Whole genome duplication potentiates inter-specific hybridisation and niche shifts in Australian burrowing frogs *Neobatrachus*

Polina Yu. Novikova[1,2,†], Ian G. Brennan[3], William Booker[4], Michael Mahony[5], Paul Doughty[6], Alan R. Lemmon[7], Emily Moriarty Lemmon[4], Levi Yant[8], Yves Van de Peer[1,2,9,10,†], J. Scott Keogh[3,†], Stephen C. Donnellan[11,12,†]

1. VIB-UGent Center for Plant Systems Biology, Ghent, Belgium
2. Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium
3. Division of Ecology & Evolution, Research School of Biology, The Australian National University, Canberra, 2600, Australia
4. Department of Biological Science, Florida State University, 319 Stadium Drive, PO Box 3064295, Tallahassee, FL 32306-4295, USA.
5. School of Environmental and Life Sciences, University of Newcastle, Callaghan, 2308, Australia
6. Western Australian Museum, Locked Bag 49, Welshpool DC, 6986, Australia
7. Department of Scientific Computing, Florida State University, 400 Dirac Science Library, Tallahassee, FL 32306-4120, USA.
8. School of Life Sciences and Future Food Beacon, University of Nottingham, Nottingham, United Kingdom
9. Bioinformatics Institute Ghent, Ghent University, Ghent, Belgium
10. Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa
11. South Australian Museum, North Terrace, Adelaide, 5000, Australia
12. School of Biological Sciences, University of Adelaide, North Terrace, Adelaide, 5005, Australia

[†] Corresponding authors: polina.novikova@psb.vib-ugent.be, yves.vandepeer@psb.vib-ugent.be, scott.keogh@anu.edu.au, steve.donnellan@samuseum.sa.gov.au

Polyploidy plays an important role in evolution because it can lead to increased genetic complexity and speciation. It also provides an extra copy buffer and increases genetic novelty. While both common and well-studied in plants, polyploidy is rare in animals, and most polyploid animals reproduce asexually. Amphibians represent a dramatic vertebrate exception, with multiple independent sexually reproducing polyploid lineages, but very few cases have been studied in any detail. The Australian burrowing frog genus *Neobatrachus* is comprised of six diploid and three polyploid species and offers a powerful model animal polyploid system. We generated exome-capture sequence data from 87 individuals representing all nine species of *Neobatrachus* to investigate species-level relationships, the origin of polyploid species, and the population genomic effects of polyploidy on genus-wide demography. We resolve the phylogenetic relationships among *Neobatrachus* species and provide further support that the three polyploid species have independent origins. We document higher genetic diversity in tetraploids, resulting from widespread gene flow specifically between the tetraploids, asymmetric inter-ploidy gene flow directed from sympatric diploids to

tetraploids, and current isolation of diploid species from each other. We also constructed models of ecologically suitable areas for each species to investigate the impact of climate variation on frogs with differing ploidy levels. These models suggest substantial change in suitable areas compared to past climate, which in turn corresponds to population genomic estimates of demographic histories. We propose that *Neobatrachus* diploids may be suffering the early genomic impacts of climate-induced habitat loss, while tetraploids appear to be avoiding this fate, possibly due to widespread gene flow into tetraploid lineages specifically. Finally, we demonstrate that *Neobatrachus* is an attractive model to study the effects of ploidy on evolution of adaptation in animals.

Introduction

Polyploidy or whole genome duplications (WGDs) play important roles in ecology and evolution[1,2]. Although widespread in plants, polyploidy events also have been detected in animals. For instance, two ancient WGDs occurred early in the vertebrate lineage[3], while more recent WGDs occurred in several animal groups, including insects, molluscs, crustaceans, fishes, amphibians and reptiles[4,8,16]. Sex chromosome systems are more frequent in animals than in plants[5], and often require dosage compensation which can be disturbed by genome duplication, at least partially explaining why polyploidy is less frequent in animals[6]. The majority of polyploid animals switch to different atypical modes of bisexual reproduction, for example, gynogenesis, where a sperm does not fertilize the egg, but rather stimulates it to develop[4,7,8]. Amphibia are the only known exception with multiple independent occurrences of diploid and sexually reproducing polyploid sister species[9], possibly due to their homomorphic (undifferentiated) sex chromosomes[10] which do not require dosage compensation in homogametic (XX or ZZ) individuals[11].

Here, we focus on a group of widely distributed, endemic Australian burrowing frogs, *Neobatrachus*. This genus comprises six diploid (*N. albipes, N. fulvus*, *N. pelobatoides*, *N. pictus*, *N. sutor*, *N. wilsmorei*; 2n=24) and three tetraploid (*N. aquilonius*, *N. kunapalari*, *N. sudellae*; 4n=48) species[12,13], all characterised by bisexual reproduction[14]. The diploid species are well defined based on external morphology, male advertisement calls and divergence at allozyme loci[15-17]. Frog call structure differs among ploidies with higher ploidy species having lower pulse rates, a trait linked to cell size increase with increasing ploidy[18]. Indeed, all tetraploid *Neobatrachus* species have lower pulse number and rate in their advertisement calls, but each of the species still have distinct calls[19,20]. This differs from the more extensively studied gray treefrog, *Hyla versicolor*, where all tetraploids are of different origin, although merged into a single species through frequent interbreeding[21], have similar calls. Tetraploid species of *Neobatrachus* were suggested to have independent origins based on mitochondrial DNA (mtDNA)[22] and to have originated through autotetraploidy rather than allotetraploidy (hybridization), as they exhibit tetrasomic inheritance and show high rates of tetravalent formation during meiosis[14,23]. It has also been observed that, while the tetraploid species are distributed sympatrically with some of the diploid species, they are also able to occupy more arid areas across Australia[2,22]. Generally, polyploidy has been associated with tolerance to harsh conditions, but it is not clear whether whole-genome duplication broadly provides a fitness advantage or simply is a consequence of elevated rates of unreduced gamete formation[2,24-26], which might be more prone to occur in extreme environments. At the same time, changing

environments may contribute to amphibian extinction rates, which continue to increase and are driven by many interdependent factors such as habitat loss, emergence and spread of diseases, invasive species and pollution[27-30].

In the current study, we use an anchored hybrid enrichment approach (AHE)[31-34] to resolve the phylogenetic relationships among *Neobatrachus* species, and to assess fine-scale intra-specific genetic population structure. We also quantify the extent of hybridization between the nine *Neobatrachus* species with a particular focus on taxa with contrasting ploidies. Finally, we combine population dynamics assessments with changes in ecologically suitable areas for each species to describe population responses to climate changes.

Results

Phylogenetic relationships, genetic population structure and gene flow in *Neobatrachus*

We first generated sequence data and alignments for 439 targeted orthologous nuclear loci of 87 *Neobatrachus* individuals spanning the entire genus as well as nine *Heleioporus* individuals as outgroups (see Methods). We then built a species tree and gene trees from the sequenced loci with ASTRAL-II[35] using RAxML[36] (Fig. 1A). These traditional bifurcating species-tree methods reveal extensive conflict between genealogies, as well as between nuclear DNA (nDNA) and mtDNA topologies, but they do all demonstrate that the three tetraploid species do not form a monophyletic group (Fig.1A, Fig. S1, Fig. S2). Multidimensional scaling (MDS) of gene tree topologies suggested that nuclear loci constitute either two or four topology clusters (Fig. S3), indicating competing signal. In the absence of informative fossil material, we estimated the approximate evolutionary timescale of *Neobatrachus* frogs for mitochondrial (Fig. 1C) loci using secondary calibrations[37]. Interspecific divergence times provided support for an Eocene-Oligocene origin of *Neobatrachus* (Fig. 1C). Clustering patterns of gene tree topologies, together with relatively deep divergence times between the species, suggest recent or ongoing gene flow.

