On the roles of landscape heterogeneity and environmental variation in determining population genomic structure in a dendritic system

Chris J. Brauer1 | Peter J. Unmack2 | Steve Smith1,3 | Louis Bernatchez4 | Luciano B. Beheregaray1

1Molecular Ecology Laboratory, College of Science and Engineering, Flinders University, Adelaide, South Australia, Australia
2Institute for Applied Ecology, University of Canberra, Canberra, Australian Capital Territory, Australia
3Department of Integrative Biology and Evolution, University of Veterinary Medicine, Vienna, Austria
4Institut de Biologie Intégrative et des Systèmes, Université Laval Québec, Québec, Quebec, Canada

Correspondence
Luciano B. Beheregaray, Molecular Ecology Laboratory, College of Science and Engineering, Flinders University, Adelaide, SA 5042, Australia.
Email: luciano.beheregaray@flinders.edu.au

Funding information
Australian Research Council, Grant/Award Number: FT130101068

Abstract
Dispersal and natural selection are key evolutionary processes shaping the distribution of phenotypic and genetic diversity. For species inhabiting complex spatial environments however, it is unclear how the balance between gene flow and selection may be influenced by landscape heterogeneity and environmental variation. Here, we evaluated the effects of dendritic landscape structure and the selective forces of hydroclimatic variation on population genomic parameters for the Murray River rainbowfish, *Melanotaenia fluviatilis* across the Murray–Darling Basin, Australia. We genotyped 249 rainbowfish at 17,503 high-quality SNP loci and integrated these with models of network connectivity and high-resolution environmental data within a riverscape genomics framework. We tested competing models of gene flow before using multivariate genotype–environment association (GEA) analysis to test for signals of adaptive divergence associated with hydroclimatic variation. Patterns of neutral genetic variation were consistent with expectations based on the stream hierarchy model and *M. fluviatilis*’ moderate dispersal ability. Models incorporating dendritic network structure suggested that landscape heterogeneity is a more important factor determining connectivity and gene flow than waterway distance. Extending these results, we also introduce a novel approach to controlling for the unique effects of dendritic network structure in GEA analyses of populations of aquatic species. We identified 146 candidate loci potentially underlying a polygenic adaptive response to seasonal fluctuations in stream flow and variation in the relative timing of temperature and precipitation extremes. Our findings underscore an emerging predominant role for seasonal variation in hydroclimatic conditions driving local adaptation and are relevant for informing proactive conservation management.

KEYWORDS

1 | INTRODUCTION

Gene flow and selection are key evolutionary processes regulating the potential for adaptive divergence among populations (Lenormand, 2002). The balance between gene flow and selection is affected by a range of factors including life history, population size, landscape heterogeneity and environmental variation. Growing evidence suggests that most adaptive genomic responses to...
environmental variation are polygenic in nature (Bernatchez, 2016; Pritchard & Di Rienzo, 2010). For species inhabiting complex spatial environments, landscape structure is also expected to greatly impact patterns of demographic connectivity and genetic diversity (Davis, Epps, Fil ictroft, & Banks, 2017; Thomaz, Christie, & Knowles, 2016). Following the emergence of the field of landscape genetics (Manel, Schwartz, Luikart, & Taberlet, 2003), significant research effort has been focused on understanding how environmental heterogeneity affects patterns of gene flow and spatial population structure (Manel & Holderegger, 2013). More recently, landscape genomics approaches have also provided information concerning the environmental determinants of adaptive population divergence. Many analytical challenges remain however, particularly when attempting to detect a signal of adaptive divergence against the backdrop of complex spatial environments such as dendritic river networks (Fourcade, Chaput-Bardy, Secondi, Fleurant, & Lemaire, 2013). Accordingly, landscape genomics studies should ideally include analyses that both maximize the likelihood of detecting a polygenic signal of adaptation and also provide the means to understand, and control for, spatial patterns of connectivity and population structure.

Recent simulations that compared genotype–environment association (GEA) approaches commonly used to identify multilocus adaptation (Forester, Lasky, Wagner, & Urban, 2018) suggested that multivariate constrained ordination methods, such as redundancy analysis (RDA), may offer the best balance between low false positive and high true positive rates. These methods are relatively robust across a range of demographic scenarios and can account for spatial genetic structure without assuming specific population models that are almost certainly violated in complex spatial environments. Forester et al. (2018) found that controlling for spatial structure in their RDA analyses was of little benefit in systems with low population structure (average global $F_{ST} = 0.05$) and was even detrimental for some scenarios. Their study did not include simulations with higher population structure however, and it seems likely that in such cases, controlling for spatial population structure may be more important, although perhaps at the cost of reduced power to identify true positives.

In addition to searching for the genomic signal of local adaptation, landscape genomics can also be used to understand how landscape structure contributes to spatial variation in connectivity and gene flow. This is particularly important in spatially heterogeneous landscapes where simple models of gene flow such as isolation by distance (IBD) may be inadequate to explain the observed population structure. Dendritic river networks are characterized by complex patterns of habitat heterogeneity, and population structure in these systems is often not well explained by IBD (Campbell Grant, Lowe, & Fagan, 2007). In such cases, the stream hierarchy model (SHM; Meffe & Vrijenhoek, 1988) may provide a more appropriate hypothesis for making predictions about the spatial distribution of genetic variation. Under the SHM, we expect to observe hierarchical genetic population structure that is also consistent with river network structure. Here, populations restricted to tributaries in adjacent catchments should exhibit reduced genetic diversity and increased population divergence relative to larger populations further downstream (Meffe & Vrijenhoek, 1988). Predictions based on the SHM have been tested in a number of empirical studies with varying levels of support (Brauer, Unmack, Hammer, Adams, & Beheregaray, 2013; Hopken, Douglas, & Douglas, 2013; Huey, Baker, & Hughes, 2006; Lean, Hammer, Unmack, Adams, & Beheregaray, 2016; Tonkin et al., 2017). However, few studies have analyzed competing models of connectivity within a riverscape genomics framework that also assesses environmental heterogeneity. In this context, incorporating the SHM into GEA analyses may improve inferences of local adaptation by better accounting for the unique spatial structure of river networks than more simple population models.

