River capture or ancestral polymorphism: an empirical genetic test in a freshwater fish using approximate Bayesian computation

MATEUS S. SOUZA[1,2], ANDRÉA T. THOMAZ[†,3,4,*] and NELSON J. R. FAGUNDES[1,2,5,*]

[1]*Postgraduate Program in Animal Biology, Institute of Biosciences, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil*
[2]*Laboratory of Medical Genetics and Evolution, Department of Genetics, Institute of Biosciences, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil*
[3]*Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada*
[4]*Department of Zoology, University of British Columbia, Vancouver, BC, Canada*
[5]*Postgraduate Program in Genetics and Molecular Biology, Institute of Biosciences, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil*

Received 17 June 2020; revised 8 August 2020; accepted for publication 14 August 2020

A headwater or river capture is a phenomenon commonly invoked to explain the absence of reciprocal monophyly of genetic lineages among isolated hydrographic basins in freshwater fish. Under the assumption of river capture, a secondary contact between populations previously isolated in different basins explains the observed genetic pattern. However, the absence of reciprocal monophyly could also arise under population isolation through the retention of ancestral of polymorphisms. Here, we applied an approximate Bayesian computation (ABC) framework for estimating the relative probability of scenarios with and without secondary contact. We used *Cnesterodon decemmaculatus* as a study model because of the multiple possible cases of river capture and the demographic parameters estimated in a previous mitochondrial DNA study that are useful for simulating scenarios to test both hypotheses using the ABC framework. Our results showed that, in general, mitochondrial DNA is useful for distinguishing between these alternative demographic scenarios with reasonable confidence, but in extreme cases (e.g. recent divergence or large population size) there is no power to discriminate between scenarios. Testing hypotheses of drainage rearrangement under a statistically rigorous framework is fundamental for understanding the evolution of freshwater fish fauna as a complement to, or in the absence of, geological evidence.

ADDITIONAL KEYWORDS: Bayesian inference – *Cnesterodon decemmaculatus* – mitochondrial DNA – model choice – statistical phylogeography.

INTRODUCTION

Obligatory freshwater fishes are organisms primarily unable to disperse actively among hydrographic basins because of physical and physiological limitations (Myers, 1938; Vari, 1988; Albert & Carvalho, 2011). However, despite these limitations, many species have broad distribution ranges across several river basins whose genetic lineages frequently lack reciprocal monophyly and can even share identical haplotypes (see Sousa *et al.*, 2008; Lee & Johnson, 2009; Schönhuth *et al.*, 2011, 2018; Bossu *et al.*, 2013; Xu *et al.*, 2014; Ramos-Fregonezi *et al.*, 2017; Eaton *et al.*, 2018).

The explanation for these incongruences among the limited dispersal capacity of the organisms, distributions in multiple basins and patterns of genetic relationship is associated with the rearrangement of drainages over geological time, which provided connections between previously isolated basins. A phenomenon commonly invoked in this context is the river or headwater capture (Bishop, 1995). River capture is characterized by the transference of a river

*Corresponding authors. E-mail: deatthomaz@gmail.com; nelson.fagundes@ufrgs.br
[†]Current address: Facultad de Ciencias Naturales, Universidad del Rosario, Bogotá DC, Colombia, 111221

(or a segment of a river) between basins caused by erosion and tectonic processes and, as a result, the biodiversity associated with the diverted river (i.e. species and genetic diversity) will then be present in the receiver basin. This has important evolutionary implications. First, genetic lineages in the diverted basin will appear paraphyletic with regard to (at least some) lineages in the receiver basin (Hubert *et al.*, 2007). Second, range expansion through river capture can be recognized by the genetic signature of a founder bottleneck in the population from the colonized basin, because it would be formed by a small sample from the population of the diverted river (Waters & Wallis, 2000; Machado *et al.*, 2018).

River captures can contribute both to a range expansion, allowing species to reach new basins, and to a secondary contact between populations of species previously present in neighbouring basins; ideally, both geological and biological evidence converge when inferring drainage rearrangement events (Waters *et al.*, 2001, 2006). However, geological evidence is not always available or is generally based on controversial morphological features (e.g. 'elbows of capture'; Bishop, 1995; Craw & Waters, 2007). In many cases, therefore, it is necessary to rely on biological data, especially species distribution and genetics, for supporting cases of drainage rearrangements.

Although secondary contact generated by river capture is a likely and frequently used explanation for lack of reciprocal monophyly of genetic lineages among neighbouring basins, it is not the only possibility. It is possible that populations from different basins remained isolated since their primary divergence but retained ancestral polymorphisms, whose persistence will depend on the time since the divergence event and their effective population sizes (Nielsen & Beaumont, 2009). In these circumstances, where two different evolutionary processes might lead to a unique genetic pattern, choosing one of these alternative demographic histories is not trivial. It requires the ability of genetic data to distinguish between historical models, which can only be achieved by properly taking into consideration uncertainties associated with the demographic parameters in the alternative models under scrutiny.

The use of model selection methods, such as approximate Bayesian computation (ABC), is a promising way to deal with this challenge (Sousa *et al.*, 2012). In an ABC framework, the evaluation of competitive models is performed through extensive simulations of the expected genetic variation under each model using demographic parameters sampled from a realistic distribution (i.e. priors). This is followed by comparisons of summary statistics between simulated and empirical data to evaluate which scenario produces simulations that best fit the observed data, thus allowing a direct estimate of the posterior probability (PP) of each competing model (Bertorelle *et al.*, 2010; Csilléry *et al.*, 2010; Beaumont, 2019).