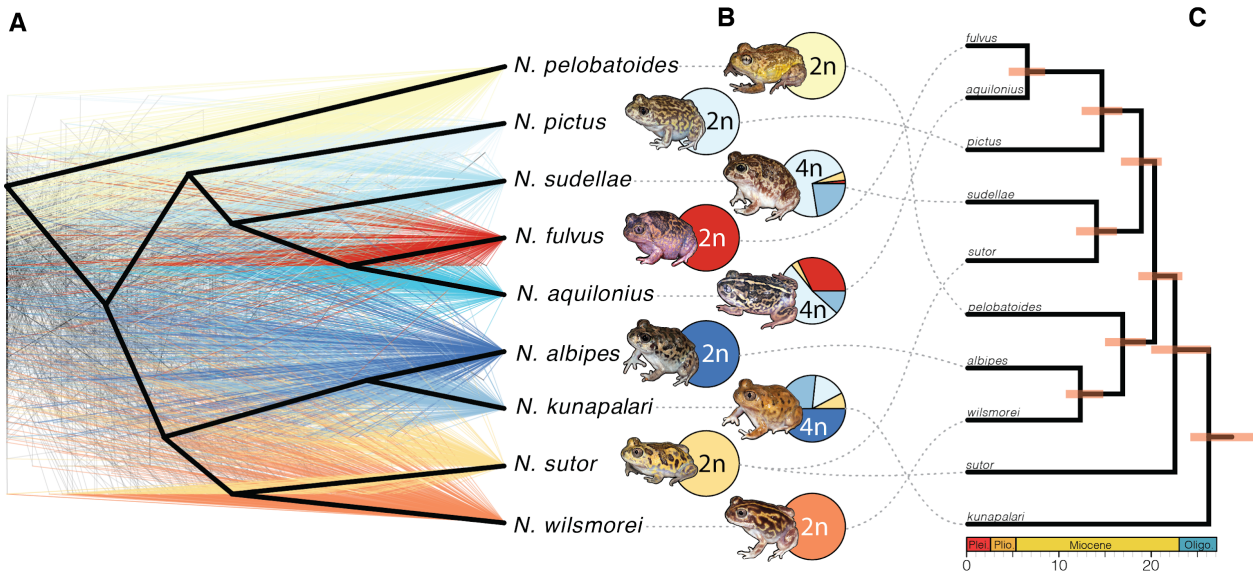

Fig. 1. Independent origins of *Neobatrachus* tetraploids and high levels of reticulation. A) Gene trees, colored by clade, for 361 nuclear loci based on 2 individuals per species show considerable incongruence and differ from the nuclear (bold black topology in A) and mitochondrial (C) species trees. Conflict between gene tree clusters (Fig. S3), the nuclear species tree, and the mitochondrial tree suggest non-bifurcating relationships between the species. B) Pie charts represent summarised admixture proportions for each species (summing assignments for each individual, Fig. S1, Fig. 2) at optimal clustering with K=7. Tetraploids (*N. sudellae*, *N. aquilonius* and *N. kunapalari*) show highly admixed ancestries. C) Dated mitochondrial tree built from coding sequences conflicts with the topology of the nuclear species tree in A. Red bars represent 95% confidence intervals on the ages of nodes, noted in millions of years before present.

To investigate this further we assessed the population structure of *Neobatrachus* by running ADMIXTURE analyses[38] on extracted polymorphism data (66,789 sites in total; see *Methods*; Fig. 1B, Fig.2 Fig. S1, Fig. S4). Overall, admixture clustering corresponded with the phylogenetic placement of the individuals on the species tree (Fig. 1, Fig. S1). Diploid species were clearly split at K=7 and did not show admixed individuals, whereas all three tetraploid species showed admixture with each other and with local diploid species (Fig. 1, Fig. 2, Fig. S1).

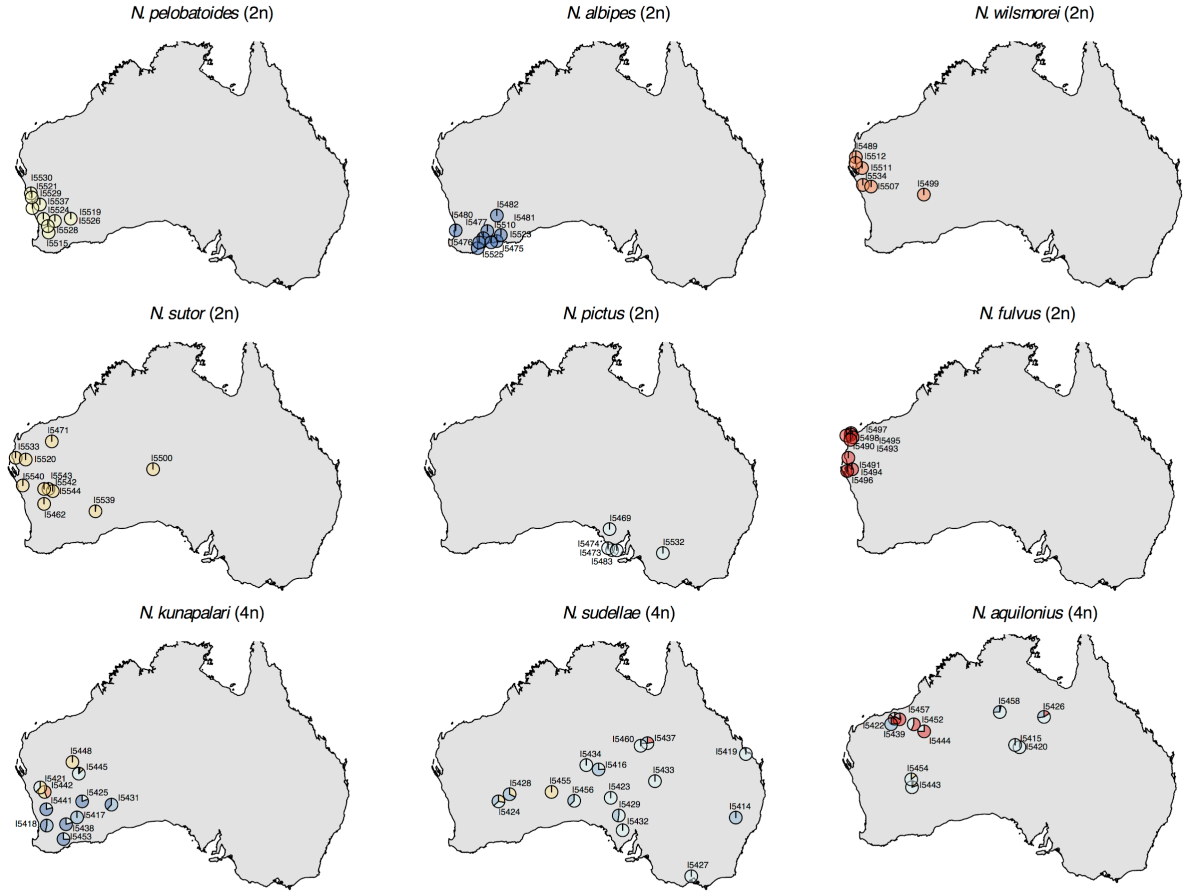

Fig.2. ADMIXTURE results (K=7) shown separately for each species. According to the sampling geographical locations of the individuals, pie charts show the probability of the assignment of the individual to one of the 7 individually colored clusters. Overlapped pie charts on the map have been moved just enough to appear separate. Diploid *Neobatrachus* species (top 6: *N. pelobatoides*, *N. albipes*,

N. wilsmorei, *N. sutor*, *N. pictus*, *N. fulvus*) are each assigned to separate clusters, while all three tetraploid species (bottom 3: *N. kunapalari*, *N. sudellae*, *N. aquilonius*) are admixed between each other and with local diploid species.

To further assess the complex demographic history of *Neobatrachus*, we performed TreeMix[39] modeling where species relationships are represented through a graph of ancestral populations (Fig. 3A). The structure of the graph is inferred from allele-frequency data and Gaussian approximation of genetic drift such that the branch lengths in the graph are proportional to the amount of drift since population split. We sequentially added up to 15 migration events, showing saturation of the model likelihood at five additional migration edges on average for 30 runs of Treemix, each with a different seed for random number generation (Fig. 3E). We show an example of the inferred introgression events and the bifurcating graph for the model with five migration events for the run that resulted in the highest maximum likelihood (Fig. 3A-D). Inferred migration events (Fig. 3B) indicate widespread directional introgression and interploidy gene flow between the polyploid species, however only two introgression events were significant (p value<0.05) in this run: from *N. sudellae* (4n) to *N. kunapalari* (4n) and from *N. sutor* (2n) to *N. kunapalari* (4n). There was some variability in inferred migration edges from run to run, so to estimate the most frequently inferred migration events, we summed the significant inferred migration edges among 30 TreeMix runs with five events allowed (Fig. 3F). Migration events were found most frequently from *N. sudellae* (4n) to *N. kunapalari* (4n) (19 of 30 runs) and from *N. sutor* (2n) to *N. kunapalari* (4n) (12 of 30 runs). Interploidy introgression events were mostly asymmetric and from diploids to tetraploids, which corresponds with our ADMIXTURE cluster assignment results (Fig. 1, Fig. 2). Inferred introgression events are broadly congruent with clusters of conflicting gene-tree topologies (Fig. S3). Each tetraploid *Neobatrachus* species (*N. aquilonius*, *N. kunapalari, N. sudellae*) is sister to a diploid species in the TreeMix graphs (Fig. 3A-B, tips highlighted in bold) as well as in the species trees (Fig. 1), which is consistent with previously suggested independent origins for the tetraploid species[22].