The Murray–Darling Basin (MDB) is an ideal system for testing the combined effects of dendritic network structure and environmental variation on patterns of biodiversity. One of the largest river basins in Australia, the MDB, covers about 14% of the continent. It spans a range of hydroclimatic environments from arid to wet, temperate to subtropical (Figure 1) and is of high ecological value with many endemic and threatened species (Murray–Darling Basin Authority 2010). The region is, however, currently undergoing rapid hydrological changes due to a combination of human development and altered climate regime (Leblanc, Tweed, Van Dijk, & Timbal, 2012), and a recent assessment of ecosystem health rated the majority of the MDB as either poor or very poor condition (Davies, Harris, Hillman, & Walker, 2010). Extensive agricultural and urban development has resulted in wholesale habitat degradation due to water abstraction, flow regulation, reduced water quality, introduced species and the widespread construction of in-stream barriers (Balcombe et al., 2011). As a result of these impacts, the MDB is arguably one of the most severely fragmented and degraded ecosystems in Australia (Kingsford, 2000), with over half of the basin’s native fish species now considered threatened (Lintermans, 2007). Riverscape genomic studies of MDB fishes to date have focused on species with either extremely low (Brauer, Hammer, & Beheregaray, 2016) or extremely high dispersal abilities (Attard et al., 2018; Harrisson et al., 2017). These studies were consistent in highlighting the importance of hydroclimatic variation in shaping patterns of adaptive divergence among populations. Support for a general effect of landscape structure on connectivity is less clear however, likely due to variations in life history, population size and dispersal capacity. In this case, examining a relatively abundant generalist species with intermediate natural dispersal ability may increase the generality of previous findings.

Rainbowfishes (Melanotaeniidae) are one of the most species-rich freshwater fish families in New Guinea and Australia, with most Australian species occurring in the subtropical or tropical north (Unmack, Allen, & Johnson, 2013). The focus of this study, the Murray River rainbowfish, *Melanotaenia fluviatilis*, occurs further south than any other rainbowfish and is the only temperate rainbowfish species. Considered a generalist species, the Murray River rainbowfish occupies a range of stream and wetland habitats and possesses...
moderate dispersal capacity (Baumgartner & Harris, 2007; McGuigan, Zhu, Allen, & Moritz, 2000). While relatively common in the northern MDB, they are less abundant in the Murray River where their southern range margin is thought to be limited by cooler winter temperatures (Crowley, Ivantsoff, & Allen, 1986). In this study, we applied genotype-by-sequencing (GBS) within a riverscape genomics framework to test two main hypotheses concerning (a) the role of habitat heterogeneity in determining spatial variation in connectivity and gene flow and (b) the environmental factors influencing adaptive divergence among populations. First, using a reduced neutral SNP data set, we assess patterns of genomewide diversity for *M. fluviatilis* in the context of expectations based on the SHM. Using multiple matrix regression, we test the hypothesis that models of gene flow that incorporate the natural dendritic hydrological structure will outperform those based on geographic distance (i.e., IBD). Second, we incorporate the SHM into a novel multivariate constrained ordination GEA approach to test the hypothesis that hydroclimatic variation contributes to adaptive divergence of *M. fluviatilis* populations across the MDB. Our integrated riverscape genomics framework provides novel insight into how landscape heterogeneity and environmental variation together modulate key evolutionary processes to shape the genomic architecture of riverine species.

### 2 | METHODS

#### 2.1 | Sampling and genomic data collection

Climate, and in particular rainfall across the MDB, is highly temporally and spatially variable. The northern MDB is characterized by unpredictable summer rainfall, while winter rainfall dominates in the south. Average annual rainfall across the basin is generally low, but ranges in extremes from >1,500 mm in the southeast highlands to <200 mm in the west (Chiew et al., 2008). A total of 249 *M. fluviatilis* samples were collected from 14 locations between 2009 and 2012. These were selected to capture maximum hydroclimatic variation across the MDB, along with potential spatial population structure within, and between the two major catchments of the Murray and Darling Rivers (Figure 1; Table 1). Fish were ethically euthanized using clove oil, snap frozen in liquid nitrogen and stored at −70°C in the Australian Biological Tissues Collection at the South Australian Museum, Adelaide.

DNA extractions were performed following a modified salting-out protocol (Sunnucks & Hales, 1996). DNA integrity and purity were assessed using gel electrophoresis and a NanoDrop 1000 spectrophotometer (Thermo Scientific), respectively. Sequencing libraries
were prepared based on a double-digest GBS approach (Poland, Brown, Sorrells, & Jannink, 2012) using the restriction enzymes PstI and Msel. Using custom individual barcodes to multiplex 48 samples per lane, libraries were randomly assigned to each of six Illumina HiSeq2000 lanes and sequenced as single-end, 100-bp reads. Raw sequencing data were demultiplexed using the process_radtags module from STACKS 1.35. dDocent 2.18 (Puritz, Hollenbeck, & Gold, 2014) was used for de novo assembly of a reference catalogue and genotyping. The resulting multisample variant call file was filtered using vcfTools (Danecek et al., 2011) to retain only bi-allelic SNP loci present in at least 90% of individuals in all populations with a minimum minor allele frequency of 0.05, before the following series of filtering steps were then applied in order to remove SNPs likely to be the result of sequencing errors, paralogs, multicopy loci and artefacts of library preparation. These steps are based on scripts available on the dDocent GitHub page (https://github.com/jpuritz/dDocent nt/). (a) Allele balance: for each locus, it should be expected that an approximately equal number of reads for the reference and alternate alleles for individuals are called as heterozygotes. Loci were therefore removed if the proportion of alternate to reference allele was <0.25 or >0.75 across all heterozygote individuals. (b) Mapping quality: as both alleles of a bi-allelic locus should start from the same restriction enzyme cut site, mapping quality scores for the two alleles should be similar. Loci with a mapping quality score ratio (alternate allele mapping score/reference allele mapping score) <90% or >110% were therefore discarded. (c) Read quality: loci with overall low read quality scores (<25% of read depth) were discarded. Additionally, Li (2014) found a predictable relationship between Illumina read quality scores and read depth, such that where loci are covered by a high number of reads, quality scores are likely to be inflated. In this case, a higher quality score threshold is required to distinguish true variants from errors. Consequently, for loci with unusually high read depths (greater than the mean depth plus three times the square root of the mean), those with quality scores less than two times their read depth were also removed. (d) Read depth: the read depth of each locus was calculated and the frequency distribution of mean depth per locus, averaged over all individuals was used as a guide to remove loci with abnormally high coverage. Individual samples were allowed a maximum of 20% missing data.