A recent phylogeographical analysis (Ramos-Fregonezi *et al.*, 2017) suggested that the lack of monophyly between basins and the shared haplotypes between neighbouring basins observed in the mitochondrial DNA (mtDNA) of *Cnesterodon decemmaculatus* (Jenyns, 1842) (Cyprinodontiformes: Poeciliidae) constitute possible cases of secondary contact by river captures. This approach is a frequent practice in phylogeographical studies looking at aquatic organisms and the influence of river captures in genetic diversity, despite the lack of a formal statistical approach to validate these inferences (Sousa *et al.*, 2012). Here, we used the ABC approach to revisit the study of Ramos-Fregonezi *et al.* (2017) and perform model selection between demographic scenarios that are commonly associated with river captures [i.e. secondary contact (SECCON)] vs. retention of ancestral polymorphism [i.e. divergence and isolation (DIVISO)] for this species. Furthermore, we used *C. decemmaculatus* as a case study to evaluate the power of mtDNA data, widely used in freshwater fishes, in distinguishing between these two competing hypotheses.

MATERIAL AND METHODS

EMPIRICAL DATASET

Cnesterodon decemmaculatus is a species widely distributed along the Pampa biome, extending its range through Brazil, Uruguay and Argentina. It has recently been the subject of two phylogeographical studies (Bruno *et al.*, 2016; Ramos-Fregonezi *et al.*, 2017). In the latter study, individuals were sampled from three basins [Uruguay River Basin (Uruguay), Negro River Basin (Negro) and Patos-Merín Lagoon Basin (Merín)] and from a group of small basins in a coastal region in southern Uruguay (Southern) (Fig. 1A). Based on analyses of the mitochondrial gene NADH dehydrogenase 2 (*ND2*), it was found that there was no reciprocal monophyly between each of these sampled basins, and in several instances two neighbouring basins shared mtDNA haplotypes (Fig. 1B). The authors discussed four cases in which the observed genetic patterns could be explained by DIVISO or SECCON generated by recent river captures: Uruguay vs. Negro (U-N), Negro vs. Merín (N-M), Negro vs. Southern (N-S) and Uruguay vs. Southern (U-S). Although the authors acknowledged that it was difficult to distinguish between the alternatives with certainty, they concluded that the data provided better evidence for the secondary contact scenario for all cases, except U-S, which was explained as a probable scenario of incomplete lineage sorting

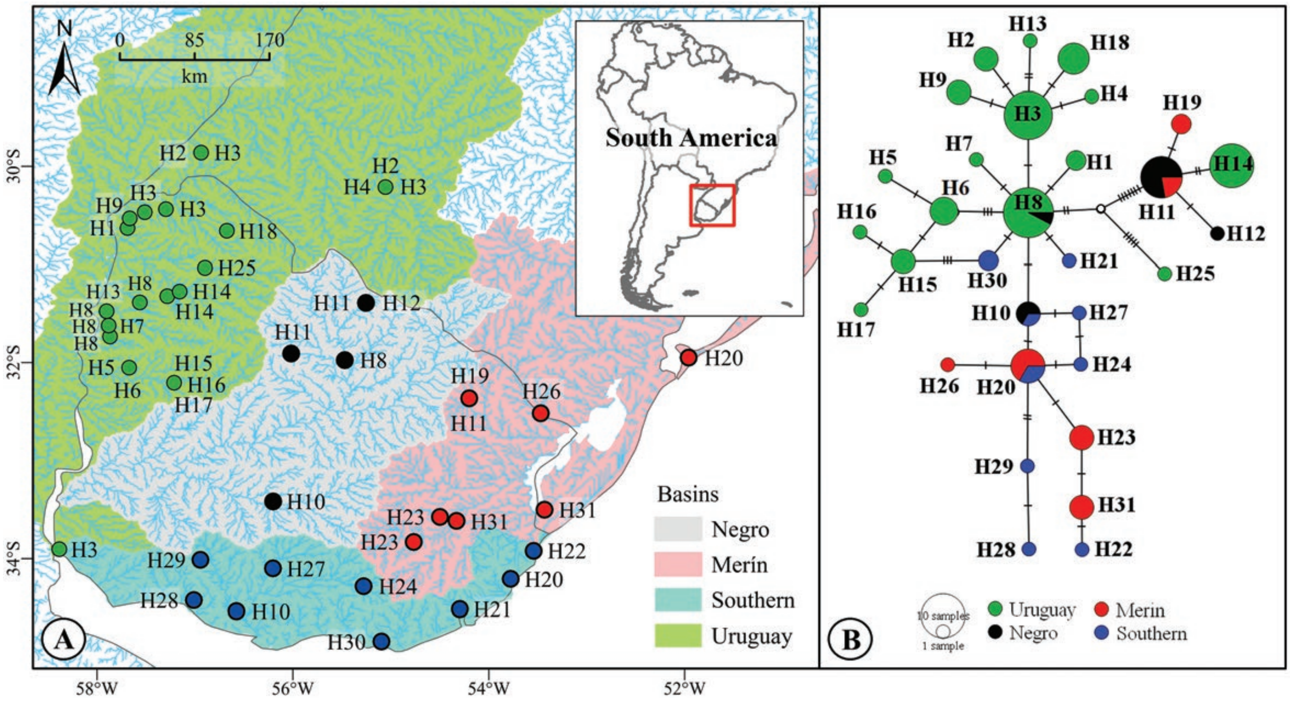

Figure 1. A, sampling sites for *Cnesterodon decemmaculatus*. The inset shows the location of the study area (highlighted in a red square) in South America. Grey lines indicate country borders, and areas in different colours indicate different basins (blue, Southern; coral, Merín; green, Uruguay; grey, Negro). Haplotypes found in each sampling site (dots) are indicated. B, median-joining network for *ND2* haplotypes. The size of the circles is proportional to haplotype frequency. Ticks on haplotype connections indicate the number of mutational steps. Original artwork based on the data from Ramos-Fregonezi *et al.* 2017.

(because of the closer relationship of haplotypes H21 and H30 to Uruguay than Southern populations; see Fig. 1). Ramos-Fregonezi *et al.* (2017) also provided DNA sequence data for the nuclear gene *Myh6*; however, as their discussion about river capture was entirely motivated by the patterns of genetic diversity presented by the mtDNA data, we did not use the *Myh6* sequences in our tests.

