Inter-basin dispersal through irrigation canals explains low genetic structure in Diplomystes cf. chilensis, an endangered freshwater catfish from Central Chile

C.P. Muñoz-Ramírez a,*, P.F. Victoriano b, c, E. Habit c, d

a Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA
b Departamento de Zoología, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Casilla 160-C, Concepción, Chile
c Centro de Investigación en Ecosistemas Patagónicos (CIEP), Coyhaique, Chile
d Facultad de Ciencias Ambientales y Centro EULA-Chile, Universidad de Concepción, Chile

A R T I C L E   I N F O

Article history:
Received 4 December 2014
Received in revised form 22 April 2015
Accepted 23 April 2015
Available online 19 May 2015

Keywords:
Biotic homogenization
Biodiversity hotspot
Approximate Bayesian computation
mtDNA
Aquatic biota

A B S T R A C T

Biotic homogenization in freshwater ecosystems is a growing concern among conservation biologists. Recent phylogeographic data has shown low genetic structure between some basins from Central Chile, suggesting that either current dispersal through irrigation canals or incomplete lineage sorting due to recent divergence might explain the observed patterns. However, these hypotheses remain untested despite their potential implications for freshwater biodiversity and conservation. We used a statistical, model-based framework (approximate Bayesian computation) to investigate the relative support for each of these hypotheses in the freshwater catfish Diplomystes cf. camposensis, an endangered species from Central Chile. Our results show strong support for the model involving current migration between basins, and rejected the model of recent divergence without migration. These results strongly suggest that irrigation canals are facilitating the dispersal between basins, posing a serious threat to biodiversity in Central Chile, an area considered a biodiversity hotspot. Finally, these results highlight the utility of model-based approaches for determining demographic processes with potential conservation implications, even with the lack of extensive molecular data.

© 2015 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Biotic homogenization is a serious threat to the integrity of endemic species and genetic pools that can result in the loss of continental, regional, and local biodiversity (Olden et al., 2004). Human constructions such as the Suez Canal and the opening of the St. Lawrence Seaway are striking examples of human activities that have facilitated the movement of hundreds of estuarine and freshwater species (Galil, 2000; Ricciardi and Maclsaac, 2000), although better known examples are provided by the introduction of invasive species at a continental or worldwide scale (Olden, 2006). These events have shown varied negative effects on the diversity of the regions impacted, including reduction of β-diversity, species extinctions, and genetic homogenization (Olden et al., 2004).

At a smaller scale, a less studied human activity that has the potential to reduce diversity in freshwater environments is the artificial connection of historically isolated river drainages. A few cases of drainage interconnections facilitating inter-basin fish dispersal have been reported in the literature. For instance, the Snowy and Murray Rivers in Australia were artificially connected as the result of a major hydroelectric project during the early 20th century, facilitating the migration of the fish species Galaxias brevipinnis, an aggressive migratory species, into the Murray basin where the species was previously unrecorded (Waters et al., 2002). Similarly in New Zealand, Esa et al. (2000) reported that the connection of Otago’s two major river systems, the Taieri and the Cluthathie, by tunnels from mining activities resulted in the ingressive hybridization of two previously allopatric galaxiid species. Although just a few studies have reported freshwater fish migration between different river basins through human mediated connections, they highlight the need for documenting these phenomena and studying its consequences for regional and local biodiversity.

