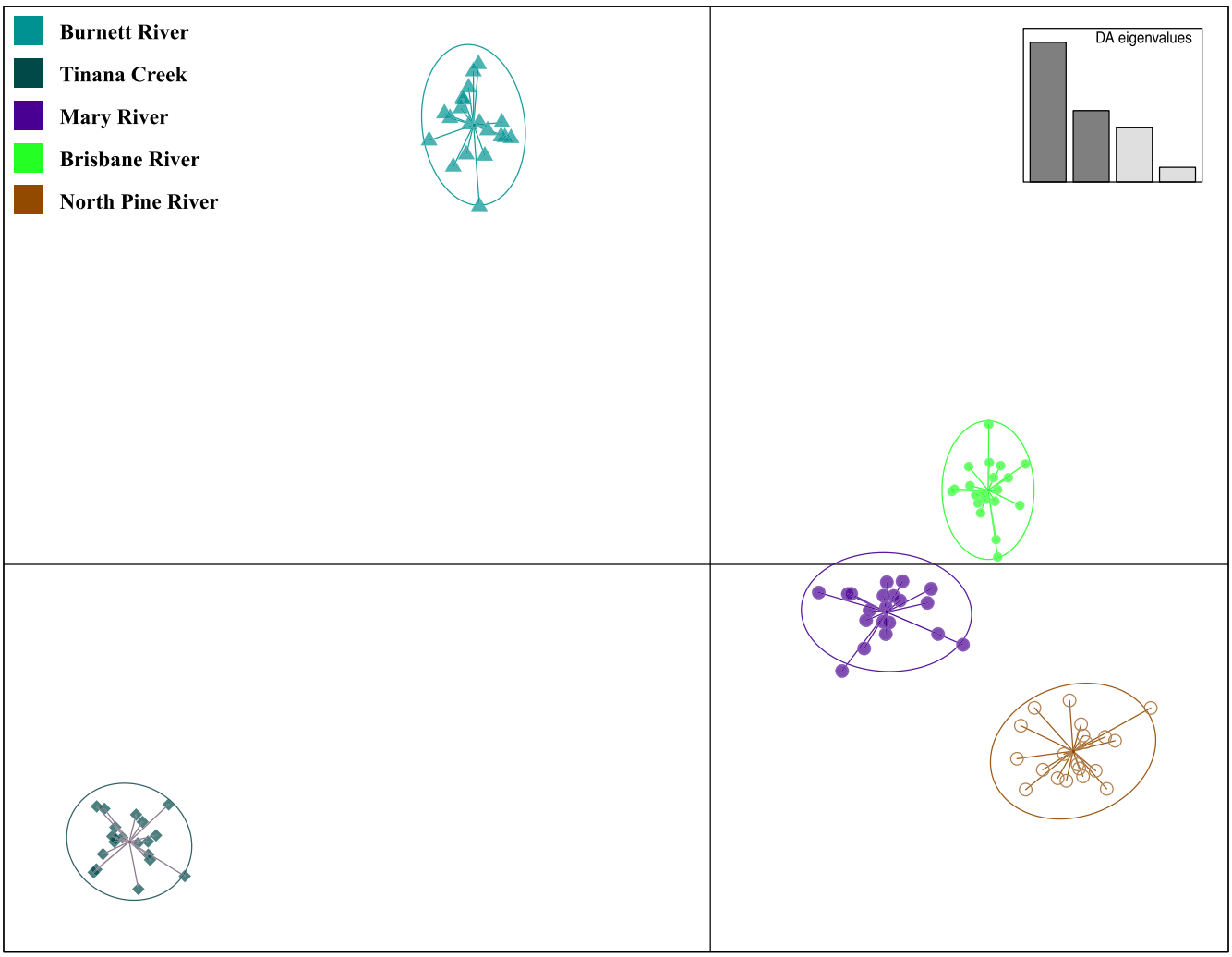


FIGURE 2 Discriminant analysis of principal components (DAPC) plot showing the relationship of 100 individual lungfish using 5196 SNPs. Colour-coded by sample site. [Colour figure can be viewed at wileyonlinelibrary.com]

substantially in the 100 SNPs dataset. With this latter panel, CEs were still much larger than was observed in the STRs dataset, but similar to what observed using mtDNA sequences (Figure S5).

The estimated prior error rates (Tables S5–S15) supported the view that the simulated SNPs dataset is more informative. STR statistics always produced the highest values of prior error rate, about 10 or 20 times higher than those resulting from SNP data in the same model comparison (Figure S4).

When applied to real data, the models without translocations (models *-a*) were never selected in the single marker datasets, the global dataset, or in the comparisons within the same topology (Figure 4; Tables S5–S15). Model *-b* (translocated origin of Brisbane and North Pine River populations from Mary River individuals) received the support only from the mtDNA dataset and only under topologies 6 and 7. The most informative datasets (SNPs panel or all markers jointly analysed) supported the translocation scenarios *-c*, *-d* or *-e* when different topologies were considered.