### Table 1

<table>
<thead>
<tr>
<th>Site</th>
<th>Location</th>
<th>N</th>
<th>H_E</th>
<th>H_O</th>
<th>%</th>
<th>F IS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBR (M)</td>
<td>Murray R., Murray Bridge</td>
<td>20</td>
<td>0.192</td>
<td>0.188</td>
<td>84.1</td>
<td>0.018</td>
</tr>
<tr>
<td>MDC (M)</td>
<td>Murray–Darling confluence</td>
<td>22</td>
<td>0.238</td>
<td>0.238</td>
<td>92.1</td>
<td>-0.003</td>
</tr>
<tr>
<td>GGL (M)</td>
<td>Murray R., Gol Gol</td>
<td>30</td>
<td>0.232</td>
<td>0.229</td>
<td>93.3</td>
<td>0.011</td>
</tr>
<tr>
<td>WAK (T)</td>
<td>Wakti Kt., Kotupna</td>
<td>16</td>
<td>0.116</td>
<td>0.113</td>
<td>36.7</td>
<td>0.018</td>
</tr>
<tr>
<td>BEN (T)</td>
<td>Broken R., Benalla</td>
<td>14</td>
<td>0.117</td>
<td>0.111</td>
<td>37.1</td>
<td>0.032</td>
</tr>
<tr>
<td>BOG (M)</td>
<td>Bogan R., Bourke</td>
<td>16</td>
<td>0.245</td>
<td>0.221</td>
<td>84.2</td>
<td>0.069</td>
</tr>
<tr>
<td>PEL (T)</td>
<td>Peel R., Caroll</td>
<td>9</td>
<td>0.166</td>
<td>0.164</td>
<td>53.7</td>
<td>0.007</td>
</tr>
<tr>
<td>GWY (T)</td>
<td>Gwydir R., Bingara</td>
<td>12</td>
<td>0.169</td>
<td>0.157</td>
<td>58.9</td>
<td>0.048</td>
</tr>
<tr>
<td>DUM (T)</td>
<td>Dumaresq R., Texas</td>
<td>14</td>
<td>0.157</td>
<td>0.147</td>
<td>54.1</td>
<td>0.037</td>
</tr>
<tr>
<td>MIB (T)</td>
<td>McIntyre Brook, Inglewood</td>
<td>20</td>
<td>0.163</td>
<td>0.164</td>
<td>57.2</td>
<td>-0.010</td>
</tr>
<tr>
<td>STG (T)</td>
<td>Canning Kt., Stonehenge</td>
<td>20</td>
<td>0.149</td>
<td>0.152</td>
<td>52.3</td>
<td>-0.025</td>
</tr>
<tr>
<td>OAK (T)</td>
<td>Oakley Kt., Bowenville</td>
<td>18</td>
<td>0.317</td>
<td>0.308</td>
<td>84.5</td>
<td>0.022</td>
</tr>
<tr>
<td>WAR (T)</td>
<td>Condamine R., Warwick</td>
<td>20</td>
<td>0.312</td>
<td>0.299</td>
<td>84.3</td>
<td>0.027</td>
</tr>
<tr>
<td>KIL (T)</td>
<td>Farm Kt., Killarney</td>
<td>18</td>
<td>0.303</td>
<td>0.275</td>
<td>83.1</td>
<td>0.061</td>
</tr>
</tbody>
</table>

Note. Streams were classified as either main channel (M) or tributaries (T) according to position in the stream network to aid with interpretation of results in context with the stream hierarchy model (Meffe & Vrijenhoek, 1988). Bonferroni corrected p-Values for F IS were all >0.05.

### 2.2 Genetic diversity and neutral population structure

Population structure and demographic parameters should normally be assessed using loci conforming to neutral expectations (Allendorf, Hohenlohe, & Luikart, 2010; Luikart, England, Tallmon, Jordan, & Taberlet, 2003). To define a putatively neutral data set, we used BAYEscan 2.1 (Foll & Gaggiotti 2008) to detect outlier loci, as it performs well where complex demographic scenarios may deviate from the underlying model (Foll & Gaggiotti 2006, 2008). The software was run for 100,000 iterations with prior odds of 10,000. Loci with a q-value <0.1 (false discovery rate [FDR] 10%) were considered outliers, and the remaining SNPs were examined for departure from expectations of Hardy–Weinberg equilibrium (HWE) using GENODIVE 2.0b27 (Meirmans & Van Tienderen, 2004). Loci out of HWE at a FDR of 10% in more than 50% of populations were subsequently removed (along with candidate adaptive loci identified in the GEA analysis; see below) and the remaining, putatively neutral SNPs were used for estimating genetic diversity, demographic parameters and population structure.

Expected heterozygosity (H_E), observed heterozygosity (H_O), percentage of polymorphic loci and inbreeding coefficient (F IS) were calculated for each sampling site based on the neutral SNPs using GENODIVE. Population differentiation was assessed by estimating pairwise FST (Weir & Cockerham, 1984) among sampling sites using GENODIVE, with significance assessed using 10,000 permutations. GENODIVE was also used to perform a hierarchical AMOVA based on F ST among major river catchments, among sites within catchments and among individuals within sites using 10,000 permutations. Missing data were replaced with alleles drawn randomly from the overall allele frequency distribution.

Population structure was examined using the spatially explicit ancestry estimation method of TESS3 (Caye, Deist, Martins, Michel, & François, 2016). This method does not make assumptions concerning HWE or linkage disequilibrium suggesting it should perform well where landscape heterogeneity may result in complex spatial patterns of dispersal and population structure. The number of ancestral
populations ($K$) was evaluated using 10 independent runs of each $K$ ($K = 1–14$) before using a cross-validation procedure to select the best value of $K$ according to the asymptote in the plot of cross-validation scores (Caye et al., 2016). Admixture coefficients were plotted using *DISTRUCT* (Rosenberg, 2004).

We used *BAYEAS* 3.0.4 (Wilson & Rannala, 2003), modified to allow analysis of large SNP data sets (https://github.com/smussma nn82/BayesAss3-SNPs) to estimate recent migration among populations. *BAYEAS* implements a Bayesian MCMC resampling method to estimate asymmetrical rates of recent migration, where migration ($m$) is the proportion of each population having migrant ancestry. First-generation migrants, or the offspring of at least one first-generation migrant, are considered as having migrant ancestry. The software was run for 10 million iterations with a 1 million iteration burn-in. Mixing parameters for allele frequencies, inbreeding coefficients and migration rate were adjusted to achieve optimum acceptance rates of 20%–40% (Wilson & Rannala, 2003). Convergence was confirmed by plotting the cumulative log likelihoods of the iterations using the program *TRACER* 1.7 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018), and five runs were performed to ensure consistency.