The dataset consisted of 99 sequences from the *ND2* gene, consisting of 62 samples from Uruguay, 11 from Negro, 15 from Merín and 11 from Southern (GenBank accession numbers: KU214332–KU214430). After downloading the sequences from GenBank, we inspected the alignment visually in MEGA X (Kumar *et al.*, 2018) and grouped them by sampling basins in DnaSP 6 (Rozas *et al.*, 2017). To enhance confidence in haplotype identity at each basin, we excluded from the dataset sample KU214333 because it contained gaps. From DnaSP 6, we exported the sequences in four ARLSUMSTAT (Excoffier *et al.*, 2010) format files, with each file containing grouped sequences from basin pairs corresponding to each of the four cases (see above). The file for U-N, for instance, contained the sequences from Uruguay and Negro. We used a median-joining network (Bandelt *et al.*, 1999) in the program PopART (Leigh & Bryant, 2015) to represent the evolutionary relationships among haplotypes.

Approximate Bayesian computation procedure

Summary statistics and simulated data

In short, the ABC procedure compares summary statistics derived from observed and simulated datasets. We used ARLSUMSTAT to calculate the observed summary statistics for each of the four cases. The choice of summary statistics is an important step in the ABC procedure (Csilléry *et al.*, 2010). Based on preliminary tests, we narrowed our analyses to eight summary statistics, which were the most informative to distinguish between scenarios: number of private polymorphic sites per population (2), number of haplotypes per population (2), Tajima's *D* for each population (2), total number of polymorphic sites over all populations (1) and genetic differentiation between populations measured by the fixation index F_{ST} (1). For all scenarios (see next subsection, '*Demographic scenarios and prior distributions*'), we simulated the evolution of a sample of *ND2*-like haplotypes with the same characteristics (e.g. sequence length, the sample size of each population) of the observed data and estimated, after each simulation, the same set of eight summary statistics described above.

Demographic scenarios and prior distributions

In the DIVISO scenario (Fig. 2A), an ancestral population originates two daughter populations *N* generations ago (*Td*), which remain isolated until the present. According to this scenario, any genetic similarity between them (e.g. sharing of lineages and haplotypes) is attributable to the retention of polymorphisms present in the ancestral population. In the SECCON scenario (Fig. 2B), after the split (*Td*) and a period of isolation between daughter populations, gene flow occurs from one population to the other (*Mig1*) during the period of one generation (*Tsc*). In this event, a proportion of the number of genetic lineages present in the source population can migrate to the other population, which, on top of ancestral polymorphism, might explain the genetic similarities between populations. Specifically, for the cases N-M (Negro–Merín), N-S (Negro–Southern) and U-S (Uruguay–Southern), we tested only one direction of migration (Fig. 2B), as indicated by the relationship between haplotypes from these basins (Fig. 1) and suggested by Ramos-Fregonezi *et al.* (2017). For the U-N (Uruguay–Negro) case, the distribution of haplotypes suggested the movement of migrants in both directions (i.e. from Uruguay to Negro and from Negro to Uruguay); therefore, we tested two additional scenarios for this case, which are variations of the SECCON scenario. In the third scenario of U-N (Fig. 2C), the source and the receiver populations of migrating lineages are inverted (SECCONinv). The fourth scenario (Fig. 2D) includes bidirectional gene flow, i.e. both populations give and receive migrants from the other population (*Mig1* and *Mig2*) at the same time point (SECCONrgf).

Prior distributions for demographic parameters, such as divergence time and effective population size, were based on the estimates from the empirical data