Once the model best accounting for the observed variation within each topology was identified, we performed a second step of selection, comparing the models supported in the previous step by the whole dataset (all markers). Models 6 and 7, in their scenario c which produced the best fit in the previous step (the scenario with Brisbane River and North Pine River that entirely descend from translocated individuals), are identical, and six (instead of seven) models were therefore compared (see Figure S1). In general, the ability of the classifier to detect the correct model in this comparison was strong, with CE between 0.5% and 13% for the SNPs or the complete datasets (Table S12). When relying on STRs only, the CE increased considerably, reaching values close to 40%.

The complete dataset favoured model 1e, that is, the simplest topology with all populations equally differentiated, and a mixed origin of Brisbane River and North Pine River from endemic ancestors and individuals introduced from both Mary and Burnett Rivers (Figure 4). The same topology was also supported by the mtDNA and SNP dataset when separately analysed, and the same translocation history was supported by the STR and SNP dataset, when separately analysed. The posterior probabilities associated with these models range between 62% and 76% depending on the genetic marker considered, four times higher than that expected a priori. We are therefore confident that model 1e is the best model

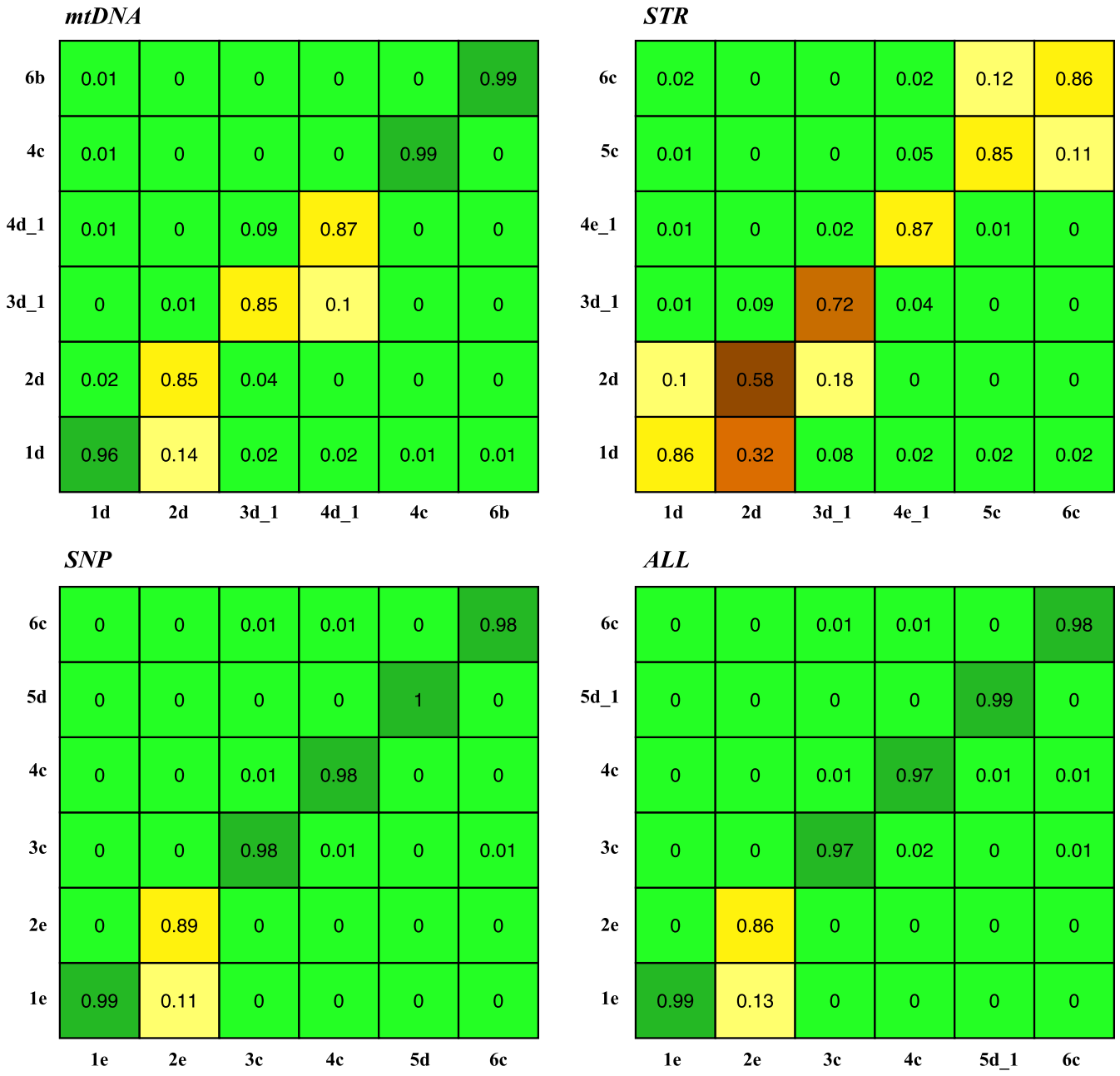


FIGURE 3 Confusion matrices (for each marker and overall) obtained when the most supported models for each topology are simulated. Each lines refer to a simulated model, and the figures reported in the different columns refer to the fraction of simulations supporting each model. High values along the top-right to bottom-left diagonal, and low values in all the other cells, support correct inferences, and colour codes are defined to visualise with the same colour similar inference quality. For the top-right to bottom-left diagonal, green to brown corresponds to decreasing fractions, and for the other cells green to brown corresponds to increasing fractions. [Colour figure can be viewed at wileyonlinelibrary.com]

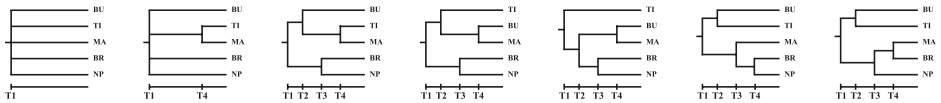
explaining the genetic variation of the lungfish populations we analysed.