### 2.3 Models of population connectivity: spatial vs. dendritic structure

The physical structure of dendritic river networks is well known to greatly affect patterns of genetic variation of stream-dwelling organisms (Fourcade et al., 2013; Hughes, Schmidt, & Finn, 2009; Morrissey & de Kerckhove, 2009; Thomaz et al., 2016). These patterns may be further influenced by the highly variable hydroclimatic conditions that characterize the MDB leading to decadal cycles of population isolation punctuated by occasional long-distance dispersal events facilitated by infrequent flooding (Attard et al., 2016; Brauer et al., 2016; Cole et al., 2016; Faulks, Gilligan, & Beheregaray, 2010a). To determine the contribution of landscape heterogeneity to observed patterns of population structure, we compared models of gene flow based on geographic distance (IBD) with models incorporating the natural dendritic hydrological structure using multiple matrix regression with randomization (MMRR; Wang, 2013). This method uses multiple regression to evaluate how genetic distance responds to multiple independent variables such as geographic or environmental distance matrices. To assess IBD, pairwise population $F_{ST}$ was regressed against pairwise waterway distances calculated with *ARCMAP* 10.3. To assess the influence of dendritic structure, we estimated $F_{ST}$ for individual stream sections following the StreamTree model of Kalinowski, Meeuwig, Narum, and Taper (2008). This method models genetic distances among populations as the sum of all pairwise genetic distances mapped to each section of the stream network independently of the length of each section (Kalinowski et al., 2008). In this way, the effects of distance are separated from the effects of landscape heterogeneity in identifying reaches of the stream network that contribute most to restricting gene flow (i.e., due to dendritic structure, in-stream barriers, tributary-main channel confluences or other unknown landscape effects). Model fit was assessed by plotting the StreamTree fitted distance against observed $F_{ST}$ and calculating the regression coefficient of determination ($R^{2}$). This model was then compared with the model of IBD, again using MMRR. All distance matrices were $z$-transformed to facilitate direct comparison of partial regression coefficients (Schielzeth, 2010) and in each case model significance was assessed using 10,000 random permutations.

### 2.4 Local hydroclimatic conditions and adaptive population divergence

Hydroclimatic variation across the MDB was summarized by performing principal component analysis (PCA) on 16 environmental variables already identified as predictors of adaptive genetic variation for freshwater fishes in this region (Attard et al., 2018; Brauer et al., 2016). These data are linked to a 9-s digital elevation model-derived stream network (~250 m resolution) and were obtained from the Australian hydrological geospatial fabric (Geoscience Australia 2011; Stein, Hutchinson, & Stein, 2014). Separate PCAs were performed for three groups of variables describing variation in (a) temperature, (b) precipitation and (c) stream flow, to aid interpretation of the results. The PCAs were carried out using the *FACTOextra* R package (Lê, Josse, & Husson, 2008) and principal components (PCs) with eigenvalues greater than one were retained as predictors for the RDA. The retained environmental PCs were subjected to a forward selection procedure using the *packfor* R package (Dray, Legendre, & Blanchet, 2016) to remove any nonsignificant ($p > 0.001$) PCs from the model. Variance inflation factor (VIF) analysis was then used to exclude highly correlated PCs using a VIF threshold of 10 (Dyer, Nason, & Garrick, 2010).

Multivariate GEA methods such as RDA are well suited to detecting small changes in allele frequencies of many covarying loci spread throughout the genome (Bourret, Dionne, Kent, Lien, & Bertrame, 2013) as expected for a polygenic response to selection (Forester et al., 2018; Le Corre & Kremer, 2012). As it is thought most ecologically important traits may evolve via polygenic adaptation (Pritchard & Di Rienzo, 2010), we employed RDA to detect associations between SNP loci and the environment, as summarized by the hydroclimatic PCs. As population structure may confound inferences of selection, we performed two RDAs exploiting complimentary methods to account for different aspects of spatial population structure. For the first RDA, we modelled broad landscape-scale spatial effects by calculating a set of spatial vectors describing the distribution of sampling sites across a range of spatial scales. Multidimensional scaling (MDS) was first applied to a matrix of pairwise waterway distances between sites to provide transformed coordinates that better represent the hydrological distance between sites. The new coordinates were then expressed as third-order orthogonal polynomials to account for nonlinear spatial patterns as expected under the SHM, following the method of Meirmans (2015). A spatial filtering procedure modified from Forester et al. (2018) was then performed to determine which polynomials to include in the model. The spatial polynomials were then assessed for correlation with the environmental variables and those with Pearson correlation...
coefficients <0.5 for all environmental PCs were retained as conditioning variables in the partial RDA model.

In performing a second RDA, we explored the possibility of using the StreamTree model to control for spatial population structure. We propose that this method offers several advantages in riverine systems by incorporating the complex patterns of spatial structure unique to dendritic river networks, as well as restrictions to connectivity due to barriers and other potentially unknown sources of resistance. We again used MDS, this time to transform the pairwise distances estimated by the StreamTree model into coordinates for input to the RDA as conditioning variables. The final partial RDA models assessed variation in individual SNP genotypes constrained by the retained environmental PCs after controlling for the effects of spatial structure. In both cases, significance of the full model, each axis and marginal significance of each environmental PC, was assessed using 1,000 permutations. The mean locus score across all loci was calculated for each significant (p < 0.05) RDA axis, and individual loci with a score greater than three standard deviations from the mean were considered candidates for selection (Forester, Jones, Joost, Landguth, & Lasky, 2015). Custom R scripts used for the environmental and spatial filtering, and the RDAs are available on Dryad: https://doi.org/10.5061/dryad.t2v8825.

2.5 Functional annotation of candidate loci

Annotation information and gene ontology (GO) terms associated with the SNP loci were examined using BLAST2GO (Conesa et al. 2005). A BLAST search and annotation of the flanking sequences for all 17,503 SNPs was performed against the NCBI nonredundant nucleotide database with the BLAST e-value threshold set to 1 × 10^{-3} and an annotation threshold e-value threshold of 1 × 10^{-6}. Enrichment of GO terms in the strong candidate data set was assessed relative to all annotated SNPs using Fisher’s exact test with a FDR of 0.05, and CATEGORIZER (Hu, Bao, & Reecy, 2008) was used to summarize GO terms assigned to the candidate loci according to the GO-Slim classification method.