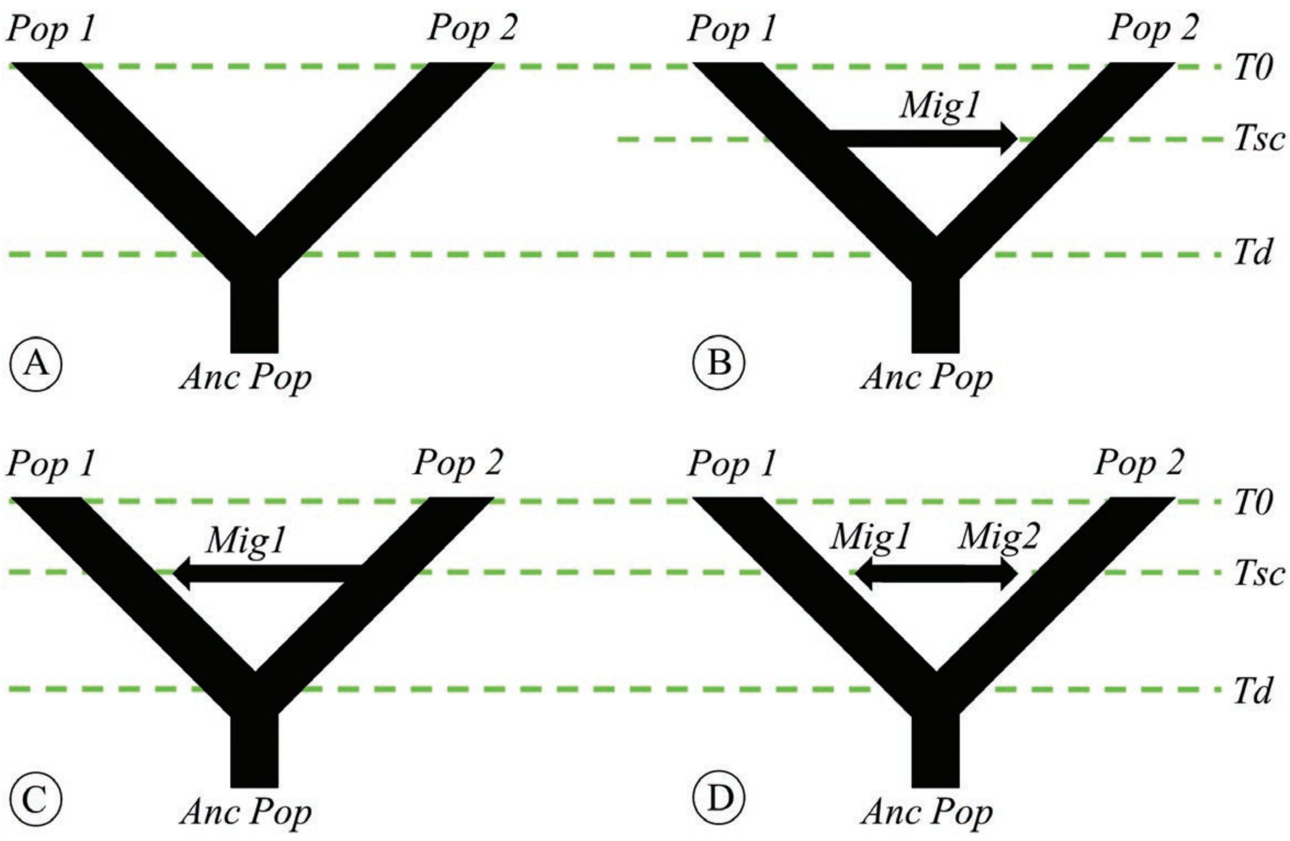

Figure 2. Coalescent demographic scenarios. A, DIVISO (divergence and isolation). B, SECCON (secondary contact). C, SECCONinv (secondary contact with inverted direction of gene flow). D, SECCONrgf (secondary contact with reciprocal gene flow). Abbreviations: *Anc Pop*, ancestral population; *Mig1* and *Mig2*, migration parameters; *Pop 1* and *Pop 2*, populations 1 and 2 (U-N: *Pop 1*, Uruguay and *Pop 2*, Negro; N-M: *Pop 1*, Merín and *Pop 2*, Negro; N-S: *Pop 1*, Negro and *Pop 2*, Southern; U-S: *Pop 1*, Southern and *Pop 2*, Uruguay); *T0*, present time; *Td*, time since divergence; *Tsc*, time since secondary contact. Migration parameters are represented backwards in time (e.g. *Mig1* in scenario B represents the proportion of genetic lineages in *Pop 1* that return to *Pop 2* in *Tsc*).

(Ramos-Fregonezi *et al.*, 2017) and are detailed in the Supporting Information (Table S1). Although the effective population sizes were estimated using a method that does not allow migration between populations, these estimates were used in SECCON scenarios because we did not expect secondary contact through river capture to have a significant impact on population sizes. We also assumed a generation time of 1 year and an mtDNA mutation rate with a normal distribution and mean of 8.6×10^{-9} substitutions per site per generation (Hrbek & Meyer, 2003), as in the study by Ramos-Fregonezi *et al.* (2017). For parameters that have not been estimated previously [i.e. the time since secondary contact (*Tsc*) and the proportion of migrant lineages (*Mig1* and *Mig2*); Supporting Information, Table S1], we used broad and uniform distributions for the priors. Note that our focus was not on refining the estimates for demographic parameters, but on the power to discriminate between DIVISO and SECCON hypotheses.

Simulations and model choice

The simulation procedure was performed in ABCtoolbox v.2.0 (Wegmann *et al.*, 2010) integrated with the coalescent simulator FASTSIMCOAL v.2.6 (Excoffier *et al.*, 2013) and ARLSUMSTAT (Excoffier *et al.*, 2010) for the computation of summary statistics. For each scenario, we performed 10^6 simulations.

All simulated and empirical datasets were analysed in R (R Core Team, 2019) using the *abc* package (Csilléry *et al.*, 2012). The PP of each scenario was estimated by applying the multinomial logistic regression approach (Fagundes *et al.*, 2007; Beaumont, 2008) based on a 1% threshold (i.e. retaining the 1% of simulations whose summary statistics were closest to the observed summary statistics). Following the approach proposed by Fagundes *et al.* (2007), the PP of the selected scenario obtained during cross-validation (see next subsection, '*Quality control*') were used to compute the probability that a given scenario is the correct one conditioned on the original PP estimate in the multinomial logistic regression step. The reasoning for this correction is that a high support for a given scenario might be misleading if this scenario has a high PP irrespective of the correct model (i.e. it is highly supported even if it is the wrong model). This idea is expressed in the following equation:

$$pr\,(SS \mid pSS = ppss) = \frac{pr\,(pSS = ppss \mid SS)}{pr\,(pSS = ppss \mid SS) + pr\,(pSS = ppss \mid AS1) + \ldots + pr\,(pSS = ppss \mid ASN)}$$

where *SS* is the selected scenario, *pSS* the probability of the selected scenario obtained during cross-validations, *ppss* the PP of the selected scenario estimated through multinomial logistic regression, *AS* the alternative or not selected scenario and *N* the number of alternative scenarios. Note that U-N has three alternative scenarios, whereas the other cases have only one.

Quality control

To evaluate the fit of scenarios to the empirical dataset, we calculated the goodness-of-fit D-statistic using the *gfit* function of the *abc* package (Csilléry *et al.*, 2012). The D-statistic is the mean distance between summary statistics from the observed dataset and summary statistics from the 1% of simulations nearest to the observed dataset. We calculated the null distribution of D using 100 pseudo-observed datasets (PODs) and the *P*-value of the goodness-of-fit test as the proportion of D from the PODs that are larger than the empirical D. High *P*-values indicate a good fit between simulated and observed data.

We also have carried out posterior predictive checks (Csilléry *et al.*, 2010) to evaluate the fit of each selected scenario. For this, 10^4 additional simulations were performed for each scenario with parameters sampled from the posterior distribution obtained through local linear regression (Beaumont *et al.*, 2002) based on a threshold of 1% (Supporting Information, Table S2). Summary statistics point estimates from the empirical datasets were plotted against the histogram of summary statistics from these simulations.

The quality of each model selection procedure was assessed by leave-one-out cross-validation. For each scenario of each case, 1000 PODs were randomly selected from the 1% of the simulations nearest to the empirical dataset. The estimate of the PP for each scenario followed the same procedure used for the empirical dataset (see above). By doing this, we were able to evaluate the rate of false-positive and false-negative inferences for each case and to calculate the average PP of the correct scenario in each case.