We performed an LDA on simulations from the best models (identified for each topology) considering a set of statistics from all the genetic markers (Figures S6 and S7). The variation generated by these models overlapped eventually along the first three LDA axes, but the projection of the simulations generated under the model selected showed that the observed data are well captured by model 1e (Figures S6 and S7).

3.3 | Parameter estimation

The parameters of model 1e were estimated using the complete dataset and the 30 first PLS components to summarise the data (Table 1; Figure S8). All but one of the R^2 values (representing the fraction of the total variance explained by the PLS components used) were between 34% and 75%, indicating that parameter estimates can be considered reliable (Neuenschwander et al., 2008).

(a)



	M1	M2	M3	M4	M5	M6	M7	Best Models
mtDNA	*d*:0.684	*d*:0.633	*d*:0.535	*d*:0.517	*c*:0.568	*b*:0.650	*b*:0.682	M1*d*:0.622
STR	*d*:0.782	*d*:0.750	*d*:0.655	*e*:0.615	*c*:0.582	*c*:0.579	*c*:0.502	M4*e*:0.765
SNP	*e*:0.849	*e*:0.779	*c*:0.434	*c*:0.428	*d*:0.590	*c*:0.509	*c*:0.430	M1*e*:0.679
All	*e*:0.722	*e*:0.681	*c*:0.506	*c*:0.493	*d*:0.562	*c*:0.466	*c*:0.458	M1*e*:0.627

(b)

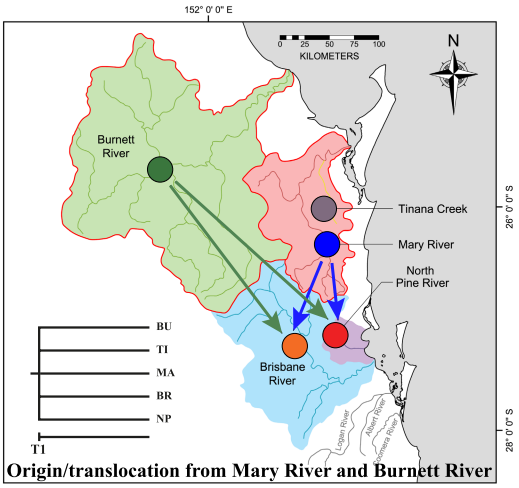


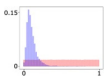
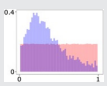
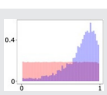
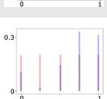
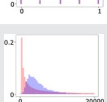
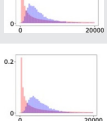
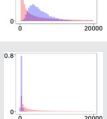
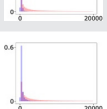
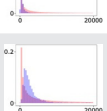
FIGURE 4 (a) Results of the hierarchical model selection procedure. Each column (from M1 to M7) shows the model best accounting for the observed variation within each of the seven topologies represented above the table. Within each cell is reported the most supported translocation scenario for that topology, with its posterior probability (*-a*: translocation origin is excluded; *-b* and *-c*: BR and NP populations have a translocation origin; *-d* and *-e*: mixed model; see Figure S1 for details). The last column shows the results of the comparison among the best models resulting from each topology for each genetic marker considered. The highlighted cell represents the model used to estimate parameters. (b) Schematic representation of the topology and demographic dynamics of the most supported model considering all the genetic markers together. BR, Brisbane River; BU, Burnett River; MA, Mary River; NP, North Pine River; TI, Tinana Creek. [Colour figure can be viewed at wileyonlinelibrary.com]

As expected, the translocation age, an almost fixed parameter with a priori distribution ranging from 1 to 5 generations, could not be estimated. Twenty-nine out of 30 PLS components calculated in the real dataset were included in the central 95% of the corresponding distributions estimated in the best 1% of the simulation under model 1e, suggesting that this model was able to generate datasets compatible with the real dataset (Figure S9).

We estimated three sets of parameters: the effective population sizes, the divergence time and the proportion of translocated individuals from Mary and Burnett Rivers to Brisbane and North Pine Rivers. The five populations showed different effective sizes (Table 1), with higher values for Mary River, Burnett River and Tinana Creek (some thousands of individuals) and lower values for North Pine and Brisbane Rivers (some hundreds of individuals). The simultaneous divergence between all five populations (assumed by the most supported topology) was estimated at 600–800 generations ago, that is, between 15,000 and 26,400 years ago. The estimated fraction of Brisbane individuals that were estimated to have had a Mary River or Burnett River origin at the translocation event was high (Table 1; Figure S8). The values with the highest probability for these two source rivers were approximately 90% and 27%, respectively, with median values equal to 79% and 33%, respectively. Assuming either the values with the highest probability or the median values, the total fraction of endemic components in the Brisbane River is estimated between 7% and 14%. The same fractions estimated for the North Pine River indicate a smaller impact of the translocations (Table 1; Figure S8), with a fraction of endemic component between 65% and 75%.