3 RESULTS

3.1 Sequencing quality and genetic diversity

After demultiplexing, a total of 645,811,728 raw sequencing reads were recovered and following quality trimming 645,506,093 reads (mean per sample = 2,592,394, min = 726,784, max = 5,927,097) were retained (Supporting Information Table S1). After filtering, 17,503 SNP loci were retained from the 537,180 variant sites present in the whole dDocent catalogue (Table 2). BAYESCAN identified 706 FST outlier loci (239 and 467 putatively under divergent and balancing selection, respectively) and after excluding these and the GEA candidate loci (see below), and filtering for HWE, 16,165 putatively neutral SNP loci remained (Table 2). These data were used for all downstream analyses excluding the GEA test which was performed using all 17,503 SNPs.

Estimates of genetic diversity varied across sites, with an average observed heterozygosity (H) of 0.205 (0.116–0.317), average observed heterozygosity of 0.198 (0.111–0.317) and an average of 68.3% (36.7%–93.3%) polymorphic loci. None of the Bonferroni corrected FST estimates were significantly different from zero (Table 1). In general, genetic variation was highest for sites from the Condamine River (OAK, WAR and KIL) and main channel sites along the Darling and Murray Rivers (BOG, MBR, MDC and GGL). On the other hand, headwater tributary sites showed the lowest diversity, particularly in the Murray River (WAK and BEN).

3.2 Population structure within and among river catchments

There was substantial population structure across the basin. Pairwise FST estimates among sampling locations were all significant (p < 0.006) and ranged from 0.003 between adjacent sites MDC and GGL to 0.489 between upper Murray River site WAK and upper Darling River site OAK (Supporting Information Table S2). Results from AMOVA fit the predictions from the SHM, with most of the total variation partitioned between the two major sub-basins (i.e., 26.3% between the Murray and Darling Rivers, p < 0.001), and with less but also significant variation partitioned between sites within the two major rivers (4.9%, p < 0.001) and among individuals within sites (68.9%, p < 0.001) (Table 3).

Based on cross-validation scores, the clustering analysis performed with tESS3 identified K = 6 as the most likely number of ancestral populations (Figure 2c; Supporting Information Figure S1); however, assessment of a range of K values revealed several levels of hierarchical structure consistent with the SHM (Figure 2: Supporting Information Figure S2). Figure 2a (K = 2) separates the three isolated Condamine River sites in the upper reaches of the Darling River from the rest of the MDB, while K = 3 separates the Murray and Darling rivers and indicates that sites downstream of the

<table>
<thead>
<tr>
<th>Step</th>
<th>SNP count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw SNP catalogue</td>
<td>537,180</td>
</tr>
<tr>
<td>Genotyped in</td>
<td></td>
</tr>
<tr>
<td>≥90% of individuals, base quality ≥30, minor allele count of 3</td>
<td>137,714</td>
</tr>
<tr>
<td>Bi-allelic only</td>
<td>113,250</td>
</tr>
<tr>
<td>Single SNP per locus</td>
<td>33,787</td>
</tr>
<tr>
<td>Sequencing errors, paralogs, multicopy loci and artefacts of library preparation</td>
<td></td>
</tr>
<tr>
<td>1) Allele balance</td>
<td>29,711</td>
</tr>
<tr>
<td>2) Mapping quality</td>
<td>27,556</td>
</tr>
<tr>
<td>3) Read quality</td>
<td>26,845</td>
</tr>
<tr>
<td>4) Read depth, MAF &gt; 0.05</td>
<td>17,503</td>
</tr>
<tr>
<td>Putatively neutral in Hardy–Weinberg equilibrium</td>
<td>16,165</td>
</tr>
</tbody>
</table>
Murray–Darling confluence are more influenced by gene flow from the Murray than the Darling River (Figure 2b).

BAYESASS indicated very low levels of recent migration among most demes with the 95% confidence intervals for only one pairwise estimate not including zero (proportion of migrant ancestry at MDC from GGL, 0.199; Supporting Information Table S3).

3.3 | Models of population connectivity

Results of the MMRR tests indicated that by accounting for landscape heterogeneity, the StreamTree model was a far better model of *M. fluviatilis* population differentiation than the simple IBD model (Figure 3; Table 4). The StreamTree model distance was a good predictor of observed genetic distance ($R^2 = 0.976$, $p < 0.0001$). In contrast, the test for IBD showed a lower, but significant relationship between $F_{ST}$ and waterway distance between sites ($R^2 = 0.337$, $p < 0.001$). Model fit was not improved by including both StreamTree distance and geographic distance, and only StreamTree distance remained significant in the full model (Table 4). Figure 4 provides a visual–spatial representation of the stream sections inferred by the StreamTree model as most restricting dispersal, with sections colour-coded according to modelled distance (yellow represents a local StreamTree distance of $<0.01$, orange: $0.01$–$0.03$ and red: $>0.03$).

3.4 | Genotype–environment association analysis

The 16 hydroclimatic variables considered for the GEA analysis included five temperature variables (average annual mean temperature, coldest month minimum temperature, hottest month maximum temperature, driest quarter mean temperature and wettest quarter mean temperature), six precipitation variables (average annual mean rainfall, driest quarter mean rainfall, wettest quarter mean rainfall, warmest quarter mean rainfall, coldest quarter mean rainfall and average rainfall erosivity) and five flow-related variables (annual mean runoff, annual runoff coefficient of variation, monthly runoff coefficient of variation, runoff perennality and runoff skewness) (Supporting Information Table S4). The first two components of each of the three environmental PCAs (temperature, precipitation and flow) explained $85.8\%$, $98.0\%$ and $88.4\%$ of the total variation, respectively (Supporting Information Table S5). The major hydroclimatic gradients across the MDB (Figure 1a–f) indicate sites in the lower Murray experience higher temperatures during the dry season than Darling River sites where maximum temperatures coincide with wetter periods (Figure 1a,b; Supporting Information Table S5). Precipitation is generally higher for headwater sites across the whole MDB, with those in the Murray River receiving most rainfall during the cooler months (Figure 1c,d; Supporting Information Table S5). Stream flow is more variable, both within and among years in the tributaries compared to those further downstream, closer to the main channel (Figure 1e,f; Supporting Information Table S5). Following VIF analysis, temperature PC1 and PC2, precipitation PC2 and flow PC2 were retained for the RDA models.