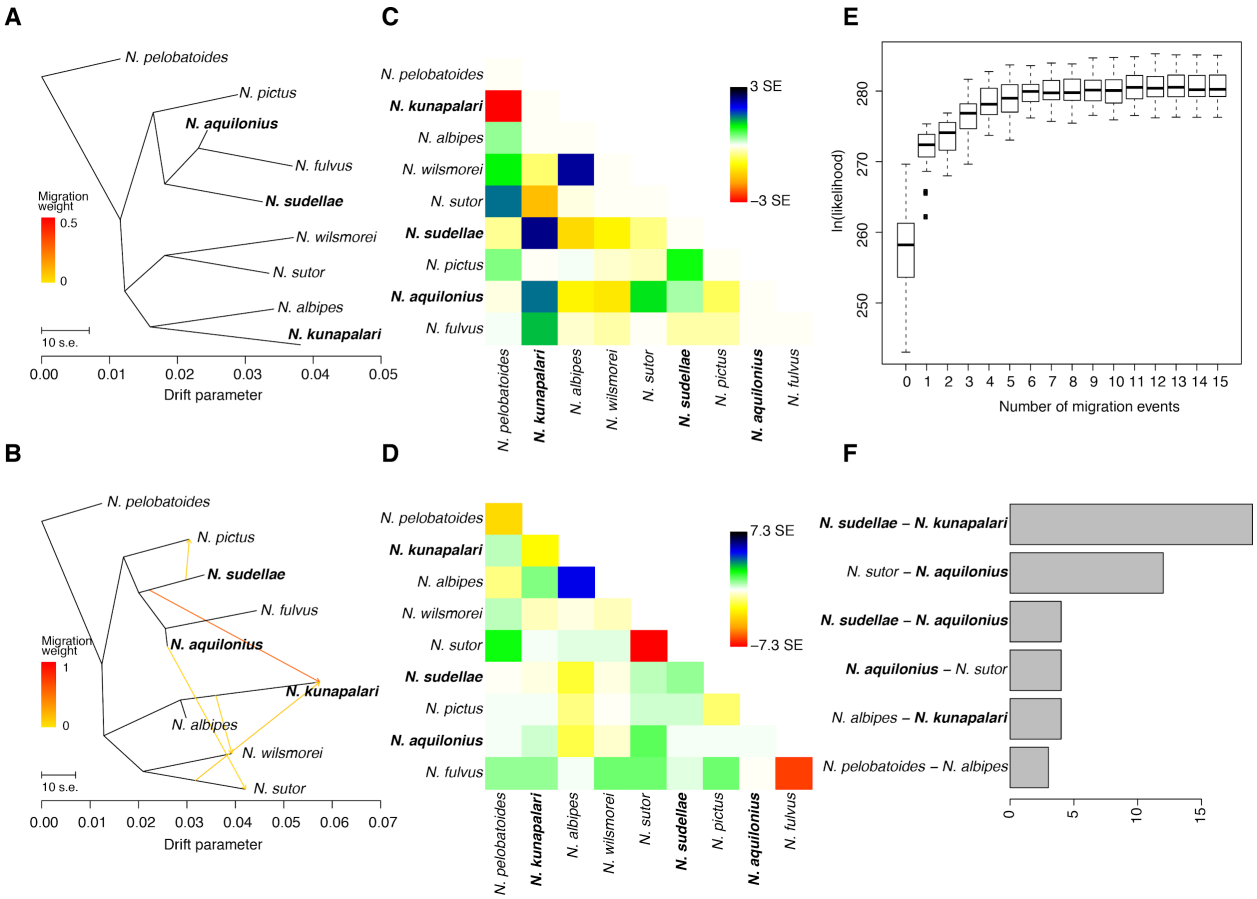

Fig. 3. Widespread introgression between *Neobatrachus* species. A) Bifurcating maximum likelihood tree produced by TreeMix. B) Example of a graph produced by TreeMix with 5 allowed migration events. C) Scaled residual fit between observed data and predicted model in A). Plot shows half of the residual covariance between each pair of populations divided by the average standard error across all pairs. Positive residuals represent populations where the model underestimates the observed covariance, meaning that populations are more closely related to each other in the data than in the modeled tree. Such population pairs are candidates for admixture events. Similarly, negative residuals indicate pairs of populations where the model overestimates the observed covariance. Overall, the residual plot of the model suggested that model fit could be improved by additional edges (migration events). D) Scaled residual fit between observed data and predicted model in B. Compared to Fig. 3C this suggests that, although the complexity of the species relatedness is not fully represented by the model, major gene flow events and their direction were probably captured. E) Boxplots of 30 runs of TreeMix (each started with a different seed for random number generation) likelihood at different numbers of allowed migration events; saturation starts after 3 additional migration edges. F) Barplot showing the number of times a particular directional migration event was inferred in 30 TreeMix runs with 5 migration events allowed. We show only the events which were inferred more than twice.

Estimation of suitable areas and demographic patterns

Tetraploid species have the highest nucleotide diversity among *Neobatrachus* species (Fig. 4A, Supplementary Table 1), which is due most likely to gene flow directed to tetraploid taxa and introgression between tetraploids of different origin. This is supported also by Fst distances (Fig 4C), where Fst distances between tetraploid species are the lowest values, while

Fst distances between tetraploids and diploid are larger, and Fst distances between the diploid lineages are highest, suggesting stronger isolation.

To estimate the dynamics in population abundance over recent times, we measured Tajima's *D,* a summary statistic that measures the lack or excess of rare alleles in a population compared to the neutral model. All of the *Neobatrachus* species have negative median values of Tajima's *D*, which suggests that none of the species are experiencing dramatic population diversity decline (Fig. 4B). We used the observed Tajima's *D* values as a proxy for each species' demographic patterns and compared them with estimated change in the suitable geographic area (Fig. 4F). In order to describe the ecological areas occupied by the different *Neobatrachus* species, as well as changes in those areas since the last glacial maximum (LGM) at around 20 Kya, we made use of occurrence data[40] (Fig. S5) and climate datasets[41]. We first performed a PCA of reduced bioclimatic variables concentrating on one of the highly correlated variables (r>0.85, Pearson correlation coefficient; *see Methods*) for individuals from the *Neobatrachus* occurrence data (Fig. S6). Using this substantially increased geographic sampling compared to our sequenced sample set, we could see moderate clustering of the individuals by species, which demonstrates that *Neobatrachus* species differ in their ecological (climatic) occupancies.