RESULTS

Scenarios including secondary contact (SECCON) presented the higher posterior probabilities for U-N, N-M and N-S. In contrast, the divergence with isolation model (DIVISO) was more probable for U-S (Table 1). The PP for the most likely scenario across the different cases ranged from > 0.95 (SECCON for N-M and N-S) to 0.525 (DIVISO for U-S), with an intermediate value of 0.626 (SECCONrgf for U-N). The goodness-of-fit tests based on the D-statistic showed a good fit (*P*-value > 0.05) to the empirical data for most of the scenarios, except DIVISO (N-M and N-S) and SECCON (N-S) (Supporting Information, Table S3). However, the posterior predictive checks showed that all selected (i.e. most likely) scenarios were able to

Table 1. Posterior probabilities of scenarios

Case	Posterior probability				Pr. selected scenario
	DIVISO	SECCON	SECCONinv	SECCONrgf	
U-N	0.002	0.067	0.365	**0.566**	0.626
N-M	0.001	**0.999**	–	–	0.953
N-S	0.219	**0.781**	–	–	0.955
U-S	**0.669**	0.331	–	–	0.525

Selected scenarios are highlighted in bold. 'Pr. selected scenario' indicates the probability that the selected scenario is the correct one given its posterior probability (PP) and the PP distributions obtained through cross-validations (see 'Simulations and model choice' in the 'Material and Methods' and Supporting Information, Fig. S2).
Abbreviations: DIVISO, Divergence and isolation; N-M, Negro vs. Merín; N-S, Negro vs. Southern; SECCON, Secondary contact; SECCONinv, Secondary contact with inverted direction of gene flow; SECCONrgf, Secondary contact with reciprocal gene flow; U-N, Uruguay vs. Negro; U-S, Uruguay vs. Southern.

reproduce the observed summary statistics, indicating a good fit between the most likely scenario and the empirical data (Supporting Information, Fig. S1).

Cross-validation checks based on 'pseudo-observed' summary statistics taken from a known scenario revealed an overall poor capacity for recovering it as the most likely scenario. For N-S, the overall error rate in the identification of the correct scenario was relatively low, at 29.9%. However, for U-N, N-M and U-S, the correct scenario was misidentified in 72.2, 43.4 and 48.8% of the tests, respectively (Supporting Information, Table S4). In N-M and U-S, the most probable scenario (i.e. the selected scenario in Table 1) was recovered with high frequency when it was either the correct scenario or not, although with an overall lower PP on average in the latter case. In U-N, SECCON was selected in most of the cross-validation tests, with a mean PP similar to the most probable scenario (SECCONrgf). Indeed, despite the poor performance of scenario selection indicated by cross-validation checks, the corrected PP estimate for the most likely scenario (conditioned on its PP during the model choice procedure) showed that, at least for cases N-M and N-S, there was a high confidence in scenario selection (Table 1; for a graphical explanation, see Supporting Information, Fig. S2).

In contrast, for U-S, the selected scenario had approximately the same PP as the alternative scenario (0.525 vs. 0.475, respectively), indicating no power to distinguish between them. For case U-N, the corrected PP of the best-favoured scenario (SECCONrgf) was 0.626. This is the only scenario including bidirectional migration, and we performed further tests to investigate whether there was power to discriminate between a single pulse of bidirectional migration vs. two temporally independent events of unidirectional migration (Supporting Information, Fig. S3). However, our results clearly showed that it was not possible to discriminate between these alternatives, which received almost equal support for this dataset (0.498 vs. 0.502, respectively; Supporting Information, Fig. S3). Overall, although it is possible to infer with confidence for U-N a scenario including secondary contact (vs. DIVISO), our data do not have enough power to discriminate among alternative scenarios of secondary contact.

DISCUSSION

In this study, we explored the use of an ABC framework for confronting two alternative demographic hypotheses that could generate the lack of reciprocal monophyly in genetic lineages between river basins; a challenge that arises frequently in phylogeographical studies of freshwater fish species. Based on genetic data from the literature and by comparing genetic diversity indices observed in empirical and simulated mtDNA datasets, we estimated the relative probability that the shared polymorphism between current isolated river basins had been originated by the maintenance of ancestral polymorphism [i.e. divergence and isolation (DIVISO)] or by recent river capture [i.e. secondary contact (SECCON)]. Our results show that it is possible to assign, with reasonable confidence in most of the cases evaluated, the genetic diversity observed in the mtDNA of *C. decemmaculatus* to one of these competing scenarios.

Besides *C. decemmaculatus*, other species of freshwater fish might have been affected by the river captures addressed in this study. The presence of four species of *Austrolebias* (Rivulidae; Loureiro *et al.*, 2011) and three species of *Brachyhypopomus* (Hypopomidae; Loureiro & Silva, 2006; Giora *et al.*, 2008; Giora & Malabarba, 2009; Serra *et al.*, 2014) in the Negro and the Patos-Merín basins is in agreement with the connection inferred in this study. It remains to be tested whether the same geological event allowed the dispersal of *Austrolebias* spp., *Brachyhypopomus* spp. and *C. decemmaculatus* between basins or whether river capture between these drainages has been a recurrent phenomenon. Also, elucidating the timing of the single or multiple connections and the extension of these events might explain the large number of endemics in the Patos-Merín system (Loureiro *et al.*, 2011) and provide insights about the directionality of the dispersal events that can help to infer geologically which basin is eroding and being captured. For example, in the case of *C. decemmaculatus* our results support the movement of migrants from Negro to Patos-Merín, in agreement with the general view that coastal drainages usually capture upland drainages (Ribeiro, 2006). However, from a spatial perspective, Loureiro *et al.* (2011) proposed two independent river capture sites that could explain the distribution of *Austrolebias* spp. in these drainages, possibly allowing the interchange of fauna in both directions owing to the opposite directionality of each capture. In addition, this unidirectional upland to coastal trend might not be absolute, because upland portions of coastal rivers (the Southern basin in our study) seem to have contributed to the Negro fish fauna, as suggested by our tests, and could be the case for other species present in both Southern and Negro, such as *Australoheros scitulus* (Říčan & Kullander, 2003) (Říčan & Kullander, 2008), *Gymnotus omarorum* Richer-de-Forges, Crampton & Albert, 2009 (Richer-de-Forges *et al.*, 2009) and *Corydoras paleatus* (Jenyns, 1842) (Tencatt *et al.*, 2016).