The spatial filtering procedure resulted in the retention of three of the nine spatial polynomials as conditioning variables for the first RDA (Supporting Information Table S6). The RDA model was globally significant ($p < 0.001$) and indicated that seasonal variation in flow, precipitation during the coldest quarter and temperature during the wettest quarter explained 5% of the total genetic variation after accounting for spatial structure, which explained 33% of the total variation. The first two RDA axes were significant ($p < 0.05$) and explained 46.3% and 29.6% of the constrained variation (portion of total genetic variation explained by the environment), respectively (Figure 5a). Permutation tests revealed that each explanatory variable was significant in the model ($p < 0.001$) with flow PC2 (seasonal variation in runoff) accounting for the highest proportion of constrained variation (35.2%), followed by temperature PC2 (maximum temperature of hottest month, 24.3%), temperature PC1 (temperature during the wettest quarter, 22.4%) and precipitation PC2 (rainfall during the coldest quarter, 18.1%). Individual locus scores for 261 SNP loci were more than three standard deviations from the mean for the RDA1 and RDA2 axes.

The StreamTree RDA was globally significant ($p < 0.001$), and environmental variation explained 4% of the total genetic variation (Figure 5b). Each of the four environmental PCs were significant in the model ($p < 0.001$) with flow PC2 (seasonal variation in runoff) again accounting for the highest proportion of constrained variation (28.8%), followed by temperature PC1 (temperature during the wettest quarter, 21.9%), temperature PC2 (maximum temperature of hottest month, 19.7%) and precipitation PC2 (rainfall during the coldest quarter, 12.0%). The StreamTree model accounted for 33% of the total variation. A total of 710 SNP loci were more than three standard deviations from the mean locus scores across the first four RDA axes, which explained 35.4%, 33.3%, 20.2% and 11.2% of the constrained variation, respectively ($p < 0.001$). Comparing results for the two RDAs revealed 146 loci were identified in both tests and these SNPs were conservatively considered as strong candidate loci contributing to adaptive divergence of *M. fluviatilis* across the MDB.

BLAST2GO reported blast hits for 3,057 of the 17,503 loci, of which 1,188 could be assigned GO terms. The 146 GEA candidate loci scored blast hits for 28 loci, of which five were assigned GO terms. Results of the Fisher’s exact test indicated no GO terms were significantly (FDR 0.05) enriched in the candidate data set. The most common terms, however, included biological processes related to metabolism (GO:0008152), signal transduction (GO:0007165), cell communication (GO:0007154) and nucleic acid metabolism (GO:0006139; GO:0006259), and molecular functions concerning

| TABLE 3 | Hierarchical analysis of molecular variance (AMOVA) based on $F_{ST}$ for *Melanotaenia fluviatilis* from the Murray–Darling Basin |
|----------|----------------------|------------------|
| Source of variation | % Variance | $p$ Value |
| Between Murray and Darling rivers | 26.3 | 0.001 |
| Among sites within rivers | 4.9 | 0.001 |
| Among individuals within sites | 68.9 | 0.001 |
catalytic activity (GO:0003824) and binding (GO:0005488) (Supporting Information Appendix S1).

4 | DISCUSSION

We were able to identify key roles for both spatial and environmental variation in shaping patterns of genetic diversity and adaptive divergence of populations inhabiting a complex dendritic network.

Spatial patterns of neutral genetic variation and population connectivity were consistent with the SHM with generally low genetic diversity in headwater populations relative to those further downstream. Hierarchical population structure congruent with river network structure was also in line with expectations of the SHM and with the species moderate dispersal ability. Our hypothesis that accounting for dendritic hydrological structure in models of gene flow would improve predictions of population differentiation over simple IBD was strongly supported, suggesting landscape

FIGURE 2  Admixture plots based on 16,165 neutral SNP loci for Melanotaenia fluviatilis from the Murray–Darling Basin (MDB) depicting (a) K = 2, (b) K = 3 and (c) the most likely number of clusters determined by cross-validation procedure using TESS3, K = 6 [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 3  Multiple matrix regression with randomization (MMRR) plots for (a) isolation by distance (IBD) and (b) StreamTree analyses. The IBD plot depicts the relationship between pairwise $F_{ST}$ based on 16,165 neutral SNPs and riverine distance between sampling sites ($R^2 = 0.337, p = 0.0002$). The StreamTree plot compares fitted distance based on the StreamTree model with the observed pairwise $F_{ST}$ values ($R^2 = 0.976, p = 0.0001$)
TABLE 4 Results of multiple matrix regression with randomization tests based on 16,165 neutral SNP loci for the relationship between pairwise genetic distance ($F_{ST}$), and geographic distance (IBD), StreamTree model distance, and a model including both geographic and StreamTree distances

<table>
<thead>
<tr>
<th>Model</th>
<th>Variable</th>
<th>Coefficient</th>
<th>$R^2$</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBD</td>
<td></td>
<td>0.616</td>
<td>0.337</td>
<td>0.0002</td>
</tr>
<tr>
<td>StreamTree</td>
<td></td>
<td>0.940</td>
<td>0.976</td>
<td>0.0001</td>
</tr>
<tr>
<td>IBD + StreamTree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IBD</td>
<td>−0.022</td>
<td>0.976</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>StreamTree</td>
<td>0.998</td>
<td></td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Note. p-Values <0.001 are indicated in bold.

FIGURE 4 Spatial representation of the dendritic stream network connecting Melanotaenia fluviatilis sampling locations in the Murray–Darling basin (shaded). Stream sections are colour-coded according to StreamTree estimated $F_{ST}$ where yellow sections contribute little to restricting dispersal, orange sections offer intermediate resistance to dispersal, and red stream sections are those that most inhibit connectivity [Colour figure can be viewed at wileyonlinelibrary.com]

heterogeneity should be routinely considered when assessing gene flow among populations inhabiting complex spatial environments. Moreover, results of the GEA analyses support the hypothesis that hydroclimatic variation is driving patterns of adaptive divergence among M. fluviatilis populations across the MDB. Seasonal variation in stream flow, along with variation in the relationship between temperature and precipitation extremes, were the primary hydroclimatic variables associated with variation of 146 candidate adaptive loci.