3.4 | Additional simulations to analyse the accuracy in estimating endemic and introduced ancestry

This set of simulations was motivated by the large fractions of introduced ancestry estimated, especially in the Brisbane River and considering the fact that historical records indicate that the number

	Median	Mode	95% HPD-LowB	95% HPD-UppB	R^2	
$P_{BU\text{-}NP}$	0.0865	0.0612	0	0.2498	.67	
$P_{MA\text{-}NP}$	0.3048	0.1998	0	0.8121	.4	
$P_{BU\text{-}BR}$	0.3273	0.2664	0.0193	0.7394	.64	
$P_{MA\text{-}BR}$	0.7949	0.8972	0.2989	1	.36	
Ttr	4.1845	4.5329	1	1.6877	.03	
N_{BU}	4076	2760	549	11,806	.76	
N_{MA}	4743	3368	824	11,998	.76	
N_{NP}	278	208	200	727	.38	
N_{BR}	356	200	200	2640	.34	
N_{TI}	2203	1483	210	8051	.75	
T1	820	630	300	1949	.54	

Note: $P_{x\text{-}y}$ represents the proportion of translocated individuals from *x* to *y*. Ttr is the translocation time, N_x the effective population size of the population *x*, T1 the simultaneous divergence time of the five populations as modelled in the selected topology. The last column reports the prior and the posterior distribution of the parameters.

TABLE 1 Point estimates (median and mode values), the 95% high posterior density (HPD-Lower and Upper bound) and the coefficient of determination (R^2) obtained for each parameter. [Colour table can be viewed at wileyonlinelibrary.com]

of introduced individuals was low. The rationale of the question addressed was to identify intrinsic factors of the methods we implemented that may have produced an overestimation of the introduced genomic components.

Quality indices were calculated to detect potential low proportions of translocation if present in our observed dataset (Table S16). The 95% coverage values indicated that, for the four parameters analysed, the true value always falls within the 95% confidence interval of the estimates. However, the other indices showed a worse situation, in particular the bias, always showing an overestimation (of about 10 times in case of the proportion of translocation from Burnett River to Brisbane River), and the Factor 2, that in all the four cases is zero (Pnb and Psb), or almost zero (Pnm and Psm). It is however worth noting that, being the simulated translocation proportion very small (0.01), any estimated proportion larger than 0.02 would not contribute to the Factor 2 statistic.

4 | DISCUSSION

Our reconstruction of the Australian lungfish translocation history suggests that 'One of the most famous things in Australia' (Longman, 1925) was originally distributed in all the main rivers where it is found today, including the Brisbane and the North Pine Rivers where the signature of the past translocation events is still recognisable. This scenario is now quantitively supported by a probabilistic analysis, and it can be regarded as a third intermediate hypothesis that excludes the previous hypotheses that individuals in the Brisbane and the North Pine Rivers are entirely descendent from either local populations (Kemp & Huynen, 2014) or reintroduced founders (Bishop et al., 2018).

The current Brisbane River genetic pool appears to be largely, though not entirely, descendent from a few individuals collected mainly in the Mary River and released by the bank manager Daniel

O'Connor more than a century ago, as a documented translocation event by the Royal Society of Queensland. The estimated endemic component comprises approximately 10% of the current gene pool, which is exceptionally low if a pre-existing endemic population existed, indicating the population is now dominated by the genes from these translocated individuals. For this situation to have developed, different non-mutually exclusive scenarios could be postulated: (1) the translocated lungfish have far outperformed any pre-existing endemic population, (2) additional undocumented translocation events could have taken place in the last century or (3) the original endemic Brisbane River population prior to the translocations was very small. The fraction of the endemic alleles is much larger, around 70%, in the North Pine River. The genetic data suggest a larger contribution (approximately three time larger) of the Mary River gene pool compared with the Burnett River, in both the Brisbane and North Pine Rivers.

These proportions, despite the large number of markers of different types, have large confidence intervals and should be interpreted cautiously. Another source of caution is the possible impact of natural selection on our estimates, which are based on the assumption that genetic variation at the analysed markers is neutral. For example, it has been shown (Ewing & Jensen, 2016; Johri et al., 2021) that neglecting background selection may produce biased inferences on the population size dynamic and the migration rate. We expect however that the impact of selection on our estimates was limited for two reasons. First, the likelihood that some of our few thousand SNPs sampled from a genome of approximately 40 giga-bases are under selection or linked to variants under selection is very limited. Second, a recent work highlights a significant relaxation of natural selection even in coding genes in lungfish species compared with other vertebrates (Fuselli et al., 2023).