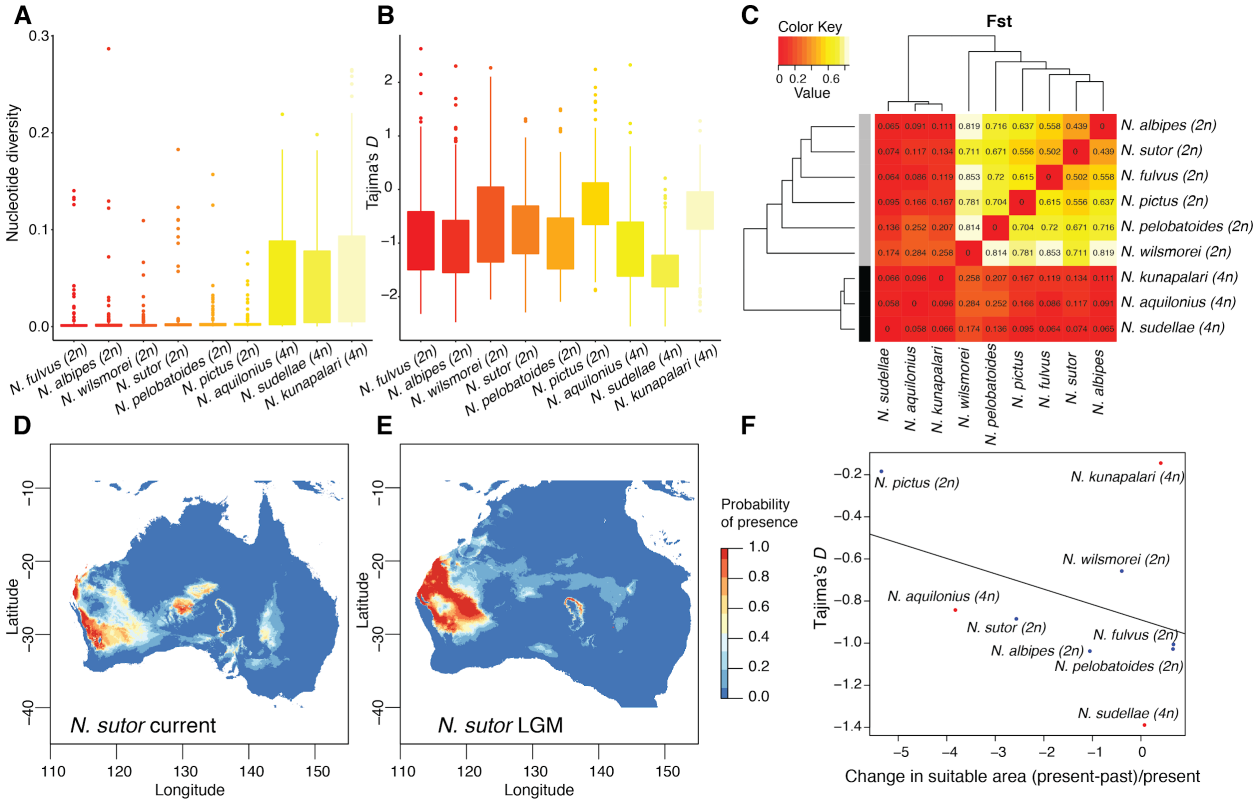

Fig. 4. Diversity and differentiation of *Neobatrachus* species and geographical suitability estimates. A) Distribution of nucleotide diversity (ordered and color scaled by the median nucleotide diversity increasing from red to yellow, left to right); B) distribution of Tajima's *D* estimator values (ordered and color scaled by the nucleotide diversity as in A). C) Heatmap and hierarchical clustering of the *Neobatrachus* lineages based on the distance matrix from pairwise median Fst values. Tetraploid species (*N. sudellae*, *N. aquilonius* and *N. kunapalari*; highlighted with black left bar) cluster together and are characterised by the lowest Fst values between each other. This, together with low Fst values between

tetraploid and diploid lineages, can probably be explained by the gene flow within the tetraploids and between the diploids and the tetraploids. Diploid lineages (highlighted with grey left bar) appear to be more isolated from each other compared to tetraploids, which is in agreement with ADMIXTURE assignment results and TreeMix estimations of possible migration events. D) Example of estimation of suitable area for *N. sutor* based on the occurrence data and current climate. E) Example of projection of suitable area for *N. sutor* based on the past climate at around 20Kya at LGM. F) Scatter plot showing relative change of the predicted suitable area at the LGM and current conditions for each species as a function of Tajima's D estimator. Pearson's correlation for all the species was -0.37 (R^2=0.0287, p-value=0.3); excluding the tetraploids -0.82 (R^2=0.596, p-value = 0.044).

We then modelled suitable areas for each species separately with MaxEnt[42], which applies machine learning maximum entropy modeling on the climate data at the geographical locations of the species occurrence data. Bioclimatic variables had different impacts on the model for each species (Fig. S7), however, appeared to be more similar for sympatric species (for example, *N. sutor* and *N. wilsmorei*) than for allopatric species (for example, *N. pictus* compared to any other diploid species). By projecting the models built on the current climate data on the past climate data (at the last glacial maximum, LGM) we could estimate the changes in the suitable habitat area since the LGM relative to the current suitable area for different species (Fig. 4D-E, Fig. S8-9). We observed a correlation between the change in the suitable habitat area and median Tajima's *D* for the diploid *Neobatrachus* species (Fig. 4F). As shrinkage in suitable habitat areas increases for a given diploid species, Tajima's *D* values also increase. This suggests that climate change may have already had a negative effect on diploid *Neobatrachus* genetic diversity via loss of suitable habitat even if populations are not obviously already contracting. Interestingly, tetraploid species appear to be the outliers to this trend, which we suggest may be due to their highly admixed genetic structure.

Discussion

We have investigated the population genomic consequences of genome duplication in a vertebrate model by generating and analysing nucleotide sequence data for 439 loci in 87 *Neobatrachus* individuals, covering the entire genus, including the three currently recognized tetraploids. The observation of non-bifurcating relationships between closely related species (Fig. 1) is now common[43-53], and is well understood to be caused by either incomplete lineage sorting or gene flow between the lineages or both. Population structure analysis by the ADMIXTURE assignment revealed that each of the diploid species form discrete clusters, consistent with their status as phylogenetically distant species. The tetraploids, however, were assigned to a mixed set of clusters, suggesting gene flow between each other and with the diploid species from overlapping geographical areas (Fig. 1, 2; Fig. S1). We extended our analysis of the population structure and potential gene flow by running TreeMix models (Fig. 3). In most replicate analyses, we inferred migration (gene flow) events from *N. sudellae* (4n) to *N. kunapalari* (4n) and from *N. sutor* (2n) to *N. aquilonius* (4n). In several analyses, we inferred migration from *N. sudellae* (4n) to *N. aquilonius* (4n), reversed migration between *N. sutor* and *N. aquilonius*, migration from *N. albipes* (2n) to *N. kunapalari* (4n) and even *N. pelobatoides* (2n) to *N. albipes* (2n). The latter, if true, may be attributed to an ancient migration event, since we do not have support from ADMIXTURE analysis for any recent mixing between the diploids.

Unidirectional gene flow from diploid into tetraploids could be achieved through a couple of scenarios. Firstly, a diploid individual could produce an unreduced gamete in a cross with a tetraploid forming a tetraploid that could backcross with the sympatric tetraploid species. While cytological evidence of the production of unreduced gametes in diploids is available in the form of triploid hybrids between diploid *Neobatrachus* species, we presently do not have direct evidence of this first scenario for inter-ploidy gene flow in *Neobatrachus* (Appendix 1, Supplementary Table 3, Fig. S10)[16,20,54-56]. Secondly, a diploid individual could cross with a tetraploid and form a triploid. A triploid producing an unreduced gamete (3n) in a cross with a diploid (normal 1n gamete) would produce a tetraploid which could backcross to a tetraploid individual producing introgression from the diploid into the tetraploid gene pool, i.e. the so-called "triploid bridge"[57]. We have cytological and molecular genetic evidence of naturally occurring triploid *Neobatrachus* and direct cytological evidence of triploid *Neobatrachus* producing balanced 3n gametes (Appendix 1). In another polyploid frog complex, *Odontophrynus*, triploids produced in laboratory crosses formed balanced haploid, diploid and triploid gametes[54], which opens the possibility of triploids crossing with diploids or tetraploids to produce tetraploids that could backcross into the sympatric tetraploid gene pool. Selective mating scenarios based on call characteristics could increase the likelihood of the backcross matings with tetraploids[30].

A mechanism for unreduced gamete formation, i.e. temperature stress, has been identified experimentally[25,26]. Similar brief stress conditions during zygote development before the first cell division could block the first division and form a tetraploid that could cross into a sympatric tetraploid gene pool[58]. Cold or heat shocks are possible for *Neobatrachus* eggs clutches at numerous locations across southern Australia that experience temperature extremes of the order of −5 °C to 48 °C[30]. Breeding by more southerly *Neobatrachus* tends to follow the onset of autumn and winter rains, typical of the Mediterranean climate in this region. Exposure to freezing or near freezing conditions in shallow egg deposition sites following cold fronts associated with rain fall events is frequently possible. *Neobatrachus* in the arid zone breed in association with heavy summer rains which may be associated with heat wave conditions.