4.1 Genomic signal of adaptive divergence

Results of the two RDAs provide consistent evidence for the importance of several key hydroclimatic variables in shaping patterns of adaptive divergence of M. fluviatilis populations across the MDB. Seasonal variation in flow (flowPC2) was the most influential factor in both models and appears particularly important in the divergence of OAK from WAR and KIL, and STG from MIB, despite their relative spatial proximity. Stream flow is highly seasonal and variable for both STG and OAK, with only ~3% of annual flow occurring during the driest 6 months at OAK and <1% for STG (average for all other sites is 11.1%). In contrast, WAR, KIL and MIB all have less seasonal flow regimes with >15% of annual runoff occurring during the driest 6 months. Variation in the temporal relationship between temperature and precipitation extremes also stands out as a central element of selection for this species. In addition to similar flow regimes, OAK and STG also share other similar hydroclimatic conditions with relatively cool minimum temperatures, hot maximum temperatures and rainfall mainly occurring during the warmest months. This is again important in driving adaptive divergence between OAK and the other two Condamine River sites with OAK experiencing warmer temperatures (maximum temperature of hottest month, OAK 31.4°C vs. WAR 29.4°C and KIL 28.4°C) and less rainfall during the coldest quarter (OAK 97.4 mm vs. WAR 108.8 mm and KIL 124.1 mm). Similarly, divergent responses to seasonal hydroclimatic variation are evident for M. fluviatilis from BOG and STG. BOG is much warmer in summer (35.6°C) and has lower rainfall (particularly during cooler periods; BOG average rainfall during the coldest quarter 64.0 mm vs. 98.3 mm for STG). Aside from the most divergent populations, a subtler seasonal climatic gradient in selection was also detected spanning sites with higher temperatures during the wettest months such as WAR, KIL, BEN and WAK (upper left quadrant; Figure 3b), transitioning to sites in the lower Murray and OAK where the temperature is much cooler during wet periods (lower right quadrant; Figure 3b). Interestingly, this pattern is particularly evident in the StreamTree-based RDA. Although extensive simulation and empirical work are required before any general conclusions may be drawn, our findings perhaps suggest that refining the models used to control for population structure may provide increased resolution and improved inferences of weaker multilocus GEA signals in complex spatial environments.

These results provide evidence supporting the generality of previous findings, suggesting that similar hydroclimatic variables shape patterns of adaptive variation for Australian freshwater fishes spanning a wide range of environments and life history strategies. For instance, golden perch (Macquaria ambigua) is a large-bodied, highly mobile species found across the MDB. A recent riverscape genomic study reported annual variation in stream flow and seasonal variation in rainfall were the most important environmental determinants of adaptive divergence for this species (Attard et al., 2018). Similarly, Harrisson et al. (2017) found seasonal variation in temperature and rainfall was likely driving a polygenic adaptive response in Murray cod (Maccullochella peeli), another large-bodied and long-lived species endemic to the MDB. On the other hand, southern pygmy perch (Nannoperca australis) is a small-bodied wetland specialist with very low capacity for dispersal. Findings of GEA analyses for this species again identified seasonal variation in rainfall as the predominate
environmental factor influencing adaptive divergence among populations (Brauer et al., 2016). Finally, Smith et al. (unpublished) found the largest number of candidate loci were linked to stream flow perenniality in their study of the subtropical rainbowfish, 

Melanotaenia duboulayi. Considered in context with these studies, our findings for 

M. fluviatilis add weight to a more general emerging paradigm where temporal variation, rather than long-term averages in hydroclimatic conditions appear the most salient agents of selection in Australian riverine ecosystems.

4.2 | Gene flow and connectivity in dendritic systems

Although IBD has been observed many times in natural populations, the strength of this relationship is often variable among, or even within species and IBD alone is often a poor predictor of spatial genetic patterns (Raeymaekers et al., 2008). One reason is that IBD models fail to capture the potential effects of landscape heterogeneity on patterns of dispersal and gene flow (Kalinowski et al., 2008). This may be especially the case for freshwater species where physical characteristics of river systems such as flow, slope and the dendritic arrangement of streams can influence dispersal independently of, and in addition to the effects of waterway distance (Castric, Bonney, & Bernatchez, 2001; Hébert, Danzman, Jones, & Bernatchez, 2000; Morrissey & de Kerckhove, 2009; Pru nier, Dubut, Loot, Tudesque, & Blanchet, 2017). Models of gene flow and connectivity that incorporate aspects of landscape heterogeneity may therefore improve understanding of spatial genetic structure in river systems (Meeuwig, Guy, Kalinowski, & Fredenberg, 2010). In the case of 

M. fluviatilis, IBD was not a good predictor of population differentiation, whereas in contrast the StreamTree model provided a strong fit with the spatial population structure across the MDB. This demonstrates that connectivity is influenced more strongly by characteristics of the stream network than by waterway distance. Specifically, stream sections connecting tributaries to the main river channel appear to contribute a disproportionally high amount to $F_{ST}$. This is predicted by the SHM (Meffe & Vrijenhoek, 1988) and suggests that connectivity among tributary populations in particular is limited by local characteristics of the stream network.

In addition to network configuration affecting spatial patterns of differentiation, the SHM also predicts reduced levels of genetic variation in tributaries relative to populations lower down in the stream network (Meffe & Vrijenhoek, 1988). The levels of diversity in the Condamine River (OAK, WAR and KIL) are higher than would be expected under the SHM. The reason for this is unclear; however, we hypothesize that the larger size of this tributary may have supported larger populations in the long-term, relative to other headwater streams. Nevertheless, our results support several other findings demonstrating that the physical structure of river systems can profoundly affect spatial patterns of genetic diversity, gene flow and metapopulation dynamics (Hébert et al., 2000; Morrissey & de Kerckhove, 2009; Paz-Vinas & Blanchet, 2015; Paz-Vinas, Loot, Stevens, & Blanchet, 2015; Thomaz et al., 2016). The stream sections that appear to most inhibit connectivity, as highlighted by the StreamTree model, are also located across the steepest hydroclimatic gradients present in the MDB. This suggests that in addition to the effects of dendritic network structure, environmental variation may also be restricting dispersal among some 

M. fluviatilis populations; a pattern that would be consistent with findings from studies of other freshwater fishes in the MDB (Faulks, Gilligan, & Beheregaray, 2010b, 2011; Lean et al., 2016).
4.3 Interactions among evolutionary processes