The main unanswered question we addressed was about the presence or not in the Brisbane River and the North Pine River of an endemic genetic component, and we believe that our analysis supports the conclusion that there is such a component. The best-fitting model selected by our hierarchical process of model comparison excludes a complete non-endemic (translocation only) ancestry in these rivers, and this result is found not only when all markers are simultaneously analysed, but also when they are separately considered. Importantly, the simulated datasets suggest that if the real scenario would have excluded an endemic component (models b and c under different population topologies), our approach would be unlikely to have missed it. Similarly, the models without translocations can be confidently excluded.

Other estimated features and parameters of the best model support a similar genetic divergence (before the translocations) of the five different rivers, estimated at 15–26 thousand years ago, and two classes of effective population sizes, smaller (some hundreds) in North Pine River and Brisbane River and larger (some thousands) in Mary River, Burnett River and Tinana Creek. The first topological inference may reflect a similar divergence due to a real split event related to the last glaciations, but more likely is just telling us that these populations diverged for enough time (possibly also with local adaptation) to reduce the power to infer a branching scheme. In our study, we considered the population topology as a nuisance parameter, to be accounted for to improve the comparison between models assuming different translocation patterns, but of no real interest. The effective population sizes we estimated for the five populations are reasonable in both absolute and relative terms, and do not point to an immediate risk of extinction. The smaller effective population size estimated for the Brisbane River is an interesting outcome given that it contains a relatively large and comparable catchment size (7015 km^2) and environment compared with the Mary River (9595 km^2). This result can possibly be explained by a largely translocated origin of the Brisbane River population. On the contrary, considering that the North Pine River is considerably smaller than the others at only 348 km^2, the smaller effective population size we estimate is expected.

Several genetic markers and statistical methods have been used in recent years to identify historical translocation events. Here, we show that an additional modelling step can be used to improve these inferences further. In fact, explicit and detailed demographic models, and the estimation of their probability through simulation, outperform descriptive approaches that require interpretation of the patterns and integration of the results produced by different analyses and different markers. As a simple example of this difference, we note here that the descriptive analyses based on the SNP dataset (see DAPC in Figure 2), where Brisbane River and North Pine River individuals show some degree of resemblance with those sampled in Mary River, is visually compatible with a translocation event, but do not exclude models without translocation or without endemic contribution that were frequently suggested for this species in previous studies (e.g., Bishop et al., 2018; Kemp & Huynen, 2014). The model comparison approach we applied here combines different sources of genomic data and quantifies probabilities and confidence intervals of possible outcomes.

The main conclusion from this work is that there is a high probability that Brisbane River and North Pine River contain a mix of endemic and translocated ancestry. This conclusion has been reached using a combined dataset, but also when the analysis was performed using only a single non-recombining marker with less than 200 segregating sites (mtDNA), a panel of 11 independent and classical multiallelic nuclear markers (STRs), or >5000 biallelic random nuclear loci (SNPs) (Figure 4a). Nevertheless, the study of the simulated datasets clearly shows that the SNP markers are more informative. For example, considering different population topologies, the fraction of datasets simulated under model *-c* (only non-endemic origin of Brisbane River and North Pine River) erroneously assigned to model *-e* (mixed origin) is always smaller than 1% for the SNP panel, but much larger values are found for the other markers (up to 5% and 10% for mtDNA and STR markers, respectively). The larger error rate found in STRs compared with the mitogenomes is rather unexpected, considering that variable sites in the mtDNA are all linked. This could be an effect of the mutational model used for STRs which was not deeply investigated in this system, or to the low genetic variation observed in these markers (Hughes et al., 2015). We did not investigate the effects on the parameter estimation of using separately the three types of markers, but it is reasonable to

expect better estimates and smaller support intervals when nuclear SNPs are considered.

4.1 | Translocations and conservation

The aim of translocation plans in modern conservation biology is to create or maintain a viable and self-sustainable population or species, when the endemic population is extinct in the wild, or the extinction risk is estimated to be high or predicted to increase (Weeks et al., 2011). When the translocation implies the rescue of an endemic group (also called restocking), the expected genetic benefits are, in the short-term, a fitness increase due to reduced inbreeding depression and decreased frequency of fixed deleterious mutations and, in the long-term, an increase in genetic variation and thus the evolutionary potential (Hedrick & Garcia-Dorado, 2016; Ralls et al., 2020). On the contrary, the introduced genetic pool may have negative effects such as outbreeding depression (Rhymer & Simberloff, 1996) or if the endemic ancestry is reduced significantly up to a point where the rescue becomes a replacement, and the specific traits and genes of the endangered group disappear. Clearly, Daniel O'Connor and the members of the Royal Society of Queensland were only considering the possible benefits of increasing the distribution of this rare and 'important' fish, presumably on the assumption that the Australian lungfish were not present at the time in the Brisbane River and the North Pine River. No records of anyone checking this assumption exist.

After more than a century, the genomic data we analysed suggest that the lungfish were already distributed in the Brisbane River and the North Pine River, and therefore that O'Connor performed a restocking intervention. The introduced animals were very likely genetically divergent from any assumed endemic population, considering the level of genetic divergence found among the Mary River, the Burnett River and the Tinana Creek that have not experienced translocation events that we are aware of. The original endemic gene pool is largely lost in the Brisbane River, but could be argued to have been maintained in the North Pine River. If the endemic populations in these rivers would have survived without the restocking is difficult to guess, but a major outbreeding depression effect can be excluded considering the current population sizes (Hughes et al., 2015).