Extensive gene flow between *Neobatrachus* species, especially between the tetraploid species, makes it difficult to estimate the true ancestral diploid population(s) for the tetraploids. Previously, it has been suggested that tetraploid *Neobatrachus* species might have independent origins[22]. Our results place tetraploid species in a polyphyletic arrangement on the species tree and on the TreeMix graphs, and suggest at least two independent origins: genetically closest diploid lineages to *N. aquilonius* and *N. sudellae* are *N. fulvus* and *N. pictus*, while the closest diploid lineage to *N. kunapalari* is *N. albipes*. An important question remains: what allows admixture between *Neobatrachus* tetraploids of potentially different origin and admixture of tetraploids with the local diploids, while the diploids seem to be currently isolated from each other? It may be that gene flow for close diploid relatives into tetraploids occurs frequently enough subsequent to the formation of the tetraploids to maintain enough phenotypic similarities for continued mating opportunities. Similarly, the tetraploid tree frogs *Hyla versicolor* of multiple origins *s*how high levels of interbreeding, however the levels of divergence between the ancestral diploid species in this case might be shallower[21]. Another example of a similar pattern was shown in plants, where polyploidy is more frequent: diploid *Arabidopsis lyrata* and *A. arenosa* could not hybridize, while tetraploidy seems to overcome the endosperm-based hybridization barrier enabling gene flow between the two species[59,60].

Neobatrachus species are widely distributed in Australia with tetraploid species occurring more in the central (drier) area compared to diploids, which is reflected on the principal component analysis of the climatic data for species occurrences (Fig. S6). Areas occupied by different *Neobatrachus* species differ only slightly in their environmental characteristics (Fig. S7). Worth mentioning that climatic variables do not entirely describe ecological niches, which could differ in other characteristics such as timing of breeding and foraging, food source preference, etc. Nevertheless, ecological niche modelling based on climate data may provide additional insights into population dynamic trends. Here, we applied the MaxEnt modelling approach to the publicly available climate and occurrence data for all nine *Neobatrachus* species, comparing the present and past suitable geographical areas. Most of the *Neobatrachus* species showed substantial changes of the suitable areas comparing current and past presence probabilities (Fig. S8-9). Interestingly, the estimated change in suitable habitat areas and population genetics estimator of demographic trends (Tajima's *D*), obtained from independent datasets, were correlated (Fig. 4F). Tetraploid species appear to be outliers from the general trend, probably due to their mixed population structure: in this case, emergence of rare alleles in the population due to migration events will affect Tajima's *D* estimator. Overall, it appears that the species with greater shrinkage of suitable area since the LGM had less negative median Tajima's *D* values, which suggests an ongoing shift from population expansion to population contraction.

Conclusion and Outlook

Neobatrachus frogs represent a group of diploid and tetraploid species with a complex ancestry. Our results, revealing gene flow between tetraploids and asymmetric inter-ploidy gene flow, open a set of important questions concerning the evolution of sexual polyploid animals. Whole-genome sequencing data for *Neobatrachus* species would not only help to refine the population structure and introgressive mixing in the genus, but also provide information on potential adaptive effect of the introgressed regions. One could hypothesize that a wide and potentially rapid spread of the tetraploids into new territories was facilitated by introgression from the locally adapted diploids[61] and that more detailed sampling of the tetraploids from parts of their ranges remote from sympatry with diploids may reveal evidence of early versus ongoing gene flow. Moreover, population-level genomic resequencing of multiple diploid and tetraploid sister species could provide insight into the unique biology of tetraploid sexual animals and effects of the tetraploidization on their evolution. The results on the changing suitable areas for the *Neobatrachus* species highlight the importance of continuous observation of their population dynamics. Monitoring the current status of biodiversity through collection of species occurrence data and population genetic data allows the prediction of population dynamics and hopefully timely response in conservation efforts in the face of rapidly changing environments[62,63]. Emerging methods of public engagement to collect occurrence and other data (video and audio; www.frogid.net.au) have potential to provide essential information on the state of frog species.

Methods

Anchored Hybrid Enrichment (AHE) phylogenomics

All the samples examined came from the Australian Biological Tissue Collection at the South Australian Museum. Details of all samples examined are presented in the Supplementary Data. We collected AHE data at Florida State University's Center for Anchored Phylogenomics (www.anchoredphylogeny.com), following the methods described in Lemmon et al.[31] and Prum et al.[64]. Briefly, after quantifying the extracted DNA using Qubit, we sonicated the DNA to a size range of 150-500bp using a Covaris Ultrasonicator. We then prepared indexed libraries using a Beckman Coulter FXp liquid-handling robot. After library QC using Qubit, we pooled the libraries in groups of 16 and enriched the library pools using an hybrid enrichment kit developed for use in Anurans[32],[33]. Finally, we sequenced the enriched library pools on two lanes of an Illumina 2500 sequencer with a PE150 protocol at the Translational Laboratory at Florida State University.

Following sequencing, we quality filtered the reads using the Casava high-chastity filter, then demultiplexed the reads using the 8bp indexes with no mismatches tolerated. In order to increase read length and correct for sequencing errors, we merged read pairs that overlapped by at least 18bp using the method of Rokyta et al.[65]. This process also removed sequencing adapters. We then performed a quasi-de novo assembly of the reads following Hamilton et al.[66], with *Pseudacris nigrita*, and *Gastrophryne carolinensis* as references. In order to reduce the potential effects of low level sample contamination, we retained only the assembly clusters containing more than 61 reads. In order to produce phased haplotypes from the assembly clusters, we applied the Bayesian approach developed by Pyron et al.[67], in which reads overlapping polymorphic sites are used to identify the likely phase of allelic variants within each locus. Because this approach was developed to accommodate any ploidy level, we were able to isolate two or four haplotypes for diploid and tetraploid individuals, respectively. We determined orthology for each locus using a neighbor-joining approach based on pairwise sequence distances, as described in Hamilton et al.[66]. We aligned homologous haplotypes using MAFFT v7.023b[68], then auto-trimmed/masked the alignments following the approach of Hamilton et al. (2017), but with *MINGOODSITES*=12, *MINPROPSAME*=0.3, and *MISSINGALLOWED*=48. Final alignments were visually inspected in Geneious R9 (Biomatters Ltd.,[69]) to ensure that gappy regions were removed and misaligned sequences were masked.

Whole mitochondrial genomes

Mitochondrial bycatch from exome capture methods can be an appreciable source of mitogenomic data[70]. To take advantage of these data, we assembled mitochondrial genomes from raw reads using custom scripts built around existing assembly and alignment software. Raw reads for each sample were assembled against the mitochondrial genome of *Lechriodus melanopyga*[71], using a baiting and mapping approach implemented in MITObim[72]. Mapped reads from each individual genomes were then aligned using MUSCLE[73], and adjusted manually. Entire concatenated mitogenomes were used to investigate sample identities, and to build an initial phylogenetic tree. To estimate divergence dates between species using

mitochondrial data, we extracted 13 protein coding regions (CDS), and partitioned them separately. We then trimmed the concatenated mitogenome phylogeny and CDS alignments down to two representatives of each taxon except for *Neobatrachus sutor*, which was not monophyletic (and therefore was trimmed down to four samples), and included two species of *Heleioporus* and *Lechriodus* as outgroups. The resulting tree and alignments were used as inputs for baseml and ultimately mcmctree, which we ran until we reached 20,000 post-burnin samples.

Phylogenetic analysis

To generate a molecular species tree, we started by reconstructing individual genealogies for each of the 439 recovered loci. We analyzed two datasets, the first included all samples (except triploids discussed in the Results section), and the second was trimmed down to just two individuals per species. We present results from the full sampling in the *Supplementary Material* (Fig. S2), and the finer sampling in the main text (Fig. 1). We used RAxML[36] to simultaneously search for the best tree and apply 100 rapid bootstraps, implementing the GTRGAMMA model of nucleotide evolution for each locus. In generating species trees, coalescent methods have been shown more accurate than concatenation, and so we used the shortcut coalescent method ASTRAL III. Shortcut coalescent methods like ASTRAL take individual gene trees as input, and are much more computationally efficient than full coalescent analyses. We used our RAxML-generated gene trees as input for ASTRAL, allowing us to make use of all our molecular data.