In contrast to patterns of neutral connectivity, our understanding of how adaptive genomic variation may be influenced by evolutionary processes other than landscape structure-mediated gene flow remains relatively limited. Findings here suggest that apart from landscape heterogeneity, there may be additional factors modulating the adaptive response to hydroclimatic selection for *M. fluviatilis*. For instance, genetic variation is much lower at headwater sites in the Murray River (WAK, BEN average 36.9% polymorphic loci) compared to those in the Darling River (PEL, GWY, DUM, MIB, STG average 55.3% polymorphic loci). The Murray River sites are at the southern limit of the distribution of not only *M. fluviatilis*, but for any member of the otherwise tropical or subtropical Melanotaeniidae (Unmack et al., 2013). Many examples exist of range margin populations exhibiting low genetic variation and associated reduced responses to selection (Bridle & Vines, 2007; Eckert, Samis, & Lougheed, 2008; Lenormand, 2002). Accordingly, several mechanisms have been proposed to explain these phenomena including gene flow from central-range populations swamping locally adapted alleles (Bridle & Vines, 2007) and potential phylogenetic constraints (Comte, Murienne, & Grenouillet, 2014). While specifically testing these hypotheses is beyond the scope of our study, findings that the genetic architecture of *M. fluviatilis* is consistent with the SHM suggest it is unlikely that there has been sufficient gene flow from maladapted central populations to cause reduced genetic variation in the Murray River. If, however, due to their tropical origins, the entire clade of Melanotaeniidae possesses limited genetic variation for traits associated with adaptation to temperate hydroclimatic conditions, it is possible that the reduced genetic variation observed here may be the result of deeper phylogenetic constraints leading to reduced fitness of *M. fluviatilis* in temperate environments. This hypothesis is further supported by the reproductive ecology of tropical rainbowfish species that are thought to reduce larval mortality by concentrating spawning effort during the more stable and benign conditions of the dry season (Pusey, Arthington, Bird, & Close, 2003). Perhaps surprisingly, the temperate *M. fluviatilis* similarly reproduce during the dry months (Humphries, Serafini, & King, 2002) despite the fact that conditions are far less predictable during the dry season in the lower MDB.

4.4 Riverscape genomics informing proactive conservation measures

The spatial distribution of genetic diversity in natural populations is shaped by a balance of evolutionary processes including gene flow among demes and natural selection in response to environmental variation within and among habitat patches. Dendritic riverscapes provide a particularly challenging environment for assessing the relative influence of these processes on spatial genetic structure and adaptive divergence of populations. By incorporating models of landscape heterogeneity with measures of environmental variation in a riverscape genomics analysis framework, it is possible to tease apart the genomic signals of each. Development of spatial statistical models that better represent the unique characteristics of dendritic river networks, however, is needed to further improve inferences in riverscape genomics studies. The StreamTree-based RDA presented in this study provides a novel and promising example incorporating the unique effects of dendritic network structure as well as restrictions to connectivity due to barriers and other potentially unknown sources of resistance. The ongoing evolution of methods to model GEAs in complex spatial environments has nevertheless already advanced our understanding of local adaptation of aquatic organisms. When combined with other genetic, demographic and environmental data, these studies provide a powerful predictive framework on which to base conservation and water management decisions. In the case of *M. fluviatilis*, adaptive divergence in response to hydroclimatic selection appears to be mediated by a combination of landscape heterogeneity, spatially variable patterns of dispersal and potentially, phylogenetic history. Translating this information into conservation management practice, however, is far from straightforward. On the one hand, anticipated warmer temperatures across the MDB (Davis et al., 2015; Kershaw, Moss, & Van Der Kaars, 2003; Morrongiello et al., 2011) could potentially benefit populations at the current southern range boundary by alleviating any genetic constraints on adaptation to temperate conditions. In contrast, predicted concurrent increases in environmental variability and unpredictability may simultaneously prove detrimental for these populations by increasing the frequency and severity of demographic fluctuations in response to extreme weather events. This highlights the difficulties faced in predicting evolutionary responses to changing environmental conditions (Webster et al., 2017). Despite these challenges, if we are to reverse the current global decline of freshwater biodiversity, proactive conservation management is needed to restore evolutionary processes across fragmented and degraded river basins (Brauer, Unmack, & Beheregaryar, 2017; Brauer et al., 2016). In this case, we argue that monitoring and, in some cases, management of populations should ideally occur before a species situation becomes critical. This will provide conservation practitioners with more options than may otherwise be available once a species has declined to the point they are formally considered threatened. In the context of this study, although *M. fluviatilis* are presently only considered threatened in the Murray River (DELWP 2018), widespread natural and anthropogenic disturbance is likely already impacting the species across the whole MDB and will continue to threaten populations in the future. We identified complex patterns of connectivity operating at a range of spatial scales and in response to several aspects of landscape heterogeneity and hydroclimatic variation. Water management practices that continue to degrade habitat and alter natural flow regimes will further disrupt metapopulation dynamics, leaving isolated populations more vulnerable to stochastic demographic decline. Additionally, the effects of human disturbance are likely already being compounded by the simultaneous and rapid changes in climate that will further threaten the persistence of many MDB species.
ACKNOWLEDGEMENTS

Collections were obtained under permits from various state fisheries agencies, and research is under Flinders University Animal Welfare Committee approval E342. We thank Minami Sasaki for laboratory assistance. We also thank Michael Hansen and two anonymous reviewers for their comments which improved the manuscript. Financial support was provided by the Australian Research Council via a Future Fellowship project to L. B. B. (FT130101068) and Discovery projects DP110101207 and DP150102903 to L. B. B. and L. B.

DATA ACCESSIBILITY

Raw demultiplexed sequences are available on NCBI SRA database (SRA accession: SRP151519). Reference sequences for the 17,503 loci, SNP genotypes, environmental data and a custom R script to replicate the RDA analysis can be accessed on Dryad: https://doi.org/10.5061/dryad.t2v8825.

AUTHOR CONTRIBUTIONS

The study was designed by L.B.B., L.B. and C.J.B. The data were analysed and generated by C.J.B. and S.S. with assistance from P.J.U., L.B. and L.B.B. The manuscript was written by C.J.B. and L.B.B. with input from S.S., P.J.U. and L.B.B. The manuscript was written by C.J.B.

ORCID

Luciano B. Beheregaray http://orcid.org/0000-0003-0944-3003

REFERENCES


**SUPPORTING INFORMATION**

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