Interestingly, *Australoheros scitulus* is also distributed in the tributaries from the left bank of the Uruguay river, which is in agreement with the river connections between the Negro and Uruguay populations suggested by Ramos-Fregonezi *et al.* (2017) and corroborated in the present study. Unlike any other pair of drainages, Negro and Uruguay have a freshwater connection at present. However, the habitat preference of *C. decemmaculatus* is restricted to small rivers, and the distance between sampling sites in which closely related haplotypes occur makes gene flow along the current watercourse unlikely (Ramos-Fregonezi *et al.*, 2017), in agreement with the result from the ABC model choice procedure. *Gymnotus omarorum* and *Corydoras paleatus*, which have similar distributions to *C. decemmaculatus*, might be other particularly interesting model species to test the spatial and temporal coincidence of genetic populations and the river capture events associated with species dispersal across drainages in this region.

It is important to keep in mind that the ability to discriminate between alternative phylogeographical hypotheses depends crucially on the power of genetic data. In the case of the ABC procedure, this will depend on the ability to obtain different sets of summary statistics under the alternative models (Csilléry *et al.*, 2010; Beaumont, 2019). In turn, how different the set of expected summary statistics will be under each hypothesis depends on evolutionary and demographic features, such as effective population sizes and how long ago divergence and secondary contact occurred. A recent divergence between large populations, for example, is likely to be a very difficult test, because one would expect a large fraction of shared ancestral polymorphism irrespective of the occurrence of secondary contact. This might be related to the lack of power associated with case U-S, which has the most recent divergence time among all cases tested (Ramos-Fregonezi *et al.*, 2017; Supporting Information, Table S1). We were not able to discriminate between simultaneous (one-step) vs. sequential (two-step) events to account for the river capture between Uruguay and Negro (Supporting Information, Fig. S3). This was attributable to a very similar set of summary statistics produced by either model considering mtDNA alone. It is likely that improving genealogical information by the inclusion of a larger number of loci will be essential for performing model choice in these more difficult cases (Felsenstein 2006; Adrion *et al.*, 2014).

Other important caveats are related to the fact that population genetic models are simplifications of the real biological world and, as such, there should be a compromise between biological realism and statistical power (Bertorelle *et al.*, 2010). In our study, we made several simplifying assumptions. For example, we regarded samples from different rivers and regions within a hydrographic basin (or even from small isolated basins, as in Southern) as members of a single panmictic population. This is a somewhat unrealistic assumption, because connectivity between different locations of a basin is constrained by the structure of the river network itself (Hughes *et al.*, 2009; Thomaz *et al.*, 2016) and by the ecological preference of the species. We also set the period in which all gene flow between basins must occur to a single generation, assuming a short-term connection between neighbouring drainages during river captures, and allowed its intensity, the proportion of migrant lineages, to vary along a broad interval (*Mig1*/*Mig2* parameter: 0.001–0.999; Supporting Information, Table S1). In cases in which there is any evidence of prolonged periods of connection between drainages (such as basins prone to recurrent flooding connectivity), it will be interesting to increase the duration of the gene flow. Concerning intensity, a weak gene flow [i.e. migration parameter(s) sampled close to the inferior limit of the prior] is likely to reproduce a situation in which migrant lineages are lost or become very rare in the receiver population, erasing the signature of a secondary contact (Burridge *et al.*, 2006) and making a SECCON simulation similar to a DIVISO simulation. An intense gene flow [i.e. migration parameter(s) sampled around the superior limit], in contrast, is likely to result in migrant lineages replacing other lineages already present in the receiver basin or making them very rare. In this case, a SECCON simulation might also be similar to a DIVISO simulation, but the secondary contact will have the genetic signature of a divergence event. Although these difficulties appear at the extremes of our prior distribution for gene flow intensity, the posterior distribution of migration parameter in cases in which SECCON scenarios were selected (U-N, N-M and N-S) shows that simulations that best fit the empirical data are most intensely sampled around intermediate values of the prior distribution (Supporting Information, Table S2). In this context, tests for assessing the fit of selected scenarios to empirical data are fundamental to show that the most favoured scenarios were able to reproduce the set of summary statistics observed from real data.

Likewise, we performed cross-validation tests for assessing the performance of the model selection procedure, in which PODs from a known scenario replaced the observed data in the model selection procedure. We found an overall elevated error rate during cross-validation, which could undermine the confidence in model selection (Table 1). However, in three out of four cases, and after conditioning on the PP obtained during cross-validation (Fagundes *et al.*, 2007), we were able to select one scenario with relatively high confidence. On the contrary, in the U-S case, our results suggest that even when a secondary contact happened as the result of a river capture, mtDNA would not have the power to discriminate between alternative scenarios with confidence (Table 1). Thus, using *C. decemmaculatus* as a model species, our results provided statistical support for the inferences made by Ramos-Fregonezi *et al.* (2017) about past connections between Negro and its adjacent basins, although the relationship between Southern drainages and the Uruguay basin could not be solved with the present data. In summary, we showed that mtDNA data might be a useful genetic marker for testing hypotheses of river capture in freshwater fishes under a statistically rigorous framework.

ACKNOWLEDGEMENTS

We thank the National Council for Scientific and Technological Development [Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)] for a master's scholarship awarded to M.S.S (Proc. 132447/2018-5). A.T.T. is a BRITE (Biodiversity Research: Integrative Training and Education) postdoctoral fellow from the Biodiversity Research Centre at the University of British Columbia. We thank Luiz Roberto Malabarba and Filipe Michels Bianchi for their helpful comments during the elaboration and execution of this study. We also thank Marcelo Coelho Miguel Gehara and two anonymous reviewers for their helpful comments on the manuscript.