To address gene-tree incongruence and investigate possible conflicting signals in our data, we used multidimensional scaling (MDS) to approximate the relative distances between gene tree topologies[74]. To prepare the data, we trimmed down gene trees to one sample per species of *Neobatrachus,* and discarded loci missing any taxa, leaving us with 361 loci. We started by simply visualizing gene-tree incongruence overlaying the topologies of all 361 loci in DensiTree (Fig.1). We then calculated the pairwise distances between all gene trees using the Robinson-Foulds metric, in the R package APE[75]. We projected the tree distances into two and three dimensions (representing tree topology space) using MDS, as visualizing and interpreting any more dimensions becomes difficult. To test if gene trees are uniformly distributed throughout tree space, or clustered, we used the partitioning around the medoids algorithm as implemented in the R package CLUSTER[76]. We chose the optimum number of clusters (k), using the gap statistic, calculated for each k = 1–10. Clusters of gene trees represent similar topologies, and so we then summarized each cluster using ASTRAL, to identify consistent differences in topology.

We estimated the evolutionary timescale of *Neobatrachus* frogs for the mitochondrial DNAm, using 13 protein coding sequences, and a concatenated RAxML phylogeny of those loci as input for mcmctree. Because no valuable fossil information for *Neobatrachus* is available, we used four secondary calibrations from the most extensive anuran time-tree to date[37]. We applied these secondary calibrations as truncated Cauchy distributions with 2.5% above and below the designated bounds, to the splits between (i) Myobatrachidae and Microhylidae (lower=120, upper=140), *Kalophrynus* and remaining microhylids (lower=55, upper=66), *Cophixalus* and *Liophryne* (lower=17, upper=24), and *Microhyla* and *Kaloula* (lower=40, upper=55). This analysis produced an estimate of the divergence between *Neobatrachus* and *Heleioporus* with a

95% CI of 25-45 Mya, which was used to inform the mcmctree analysis of mtDNA coding sequences. We ran analyses until we reached 20,000 post-burnin samples.

Population structure

Maximum likelihoods of individual ancestries were estimated with ADMIXTURE[38] for 66,789 biallelic sites combining all 439 loci and allowing for maximum 20 out of 87 *Neobatrachus* individuals (excluding the outgroup species) to have missing data at each site. In order to include tetraploid samples in ancestry assignment we randomly chose two alleles for each site. We also applied minor allele frequency threshold of two percent. Ancestral population assignment showed three local minima of cross-validation errors at K equals 3, 7 and 9 (Fig. S4), with K=7 being the lowest, which we chose for the subsequent analysis as the optimal solution.

Misidentifications in the dataset

Both phylogenetic and admixture assignments suggested that several individuals had been misidentified in the field, which is expected for morphologically similar species and in particular for some of the diploid tetraploid species pairs, e.g. *fulvus* and *N. aquilonius*[15]. Field sampling can be accompanied by a certain level of honest mistakes in species identification, especially for sympatric species. However, a high level of incompletely sorted polymorphisms in recently split lineages or recent hybridization events could also result in uncertain positioning of an individual. We carefully curated the dataset and made a decision to rename some of the misidentified samples or completely exclude them from the analyses based on the amount of the missing data in the assembly, ploidy estimations from the sequencing data, identification from mitochondrial sequences and on the clear placement in a different clade. Below we describe our workflow for manual curation of the dataset to exclude or rename uncertain individuals without compromising too much on the potentially real shared variation.

The multiple sequence alignment resulting from the AHE workflow contains different amounts of informative sequence and gaps for each individual. First, we calculated the informative sequence fraction (no gaps) for each individual compared to the multiple sequence alignment length and applied a threshold of at least 0.2 of informative fraction for each individual to qualify for the subsequent analysis. Based on these criteria we excluded 6 samples (Supplementary Data).

Second, we estimated the ploidy of each sample using the nQuire software[77] on the next generation sequencing data mapped to one of the outgroup species *Heleioporus australiacus* (I5549) AHE assembly as a reference. As a preparation step for nQuire, we mapped reads to the reference using the BWA-MEM algorithm from BWA[78] (version 0.7.17), used Samtools[79] (version 1.6) to sort and index the mapping and removed potential duplicates from the PCR amplification step of library preparation with picard-tools (http://broadinstitute.github.io/picard/). We used the denoised input of base frequencies generated with default parameters for the Gaussian Mixture Model utilized in nQuire to estimate ploidy levels on the basis of frequency distributions at biallelic sites. The resulted estimations can be found in Supplementary Data. We excluded 5 samples placed in a different clade compared to the rest of the samples in the corresponding lineage, where ploidy estimation confirmed their misidentification. Finally, we

renamed 2 samples to a different species name with which it clustered, in cases when initial ploidy and estimated ploidy corresponded to each other (Supplementary Data).

We have also excluded from further analysis the sample I5442 initially identified as *N. kunapalari*, which was estimated to be a triploid and showed high levels of admixture between a diploid *N. wilsmorei* and potentially *N. kunapalari* or *N. sudellae*. In fact, triploid individuals are known in natural populations of *Neobatrachus*[17,19,20,80,81] (Appendix 1), and could provide an explanation for the gene flow between species of different ploidy through a "triploid bridge", when a 3n individual formed in a cross between 2n and 4n individuals and could produce a balanced haploid, diploid or triploid gametes and in the two later cases cross to a diploid and produce 4n individuals that can backcross into the sympatric 4n population[57,82,83] .

Introgression inference

The graphs representing ancestral bifurcations and migration events were produced using Treemix V.1.12[39]. Input data contained 5092 biallelic sites called at 439 loci among 9 *Neobatrachus* species with at least 20% of the data to be present at each species at each site. Position of the root was set to *N. pelobatoides* as the nuclear species tree suggested (Fig. 1A, black). To account for linkage disequilibrium we grouped SNPs in windows of size 10 using -k flag. We also generated bootstrap replicates using -bootstrap flag and subsequently allowed up to 15 migration events with flag -m. We ran TreeMix software with 30 different random number generated seeds. For graph and residuals visualisation we used R script plotting_funcs.R from the Treemix package.

Summary statistics and demographic tendencies

We calculated summary statistics (Fig. 4A-C) with the R package "PopGenome"[84] for the loci with more than one segregating biallelic site separately for each species for within-species statistics (nucleotide diversity, Tajima's *D*, Supplementary Table 1) and in a pairwise mode for between-species statistics (Fst).

Species distribution modelling

Bioclimatic variables were obtained from worldclim project[41] with 2.5 minutes resolution for reconstructed climate data at Last Glacial Maximum around 20Kya, averaged conditions across 1960-1990 and the most recent available conditions averaged across 1970-2000. Software DIVA-GIS 7.5 was used to trim the data to area longitude from 110 to 155 and latitude from -40 to -9. Bioclimatic variables were excluded if they were highly correlated (r>0.85, Pearson correlation coefficient) in all 3 climatic data sets, leaving for further analysis 6 bioclimatic variables in total: BIO9 = Mean Temperature of Driest Quarter, BIO10 = Mean Temperature of Warmest Quarter, BIO12 = Annual Precipitation, BIO17 = Precipitation of Driest Quarter, BIO18 = Precipitation of Warmest Quarter, BIO19 = Precipitation of Coldest Quarter.

To model the species suitable area we used software MaxEnt (v. 3.4.1.), which predicts species distribution from climate data using the species occurrences employing machine learning technique called maximum entropy modeling[42]. Here we used *Neobatrachus* species

occurrence data from amphibiaweb.org[40], which includes 189 entries for *N. albipes*, 282 for *N. aquilonius*, 87 for *N. fulvus*, 588 for *N. kunapalari*, 802 for *N. pelobatoides*, 699 for *N. pictus*, 707 for *N. sudellae*, 639 for *N. sutor* and 282 for *N. wilsmorei*. We used 75% of occurrence points for each species for model training and 25% for model testing with 1,000,000 background points and 10 replicates. We have trained the model on bioclimatic variables (reduced to 6 in total, described earlier) averaged across conditions 1960-1990; and then projected that model to the same set of environmental variables from the Last Glacial Maximum. The average test AUC (area under the Receiving Operator Curve) for the replicate runs for all the species was more than 0.9 (Supplementary Table 2), indicating a high performance of the models. In order to estimate which bioclimatic variable is the most important in the models we performed a jackknife test, where model performance was estimated without a particular variable and only with this particular variable in turn (Fig. S7).

We used cloglog output format of MaxEnt, which gives an estimate between 0 and 1 of probability of presence of the species in the area. In order to determine the relative change of the suitable area we used the point-wise mean values from the 10 replicates for model predictions on current and past climate (Fig. S8,9). We extracted the suitable area for both climate conditions with R library 'raster'[85] with 0.8 presence probability threshold and estimated change in the suitable area relative to the current suitable area as (current-past)/current.