The implications of these new findings will inform future genetic management considerations for the species. Currently, the Australian lungfish is conserved under national and international conservation legislation. The primary Australian legislation, the Environmental Protection and Biodiversity Conservation Act (EPBC Act, 1999), considers all existing populations of a threatened species, irrespective of endemic or translocated origin, as a significant sub-population for the species as a whole. Under this legislation, therefore, the identification of a mixed origin, endemic and translocated, of the Brisbane River or North Pine River populations is inconsequential in that all sub-populations form an important management unit to sustain the species. However, considering that the local habitats of the five main lungfish populations are quite different, and genetic divergence is high, we expect some level of local adaptation to these different conditions supporting independent management. And if this is particularly true for the rivers where translocation can be excluded, we believe that also the likely admixed populations in the Brisbane and the North Pine Rivers should be followed with great attention analysing the genetic divergence and the introgression process at different genes. More practically, the value of the Brisbane River and North Pine River lungfish as insurance populations to the endemic populations of the Burnett River, Mary River and Tinana Creek, requires careful consideration from a genetic standpoint. These mixed gene pools certainly need to be managed as separate populations but should not to be mixed with the known endemic populations of the Burnett River, Mary River and Tinana Creek, without careful consideration of the genetic implications. Outbreeding depression and its classical mechanisms (such as loss of adaptation and genetic incompatibilities, also called extrinsic and intrinsic outbreeding depression) can produce fitness reduction, and it has been theoretically and empirically supported for a long time (e.g., Allendorf et al., 2022; Rhymer & Simberloff, 1996). In addition, the risk of introducing deleterious alleles when individuals from a large and genetically variable population are used to rescue a small one has been recently suggested and should be considered (Kyriazis et al., 2021). Restocking may boost adaptive genetic variability and provide some insurance against unpredictable future conditions, but preliminary analyses on local adaptation patterns can reduce the risk of outbreeding depressions, and genomic typing of introduced animals is a promising approach to exclude individuals with large inbreeding coefficients and large genetic load (Speak et al., 2023; van Oosterhout, 2020). In the case of *N. forsteri* these complex genetic considerations will not come easily, given their low genetic diversity and mixed genetic provenance of the Brisbane and North Pine River populations. This warrants caution when considering moving mixed and depauperate genotypes to or from the Mary and Burnett Rivers.

Finally, further insights may be gained by investigating the genetic diversity of those small remnant populations of the presumed translocated lungfish from 1896 still in existence in a number of other river catchments in the Brisbane area, including the Enoggera Creek, Coomera River and Logan River systems. Their genetic analysis could be of great value to better reconstruct the evolutionary and translocation history of this species, and to plan future interventions.

AUTHOR CONTRIBUTIONS

SG, JMH and GB designed the study. DJS produced the SNP data. RB and SG analysed the data. RB, SG and GB wrote the manuscript. All the authors discussed and interpreted the results and contributed to the writing.

ACKNOWLEDGEMENTS

GB, SF and RB are deeply grateful to Jane Hughes and Dan Schmidt for their hospitality at Griffith University, Queensland, and for their friendship. Part of this study was funded by the University of Ferrara. RB was supported by MUR grant 201794ZXTL. JMH and DJS were supported by Griffith University, Queensland.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

Individual genotype data from RADseq are available on Zenodo (https://doi.org/10.5281/zenodo.7646503). Scripts for running the simulations and the ABC analyses are available on Zenodo (https://doi.org/10.5281/zenodo.7646503).

ORCID

Roberto Biello https://orcid.org/0000-0002-5916-884X
Silvia Ghirotto https://orcid.org/0000-0003-2522-9277
Daniel J. Schmidt https://orcid.org/0000-0001-5638-497X
Silvia Fuselli https://orcid.org/0000-0002-1442-6171
Giorgio Bertorelle https://orcid.org/0000-0002-2498-2702

REFERENCES

Allendorf, F. W., Funk, W. C., Aitken, S. N., Byrne, M., & Luikart, G. (2022). *Conservation and the genomics of populations* (3rd ed.). Oxford University Press.

Antonelli, A., & Perrigo, A. (2018). The science and ethics of extinction. *Nature Ecology & Evolution*, *2*(4), 581. https://doi.org/10.1038/s41559-018-0500-z

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*(4), 2025–2035. https://doi.org/10.1093/genetics/162.4.2025

Bishop, C. R., Hughes, J. M., & Schmidt, D. J. (2018). Mitogenomic analysis of the Australian lungfish (*Neoceratodus forsteri*) reveals structuring of indigenous riverine populations and late Pleistocene movement between drainage basins. *Conservation Genetics*, *19*, 587–597. https://doi.org/10.1007/s10592-017-1034-7

Blum, M. G., & François, O. (2010). Non-linear regression models for approximate Bayesian computation. *Statistics and Computing*, *20*, 63–73. https://doi.org/10.1007/s11222-009-9116-0

Brooks, S., Espinoza, T., Kennard, M., Arthington, A., & Roberts, D. (2019). *Neoceratodus forsteri*. The IUCN Red List of Threatened Species.