REFERENCES

Adrion JR, **Kousathanas A**, **Pascual M**, **Burrack HJ**, **Haddad NM**, **Bergland AO**, **Machado H**, **Sackton TB**, **Schlenke TA**, **Watada M**, **Wegmann D**, **Singh ND. 2014.** *Drosophila suzukii*: the genetic footprint of a recent, worldwide invasion. *Molecular Biology and Evolution* **31:** 3148–3163.

Albert JS, **Carvalho TP. 2011.** Neogene assembly of modern faunas. In: Albert JS, Reis RE, eds. *Historical biogeography of Neotropical freshwater fishes*. Berkeley and Los Angeles: University of California Press, 118–136.

Bandelt HJ, **Forster P**, **Rohl A. 1999.** Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16:** 37–48.

Beaumont MA. 2008. Joint determination of topology, divergence time and immigration in population trees. In: Matsumura S, Forster P, Renfrew C, eds. *Simulation, genetics, and human prehistory*. Cambridge: McDonald Institute for Archaeological Research, 135–154.

Beaumont MA. 2019. Approximate Bayesian computation. *Annual Review of Statistics and Its Application* **6:** 379–403.

Beaumont MA, **Zhang W**, **Balding DJ. 2002.** Approximate Bayesian computation in population genetics. *Genetics* **162:** 2025–2035.

Bertorelle G, **Benazzo A**, **Mona S. 2010.** ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology* **19:** 2609–2625.

Bishop P. 1995. Drainage rearrangement by river capture, beheading and diversion. *Progress in Physical Geography* **19:** 449–473.

Bossu CM, **Beaulieu JM**, **Ceas PA**, **Near TJ. 2013.** Explicit tests of palaeodrainage connections of southeastern North America and the historical biogeography of Orange throat Darters (Percidae: *Etheostoma: Ceasia*). *Molecular Ecology* **22:** 5397–5417.

Bruno MC, **Mapelli FJ**, **Casciotta JR**, **Almirón AE**, **Lizarralde MS. 2016.** Phylogeography of *Cnesterodon decemmaculatus* (Cyprinodontiformes: Poeciilidae) in Southern Pampas, Argentina: ancient versus recent patterns in freshwater fishes. *Environmental Biology of Fishes* **99:** 293–307.

Burridge CP, **Craw D**, **Waters JM. 2006.** River capture, range expansion, and cladogenesis: the genetic signature of freshwater vicariance. *Evolution* **60:** 1038–1049.

Craw D, **Waters J. 2007.** Geological and biological evidence for regional drainage reversal during lateral tectonic transport, Marlborough, New Zealand. *Journal of the Geological Society* **164:** 785–793.

Csilléry K, **Blum MGB**, **Gaggiotti OE**, **François A. 2010.** Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* **25:** 410–418.

Csilléry K, **François O**, **Blum MGB. 2012.** abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution* **3:** 475–479.

Eaton KR, **Loxterman JL**, **Keeley ER. 2018.** Connections and containers: using genetic data to understand how watershed evolution and human activities influence cutthroat trout biogeography. *PLoS One* **13:** e0202043.

Excoffier L, **Dupanloup I**, **Huerta-Sánchez E**, **Sousa VC**, **Foll M. 2013.** Robust demographic inference from genomic and SNP data. *PLoS Genetics* **9:** e1003905.

Excoffier L, **Lischer HEL. 2010.** Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10:** 564–567.

Fagundes NJR, **Ray N**, **Beaumont M**, **Neuenschwander S**, **Salzano FM**, **Bonatto SL**, **Excoffier L. 2007.** Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 17614–17619.

Felsenstein J. 2006. Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution* **23:** 691–700.

Giora J, **Malabarba LR. 2009.** *Brachyhypopomus gauderio*, new species, a new example of underestimated species diversity of electric fishes in the southern South America (Gymnotiformes: Hypopomidae). *Zootaxa* **2093:** 60–68.

Giora J, **Malabarba LR**, **Crampton W. 2008.** *Brachyhypopomus draco*, a new sexually dimorphic species of Neotropical electric fish from southern South America (Gymnotiformes: Hypopomidae). *Neotropical Ichthyology* **6:** 159–168.

Hrbek T, **Meyer A. 2003.** Closing of the Tethys Sea and the phylogeny of Eurasian killifishes (Cyprinodontiformes: Cyprinodontidae). *Journal of Evolutionary Biology* **16:** 17–36.

Hubert N, **Duponchelle F**, **Nuñez J**, **Garcia-Davila C**, **Paugy D**, **Renno, J. 2007.** Phylogeography of the piranha genera *Serrasalmus* and *Pygocentrus*: implications for the diversification of the Neotropical ichthyofauna. *Molecular Ecology* **16:** 2115–2136.

Hughes JM, **Schmidt DJ**, **Finn DS. 2009.** Genes in streams: using DNA to understand the movement of freshwater fauna and their riverine habitat. *BioScience* **59:** 573–583.

Kumar S, **Stecher G**, **Li M**, **Knyaz C**, **Tamura K. 2018.** MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35:** 1547–1549.

Lee JB, **Johnson JB. 2009.** Biogeography of the livebearing fish *Poecilia gillii* in Costa Rica: are phylogeographical breaks congruent with fish community boundaries? *Molecular Ecology* **18:** 4088–4101.

Leigh JW, **Bryant D. 2015.** PopART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6**:1110–1116.

Loureiro M, **Duarte A**, **Zarucki M. 2011.** A new species of *Austrolebias* Costa (Cyprinodontiformes: Rivulidae) from northeastern Uruguay, with comments on distribution patterns. *Neotropical Ichthyology* **9:** 335–342.