Data availability

The raw genomic reads generated in this study were deposited in the NCBI SRA under the BioProject PRJNA507583. BioSample SRA number for each of the individual can be found in the Supplementary Data.

Acknowledgments

The Australian Research Council Discovery grant DP120104146 awarded to JSK and SCD supported the sequencing work. P.Y.N acknowledges postdoctoral fellowship from The Research Foundation – Flanders (FWO), 12S9618N. This work was also supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme [grant number ERC-StG 679056 HOTSPOT], via a grant to L.Y. We thank Sarah Catalano and Steven Myers for DNA extraction and shipping samples, Mitzy Pepper for laboratory assistance, the Western Australian Museum and Dale Roberts for some of the tissue samples and Stephen Mahony for some of the frog images. We thank Sean Holland and Michelle Kortyna at Florida State University's Center for Anchored Phylogenomics for assistance with data collection and analysis.

Competing financial interests

The authors declare no competing financial interests.

References

1 Soltis, D. E., Visger, C. J. & Soltis, P. S. The polyploidy revolution then...and now: Stebbins revisited. *Am J Bot* **101**, 1057-1078, doi:10.3732/ajb.1400178 (2014).

2 Van de Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat Rev Genet* **18**, 411-424, doi:10.1038/nrg.2017.26 (2017).

3 Dehal, P. & Boore, J. L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* **3**, e314, doi:10.1371/journal.pbio.0030314 (2005).

4 Otto, S. P. & Whitton, J. Polyploid incidence and evolution. *Annu Rev Genet* **34**, 401-437, doi:10.1146/annurev.genet.34.1.401 (2000).

5 Ming, R., Bendahmane, A. & Renner, S. S. Sex chromosomes in land plants. *Annu Rev Plant Biol* **62**, 485-514, doi:10.1146/annurev-arplant-042110-103914 (2011).

6 Wertheim, B., Beukeboom, L. W. & van de Zande, L. Polyploidy in animals: effects of gene expression on sex determination, evolution and ecology. *Cytogenet Genome Res* **140**, 256-269, doi:10.1159/000351998 (2013).

7 Stenberg, P. & Saura, A. Meiosis and its deviations in polyploid animals. *Cytogenet Genome Res* **140**, 185-203, doi:10.1159/000351731 (2013).

8 Neiman, M., Sharbel, T. F. & Schwander, T. Genetic causes of transitions from sexual reproduction to asexuality in plants and animals. *J Evol Biol* **27**, 1346-1359, doi:10.1111/jeb.12357 (2014).

9 Mable, B. K., Alexandrou, M. A. & Taylor, M. I. Genome duplication in amphibians and fish: an extended synthesis: Polyploidy in amphibians and fish. *Journal of Zoology* **284**, 151-182, doi:10.1111/j.1469-7998.2011.00829.x (2011).

10 Miura, I. Sex Determination and Sex Chromosomes in *Amphibia*. *Sex Dev* **11**, 298-306, doi:10.1159/000485270 (2017).

11 Schmid, M. & Steinlein, C. Sex chromosomes, sex-linked genes, and sex determination in the vertebrate class amphibia. *EXS*, 143-176 (2001).

12 Roberts, J. D. Taxonomic status of the Australian burrowing frogs *Neobatrachus sudelli*, *N. centralis* and *Neoruinosus* and clarification of the type specimen of *N. albipes*. *Records of the Western Australian Museum* **25**, 455–458 (2010).

13 Frost, D. R. Amphibian Species of the World: an Online Reference. Version 6.0 (Date of access). Electronic Database accessible at http://research.amnh.org/herpetology/amphibia/index.html. *American Museum of Natural History, New York, USA.* (2016).

14 Mahony, M. J. & Robinson, E. S. Polyploidy in the australian leptodactylid frog genus *Neobatrachus*. *Chromosoma* **81**, 199-212 (1980).

15 Mahony, M. & Roberts, J. D. Two new species of desert burrowing frogs of the genus *Neobatrachus (Anura:Myobatrachidae)* from Western Australia. *Records of the Western Australian Museum* **13**, 155-170 (1986).

16 Roberts, J. D., Mahony, M., Kendrick, P. & Majors, C. M. A new species of burrowing frog, *Neobatrachus (Anura:Myobatrachidae)*, from the eastern wheatbelt of Western Australia. *Records of the Western Australian Museum* **15**, 23-32 (1991).

17 Mahony, M., Donnellan, S. C. & Roberts, J. D. An Electrophoretic Investigation of Relationships of Diploid and Tetraploid Species of Australian Desert Frogs *Neobatrachus (Anura : Myobatrachidae)*. *Australian Journal of Zoology* **44**, 639-650 (1996).

18 Keller, M. J. & Gerhardt, H. C. Polyploidy alters advertisement call structure in gray treefrogs. *Proc Biol Sci* **268**, 341-345, doi:10.1098/rspb.2000.1391 (2001).

19 Roberts, J. D. Call Evolution in Neobatrachus (Anura: Myobatrachidae): Speculations on Tetraploid Origins. *Copeia* **1997**, 791-801, doi:10.2307/1447296 (1997).
20 Roberts, J. D. & Edwards, D. in *On the Ecology of Australia's Arid Zone* 149-180 (2018).
21 Holloway, A. K., Cannatella, D. C., Gerhardt, H. C. & Hillis, D. M. Polyploids with different origins and ancestors form a single sexual polyploid species. *Am Nat* **167**, E88-101, doi:10.1086/501079 (2006).
22 Mable, B. K. R., J.D. Mitochondrial DNA evolution of tetraploids in the genus *Neobatrachus (Anura: Myobatrachidae)*. *Copeia* **4**, 680-689 (1997).
23 Schmid, M., Evans, B. J. & Bogart, J. P. Polyploidy in *Amphibia*. *Cytogenetic and Genome Research* **145**, 315-330, doi:10.1159/000431388 (2015).
24 Comai, L. The advantages and disadvantages of being polyploid. *Nat Rev Genet* **6**, 836-846, doi:10.1038/nrg1711 (2005).
25 Mason, A. S. & Pires, J. C. Unreduced gametes: meiotic mishap or evolutionary mechanism? *Trends Genet* **31**, 5-10, doi:10.1016/j.tig.2014.09.011 (2015).
26 Pandian, T. J. K., R. Ploidy induction and sex control in fish. *Hydrobiologia* **384**, 167-243 (1998).
27 Stuart, S. N. *et al.* Status and trends of amphibian declines and extinctions worldwide. *Science* **306**, 1783-1786, doi:10.1126/science.1103538 (2004).
28 Collins, J. P. Amphibian decline and extinction: what we know and what we need to learn. *Dis Aquat Organ* **92**, 93-99, doi:10.3354/dao02307 (2010).
29 Hudson, M. A. *et al.* Dynamics and genetics of a disease-driven species decline to near extinction: lessons for conservation. *Sci Rep* **6**, 30772, doi:10.1038/srep30772 (2016).
30 O'Hanlon, S. J. *et al.* Recent Asian origin of chytrid fungi causing global amphibian declines. *Science* **360**, 621-627, doi:10.1126/science.aar1965 (2018).
31 Lemmon, A. R., Emme, S. A. & Lemmon, E. M. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst Biol* **61**, 727-744, doi:10.1093/sysbio/sys049 (2012).
32 Barrow, L. N., Lemmon, A. R. & Lemmon, E. M. Targeted Sampling and Target Capture: Assessing Phylogeographic Concordance with Genome-wide Data. *Syst Biol* **67**, 979-996, doi:10.1093/sysbio/syy021 (2018).
33 Heinicke, M. P., Lemmon, A. R., Lemmon, E. M., McGrath, K. & Hedges, S. B. Phylogenomic support for evolutionary relationships of New World direct-developing frogs (Anura: Terraranae). *Mol Phylogenet Evol* **118**, 145-155, doi:10.1016/j.ympev.2017.09.021 (2018).
34 Yuan, Z.-Y. *et al.* Natatanuran frogs used the Indian Plate to step-stone disperse and radiate across the Indian Ocean. *National Science Review*, nwy092-nwy092, doi:10.1093/nsr/nwy092 (2018).
35 Mirarab, S. & Warnow, T. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* **31**, i44-52, doi:10.1093/bioinformatics/btv234 (2015).
36 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).
37 Feng, Y. J. *et al.* Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous-Paleogene boundary. *Proc Natl Acad Sci U S A* **114**, E5864-E5870, doi:10.1073/pnas.1704632114 (2017).
38 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655-1664, doi:10.1101/gr.094052.109 (2009).