Burridge, C. P., Craw, D., Fletcher, D., & Waters, J. M. (2008). Geological dates and molecular rates: Fish DNA sheds light on time dependency. *Molecular Biology and Evolution*, *25*(4), 624–633. https://doi.org/10.1093/molbev/msm271

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*, 3124–3140. https://doi.org/10.1111/mec.12354

Espinoza, T., Marshall, S. M., & McDougall, A. J. (2013). Spawning of the endangered Australian lungfish (*Neoceratodus forsteri*) in a heavily regulated river: A pulse for life. *River Research and Applications*, *29*(10), 1215–1225. https://doi.org/10.1002/rra.2607

Estoup, A., Jarne, P., & Cornuet, J. M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, *11*(9), 1591–1604. https://doi.org/10.1046/j.1365-294X.2002.01576.x

Ewing, G. B., & Jensen, J. D. (2016). The consequences of not accounting for background selection in demographic inference. *Molecular Ecology*, *25*, 135–141. https://doi.org/10.1111/mec.13390

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*(10), e1003905. https://doi.org/10.1371/journal.pgen.1003905

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Frentiu, F. D., Ovenden, J. R., & Street, R. (2001). Australian lungfish (*Neoceratodus forsteri*: Dipnoi) have low genetic variation at allozyme and mitochondrial DNA loci: A conservation alert? *Conservation Genetics*, *2*, 63–67. https://doi.org/10.1023/A:1011576116472

Fuselli, S., Greco, S., Biello, R., Palmitessa, S., Lago, M., Meneghetti, C., McDougall, C., Trucchi, E., Rota Stabelli, O., Biscotti, A. M., Schmidt, D. J., Roberts, D. T., Espinoza, T., Hughes, J. M., Ometto, L., Gerdol, M., & Bertorelle, G. (2023). Relaxation of natural selection in the evolution of the Giant lungfish genomes. *Molecular Biology and Evolution*, *40*(9), msad193. https://doi.org/10.1093/molbev/msad193

Hamilton, G., Currat, M., Ray, N., Heckel, G., Beaumont, M., & Excoffier, L. (2005). Bayesian estimation of recent migration rates after a spatial expansion. *Genetics*, *170*(1), 409–417. https://doi.org/10.1534/genetics.104.034199

Hedrick, P. W., & Garcia-Dorado, A. (2016). Understanding inbreeding depression, purging, and genetic rescue. *Trends in Ecology & Evolution*, *31*(12), 940–952. https://doi.org/10.1016/j.tree.2016.09.005

Hughes, J. M., Schmidt, D. J., Huey, J. A., Real, K. M., Espinoza, T., McDougall, A., Kind, P. K., Brooks, S., & Roberts, D. T. (2015). Extremely low microsatellite diversity but distinct population structure in a long-lived threatened species, the Australian lungfish *Neoceratodus forsteri* (Dipnoi). *PLoS One*, *10*(4), e0121858. https://doi.org/10.1371/journal.pone.0121858

Johri, P., Riall, K., Becher, H., Excoffier, L., Charlesworth, B., & Jensen, J. D. (2021). The impact of purifying and background selection on the inference of population history: Problems and prospects. *Molecular Biology and Evolution*, *38*(7), 2986–3003. https://doi.org/10.1093/molbev/msab050

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 1–15. https://doi.org/10.1186/1471-2156-11-94

Kemp, A. (1986). The biology of the Australian lungfish, *Neoceratodus forsteri* (Krefft 1870). *Journal of Morphology*, *190*(S1), 181–198.

Kemp, A. (1997). A revision of Australian Mesozoic and Cenozoic lungfish of the family Neoceratodontidae (Osteichthyes: Dipnoi), with a description of four new species. *Journal of Paleontology*, *71*(4), 713–733.

Kemp, A. (2018). Recruitment failure in the Australian lungfish, *Neoceratodus forsteri* (Osteichthyes: Dipnoi), in south-east Queensland. *Pacific Conservation Biology*, *25*(3), 283–298.

Kemp, A., & Huynen, L. (2014). Occurrence of lungfish in the Brisbane River, Queensland, Australia dates back to 3850 yr BP. *Journal of Archaeological Science*, *52*, 184–188.

Kind, P. K. (2011). The general natural history of the Australian lungfish *Neoceratodus forsteri* (Krefft 1870). In J. M. Jørgensen & J. Joss (Eds.), *Biology of lungfishes* (vol. 2010, pp. 61–95). CRC Press.

Kyriazis, C. C., Wayne, R. K., & Lohmueller, K. E. (2021). Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. *Evolution Letters*, *5*(1), 33–47. https://doi.org/10.1002/evl3.209

Lissone, I. (2003). *Conservation genetics and the Australian lungfish* Neoceratodus forsteri *(Krefft 1870); a spatio-temporal study of population structure*. MSc thesis, University of the Sunshine Coast, Sippy Downs, Queensland.

Lissone, I., Shapcott, A., & Ovenden, J. (2004). The use of RAFs enables determination of genetic structure within and among catchments in the Australian lungfish *Neoceratodus forsteri*. In *51st annual meeting*

of genetics Society of Australia Conference Abstracts (p. 32). Genetics Society of Australia.