Loureiro M, **Silva A. 2006.** A new species of *Brachyhypopomus* (Gymnotiformes, Hypopomidae) from Northeast Uruguay. *Copeia* **2006:** 665–673.

Machado CB, **Galleti PM Jr**, **Carnaval AC. 2018.** Bayesian analyses detect a history of both vicariance and geodispersal in Neotropical freshwater fishes. *Journal of Biogeography* **45:** 1313–1325.

Myers GS. 1938. Fresh water fishes and west Indian Zoogeography. *Annual Report of the Board of Regents of the Smithsonian Institution* **92:** 339–364.

Nielsen R, **Beaumont MA. 2009.** Statistical inferences in phylogeography. *Molecular Ecology* **18:** 1034–1047.

R Core Team. **2019.** *R: a language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org/

Ramos-Fregonezi AMC, **Malabarba LR**, **Fagundes NJR. 2017.** Population genetic structure of *Cnesterodon decemmaculatus* (Poeciliidae): a freshwater look at the Pampa Biome in southern South America. *Frontiers in Genetics* **8:** 214.

Ribeiro AC. 2006. Tectonic history and the biogeography of the freshwater fishes from the coastal drainages of eastern Brazil: an example of faunal evolution associated with a divergent continental margin. *Neotropical Ichthyology* **4:** 225–246.

Říčan O, **Kullander SO. 2008.** The *Australoheros* (Teleostei: Cichlidae) species of the Uruguay and Paraná River drainages. *Zootaxa* **1724:** 1–51.

Richer-de-Forges MM, **Crampton WGR**, **Albert JS. 2009.** A new species of *Gymnotus* (Gymnotiformes, Gymnotidae) from Uruguay: description of a model species in neurophysiological research. *Copeia* **2009:** 538–544.

Rozas J, **Ferrer-Mata A**, **Sánchez-DelBarrio JC**, **Guirao-Rico S**, **Librado P**, **Ramos-Onsins SE**, **Sánchez-Gracia A. 2017.** DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* **34:** 3299–3302.

Schönhuth S, **Blum MJ**, **Lozano-Vilano L**, **Neely DA**, **Varela-Romero A**, **Espinosa H**, **Perdices A**, **Mayden RL. 2011.** Inter-basin exchange and repeated headwater capture across the Sierra Madre occidental inferred from the phylogeography of Mexican stonerollers. *Journal of Biogeography* **38:** 1406–1421.

Schönhuth S, **Gagne RB**, **Alda F**, **Neely DA**, **Mayden RL**, **Blum MJ. 2018.** Phylogeography of the widespread creek chub *Semotilus atromaculatus* (Cypriniformes: Leuciscidae). *Journal of Fish Biology* **93:** 778–791.

Serra S, **Bessonart J**, **Mello FT**, **Duarte A**, **Malabarba LR**, **Loureiro M. 2014.** *Peces del Río Negro*. Montevideo: Ministerio de Agricultura, Ganadería y Pesca, Dirección Nacional de Recursos Acuáticos. AQ21AQ16

Sousa VC, **Beaumont MA**, **Fernandes P**, **Coelho MM**, **Chikhi L. 2012.** Population divergence with or without admixture: selecting models using an ABC approach. *Heredity* **108:** 521–530.

Sousa V, **Penha F**, **Collares-Pereira MJ**, **Chikhi L**, **Coelho MM. 2008.** Genetic structure and signature of population decrease in the critically endangered freshwater cyprinid *Chondrostoma lusitanicum*. *Conservation Genetics* **9:** 791–805.

Tencatt LFC, **Britto MR**, **Pavanelli CS. 2016.** Revisionary study of the armored catfish *Corydoras paleatus* (Jenyns, 1842) (Siluriformes: Callichthyidae) over 180 years after its discovery by Darwin, with description of a new species. *Neotropical Ichthyology* **14:** 1–20.

Thomaz AT, **Christie MR**, **Knowles LL. 2016.** The architecture of river networks can drive the evolutionary dynamics of aquatic populations. *Evolution* **70:** 731–739.

Vari RP. 1988. The Curimatidae, a lowland neotropical fish family (Pisces: Characiformes); distribution, endemism and phylogenetic biogeography. In: Vanzolini PE, Heyer WR, eds. *Proceedings of a workshop on neotropical distribution patterns*. Rio de Janeiro: Academia Brasileira de Ciências, 343–377.

Waters JM, **Allibone RM**, **Wallis GP. 2006.** Geological subsidence, river capture, and cladogenesis of galaxiid fish lineages in central New Zealand. *Biological Journal of the Linnean Society* **88:** 367–376.

Waters JM, **Craw D**, **Youngson JH**, **Wallis GP. 2001.** Genes meet geology: fish phylogeographic pattern reflects ancient, rather than modern, drainage connections. *Evolution* **55:** 1844–1851.

Waters JM, **Wallis GP. 2000.** Across the Southern Alps by river capture? Freshwater fish phylogeography in South Island, New Zealand. *Molecular Ecology* **9:** 1577–1582.

Wegmann D, **Leuenberger C**, **Neuenschwander S**, **Excoffier L. 2010.** ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* **11:** 116.

Xu W, **Yin W**, **Chen A**, **Li J**, **Lei G**, **Fu C. 2014.** Phylogeographical analysis of a cold-temperate freshwater fish, the Amur sleeper (*Perccottus glenii*) in the Amur and Liaohe River Basins of northeast Asia. *Zoological Science* **31:** 671–679.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Posterior predictive checks.
Figure S2. Probability that the selected scenario is the correct one.
Figure S3. Additional scenario selection for Uruguay vs. Negro (U-N).
Table S1. Prior distribution for demographic parameters.
Table S2. Distribution of demographic parameters used for simulations aimed to perform the posterior predictive checks.
Table S3. Goodness-of-fit D-statistic for each scenario based on 100 pseudo-observed datasets (PODs) and the empirical data.
Table S4. Cross-validation of scenario selection based on 1000 pseudo-observed datasets (PODs) taken from the 1% of simulations nearest to the observed summary statistics.