39 Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* **8**, e1002967, doi:10.1371/journal.pgen.1002967 (2012).
40 AmphibiaWeb. Information on amphibian biology and conservation. *Berkeley, California: AmphibiaWeb. Available:* http://amphibiaweb.org/. (2016).
41 Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978, doi:doi:10.1002/joc.1276 (2005).
42 Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231-259, doi:10.1016/j.ecolmodel.2005.03.026 (2006).
43 Novikova, P. Y. *et al.* Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nat Genet* **48**, 1077-1082, doi:10.1038/ng.3617 (2016).
44 Green, R. E. *et al.* A Draft Sequence of the Neandertal Genome. *Science* **328**, 710-722, doi:10.1126/science.1188021 (2010).
45 Nishihara, H., Maruyama, S. & Okada, N. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. *Proc Natl Acad Sci U S A* **106**, 5235-5240, doi:10.1073/pnas.0809297106 (2009).
46 Hallstrom, B. M. & Janke, A. Mammalian evolution may not be strictly bifurcating. *Mol Biol Evol* **27**, 2804-2816, doi:10.1093/molbev/msq166 (2010).
47 Garrigan, D. *et al.* Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. *Genome Res* **22**, 1499-1511, doi:10.1101/gr.130922.111 (2012).
48 Martin, S. H. *et al.* Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* **23**, 1817-1828, doi:10.1101/gr.159426.113 (2013).
49 Jonsson, H. *et al.* Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc Natl Acad Sci U S A* **111**, 18655-18660, doi:10.1073/pnas.1412627111 (2014).
50 Fontaine, M. C. *et al.* Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* **347**, 1258524, doi:10.1126/science.1258524 (2015).
51 Lamichhaney, S. *et al.* Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371-375, doi:10.1038/nature14181 (2015).
52 Suh, A., Smeds, L. & Ellegren, H. The Dynamics of Incomplete Lineage Sorting across the Ancient Adaptive Radiation of Neoavian Birds. *PLoS Biol* **13**, e1002224, doi:10.1371/journal.pbio.1002224 (2015).
53 Pease, J. B., Haak, D. C., Hahn, M. W. & Moyle, L. C. Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. *PLoS Biol* **14**, e1002379, doi:10.1371/journal.pbio.1002379 (2016).
54 Becak, M. L. & Becak, W. Further studies on polyploid amphibians (*Ceratophrydidae*). 3. Meiotic aspects of the interspecific triploid hybrid: *Odontophrynus cultripes* (2n=22) x *O. americanus* (4n=44). *Chromosoma* **31**, 377-385 (1970).
55 Main, A. R. *Comparisons of breeding biology and isolating mechanisms in Western Australian frogs*. (Melbourne Univ. Press 1962).
56 Nishioka, M. & Ueda, H. Studies on polyploidy in Japanese frogs. *Sci Rep Lab Amphibian Biol. Hiroshima Univ.* **6**, 207–252 (1983).
57 Bogart, J. P. & Bi, K. Genetic and genomic interactions of animals with different ploidy levels. *Cytogenet Genome Res* **140**, 117-136, doi:10.1159/000351593 (2013).
58 Geach, T. J., Stemple, D. L. & Zimmerman, L. B. Genetic analysis of *Xenopus tropicalis*. *Methods Mol Biol* **917**, 69-110, doi:10.1007/978-1-61779-992-1_5 (2012).

59 Lafon-Placette, C. *et al.* Endosperm-based hybridization barriers explain the pattern of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in Central Europe. *Proc Natl Acad Sci U S A* **114**, E1027-E1035, doi:10.1073/pnas.1615123114 (2017).
60 Lafon-Placette, C. & Kohler, C. Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. *Mol Ecol* **25**, 2620-2629, doi:10.1111/mec.13552 (2016).
61 Schmickl, R., Marburger, S., Bray, S. & Yant, L. Hybrids and horizontal transfer: introgression allows adaptive allele discovery. *J Exp Bot* **68**, 5453-5470, doi:10.1093/jxb/erx297 (2017).
62 McCartney-Melstad, E. & Shaffer, H. B. Amphibian molecular ecology and how it has informed conservation. *Mol Ecol* **24**, 5084-5109, doi:10.1111/mec.13391 (2015).
63 McCartney-Melstad, E., Gidis, M. & Shaffer, H. B. Population genomic data reveal extreme geographic subdivision and novel conservation actions for the declining foothill yellow-legged frog. *Heredity (Edinb)* **121**, 112-125, doi:10.1038/s41437-018-0097-7 (2018).
64 Prum, R. O. *et al.* A comprehensive phylogeny of birds (*Aves*) using targeted next-generation DNA sequencing. *Nature* **526**, 569, doi:10.1038/nature15697 (2015).
65 Rokyta, D. R., Lemmon, A. R., Margres, M. J. & Aronow, K. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics* **13**, 312, doi:10.1186/1471-2164-13-312 (2012).
66 Hamilton, C. A., Lemmon, A. R., Lemmon, E. M. & Bond, J. E. Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. *BMC Evol Biol* **16**, 212, doi:10.1186/s12862-016-0769-y (2016).
67 Pyron, R. A., Hsieh, F. W., Lemmon, A. R., Lemmon, E. M. & Hendry, C. R. Integrating phylogenomic and morphological data to assess candidate species-delimitation models in brown and red-bellied snakes (*Storeria*). *Zoological Journal of the Linnean Society* **177**, 937-949, doi:doi:10.1111/zoj.12392 (2016).
68 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).
69 Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649, doi:10.1093/bioinformatics/bts199 (2012).
70 Taucce, P. P. G. *et al.* The mitochondrial genomes of five frog species of the Neotropical genus *Ischnocnema (Anura: Brachycephaloidea: Brachycephalidae)*. *Mitochondrial DNA Part B* **3**, 915-917, doi:10.1080/23802359.2018.1501312 (2018).
71 Irisarri, I. *et al.* The origin of modern frogs (*Neobatrachia*) was accompanied by acceleration in mitochondrial and nuclear substitution rates. *BMC Genomics* **13**, 626, doi:10.1186/1471-2164-13-626 (2012).
72 Hahn, C., Bachmann, L. & Chevreux, B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research* **41**, e129-e129, doi:10.1093/nar/gkt371 (2013).
73 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).
74 Hillis, D. M., Heath, T. A. & St John, K. Analysis and visualization of tree space. *Syst Biol* **54**, 471-482, doi:10.1080/10635150590946961 (2005).
75 Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics (Oxford, England)* **20**, 289-290 (2004).
76 Maechler, M., Rousseeuw, P., Struyf, A. & Hubert, M. cluster: Cluster Analysis Basics and Extension. (2018).

77 Weiss, C. L., Pais, M., Cano, L. M., Kamoun, S. & Burbano, H. A. nQuire: a statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics* **19**, 122, doi:10.1186/s12859-018-2128-z (2018).
78 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
79 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
80 Roberts, J. D. Geographic Variation in Calls of Males and Determination of Species Boundaries in Tetraploid Frogs of the Australian Genus *Neobatrachus* (*Myobatrachidae*). **45**, 95-112, doi:10.1071/zo96006 (1997).
81 Mahony, M. Cytogenetic studies on Australian frogs of the family *Myobatrachidae*. Ph.D. thesis, Macquarie University, Sydney, Australia. (1986).
82 Choleva, L. & Janko, K. Rise and persistence of animal polyploidy: evolutionary constraints and potential. *Cytogenet Genome Res* **140**, 151-170, doi:10.1159/000353464 (2013).
83 Okamoto, T., Ohnishi, Y. & Toda, E. Development of polyspermic zygote and possible contribution of polyspermy to polyploid formation in angiosperms. *J Plant Res* **130**, 485-490, doi:10.1007/s10265-017-0913-9 (2017).
84 Pfeifer, B., Wittelsburger, U., Ramos-Onsins, S. E. & Lercher, M. J. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol* **31**, 1929-1936, doi:10.1093/molbev/msu136 (2014).
85 Hijmans, R. J., van Etten J. raster: Geographic analysis and modeling with raster data. R package version 2.0-12. (2012).