Longman, A. H. (1925). *The Brisbane courier*, 26 November, page 6.

Marino, I. A., Benazzo, A., Agostini, C., Mezzavilla, M., Hoban, S. M., Patarnello, T., Zane, L., & Bertorelle, G. (2013). Evidence for past and present hybridization in three Antarctic icefish species provides new perspectives on an evolutionary radiation. *Molecular Ecology*, *22*(20), 5148–5161. https://doi.org/10.1111/mec.12458

Marshall, S. M., Espinoza, T., & Mcdougall, A. J. (2015). Effects of water level fluctuations on spawning habitat of an endangered species, the Australian lungfish (*Neoceratodus forsteri*). *River Research and Applications*, *31*(5), 552–562. https://doi.org/10.1002/rra.2754

Meyer, A., Schloissnig, S., Franchini, P., Du, K., Woltering, J. M., Irisarri, I., Wong, W. Y., Nowoshilow, S., Kneitz, S., Kawaguchi, A., Fabrizius, A., Xiong, P., Dechaud, C., Spaink, H. P., Volff, J. N., Simakov, O., Burmester, T., Tanaka, E. M., & Schartl, M. (2021). Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature*, *590*(7845), 284–289. https://doi.org/10.1038/s41586-021-03198-8

Neuenschwander, S., Largiadèr, C. R., Ray, N., Currat, M., Vonlanthen, P., & Excoffier, L. (2008). Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): Inference under a Bayesian spatially explicit framework. *Molecular Ecology*, *17*(3), 757–772. https://doi.org/10.1111/j.1365-294X.2007.03621.x

O'Connor, D. (1897). Report on preservation of *Ceratodus*. *Proceedings of the Royal Society of Queensland*, *12*, 101–102.

Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Pudlo, P., Marin, J. M., Estoup, A., Cornuet, J. M., Gautier, M., & Robert, C. P. (2016). Reliable ABC model choice via random forests. *Bioinformatics*, *32*(6), 859–866. https://doi.org/10.1093/bioinformatics/btv684

Ralls, K., Sunnucks, P., Lacy, R. C., & Frankham, R. (2020). Genetic rescue: A critique of the evidence supports maximizing genetic diversity rather than minimizing the introduction of putatively harmful genetic variation. *Biological Conservation*, *251*, 108784. https://doi.org/10.1016/j.biocon.2020.108784

Rhymer, J. M., & Simberloff, D. (1996). Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, *27*(1), 83–109.

Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RADseq short-read data using stacks. *Nature Protocols*, *12*, 2640–2659. https://doi.org/10.1038/nprot.2017.123

Schmidt, D. J., Fallon, S., Roberts, D. T., Espinoza, T., McDougall, A., Brooks, S. G., Kind, P. K., Bond, N. R., Kennard, M. J., & Hughes, J. M. (2018). Monitoring age-related trends in genomic diversity of Australian lungfish. *Molecular Ecology*, *27*(16), 3231–3241. https://doi.org/10.1111/mec.14791

Speak, S., Birley, T., Bortoluzzi, C., Clark, M., Percival-Alwyn, L., Morales, H., & van Oosterhout, C. (2023). Genomics-informed captive breeding can reduce inbreeding depression and the genetic load in zoo populations. *Authorea*. https://doi.org/10.22541/au.169566325.52568264/v1

Steele, J. G. (1972). *The explorers of the Moreton Bay district, 1770–1830*. University of Queensland Press.

van Oosterhout, C. (2020). Mutation load is the spectre of species conservation. *Nature Ecology & Evolution*, *4*(8), 1004–1006. https://doi.org/10.1038/s41559-020-1204-8

Weeks, A. R., Sgro, C. M., Young, A. G., Frankham, R., Mitchell, N. J., Miller, K. A., Byrne, M., Coates, D. J., Eldridge, M. D., Sunnucks, P., Breed, M. F., James, E. A., & Hoffmann, A. A. (2011). Assessing the benefits and risks of translocations in changing environments: A genetic perspective. *Evolutionary Applications*, *4*(6), 709–725. https://doi.org/10.1111/j.1752-4571.2011.00192.x

Wegmann, D., Leuenberger, C., & Excoffier, L. (2009). Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics*, *182*(4), 1207–1218. https://doi.org/10.1534/genetics.109.102509

Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, *11*(1), 1–7. https://doi.org/10.1186/1471-2105-11-116

Woltering, J. M., Irisarri, I., Ericsson, R., Joss, J. M., Sordino, P., & Meyer, A. (2020). Sarcopterygian fin ontogeny elucidates the origin of hands with digits. *Science Advances*, *6*(34), eabc3510. https://doi.org/10.1126/sciadv.abc3510

SUPPORTING INFORMATION

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

How to cite this article: Biello, R., Ghirotto, S., Schmidt, D. J., Fuselli, S., Roberts, D. T., Espinoza, T., Hughes, J. M., & Bertorelle, G. (2024). Unravelling the mystery of endemic versus translocated populations of the endangered Australian lungfish (*Neoceratodus forsteri*). *Molecular Ecology*, *33*, e17266. https://doi.org/10.1111/mec.17266