The area of Central Chile concentrates the highest human population density of the country, as well as a strong agricultural activity due to optimal climatic conditions (Fuentes and Hajek, 1979). The area is drained by small and parallel river basins that originate in the Andes and discharge into the Pacific Ocean, after running just a few hundreds of kilometers. Due to the constant demand for water from
increasing agricultural activity, a number of projects have altered the natural hydrological landscape of this area by expanding the irrigation network. Some of these alterations have resulted in the current connection of previously isolated drainages, such as the Teno–Chimbarongo Canal in Central Chile that links the Rapel and Mataquito basins, and the Zahartu Canal connecting the Itata and Biobío basins. Despite the potential threat to the biodiversity of the region, there is a complete lack of studies addressing this issue, and no reports of inter-basin dispersal through these canals have been reported up to date. However, there is some evidence suggesting that dispersal through these uncovered, stream-like canals might be possible. Along with the documented presence of fish in some of these canals, which may be using them as an extension of their natural habitats (Habit and Parra, 2001; Habit, 1994), a recent phylogeographic study on a catfish genus has found low genetic structure and shared haplotypes between these basins (Muñoz-Ramírez et al., 2014). This study, which analyzed patterns of genetic variation from sequences of mitochondrial DNA, showed that even though a high amount of genetic diversity was partitioned across river basins with unique lineages in most drainages across the distribution of this group (Muñoz-Ramírez et al., 2014, some basins in Central Chile exhibited low inter-basin genetic structure, particularly between the Rapel and Mataquito basins (Fig. 1).

Two hypotheses were suggested in Muñoz-Ramírez et al. (2014) to explain the lack of genetic structure in these basins. One points out that current migrations between these basins might be occurring through irrigation canals (e.g. the Teno–Chimbarongo Canal, Fig. 1), reducing genetic structure by genetic homogenization. The second suggested that the lack of genetic differentiation could be explained by incomplete lineage sorting resulting from a relatively recent divergence following the last Pleistocene glaciation. Opportunities for headwater rearrangements and river captures might have occurred when glaciers melted approximately 10,000 years ago (Ruzzante et al., 2006).

The consequences of potential connections between historically isolated basins might bring unforeseen impacts to the diversity of this region, which is considered an important hotspot of biodiversity (Myers et al., 2000). Thus, it is urgent to know whether these basins are exchanging fauna. However, testing these hypotheses with molecular data requires the use of methods that rigorously assess the relative support of these alternative scenarios and not simply rely on ad hoc explanations of the molecular patterns (Knowles, 2003). Current model-based approaches have proven to be powerful tools to test alternative demographic models, allowing the assessment of competing historical scenarios and the estimation of important demographic parameters (Bertorelle et al., 2010; Csilléry et al., 2010; Fagundes et al., 2007). In this study, we use a model-based simulation approach (approximate Bayesian computation) that allows the probabilistic comparison of alternative demographic models and the simultaneous estimation of demographic parameters to test whether current dispersal through irrigation canals or incomplete lineage sorting resulting from recent divergence better explain the observed genetic patterns in an endangered catfish, Diplomystes cf. chilensis.

Material and methods

Study system

Central Chile is considered a hotspot of biodiversity (Myers et al., 2000). This area is also one of the most threatened in Chile as it concentrates high levels of urbanization, forestry, and agriculture (Aronson et al., 1998; Azócar et al., 2007). One of the rarest, but also interesting group of fish inhabiting these rivers are the diplomystids, a group of endangered catfish recognized as one of the earliest branching lineages in the diverse order Siluriformes (Arratia, 1987; Sullivan et al., 2006). Diplomystids species are rare, have low fecundity (Vila et al., 1996), and exhibit generally low genetic diversity compared with other co-distributed fish species (Muñoz-Ramírez et al., 2014; Victoriano et al., 2012). All Chilean species are considered Endangered (CONAMA 2008). Their geographic distribution has decreased in recent times as they have gone extinct from two basins in the north part of their distribution (Aconcagua and Maipo basins; Muñoz-Ramírez et al., 2010), presumably as a consequence of habitat deterioration and the introduction of exotic species (Arratia, 1987). In this regard, irrigation canals may pose
a new type of threat for their survival and a challenge for their conservation which needs to be investigated.

**Haplotype network**

We reconstruct a haplotype network focusing on the northern distribution of the genus to highlight the shallow relationships between lineages from different basins in this region (Fig. 2). This was done by using the same sequences (GenBank’s accession numbers JX648663–JX649064) and following the same methodology as in Muñoz-Ramírez et al. (2014), using the software Network 4.610 (Bandelt et al., 1999) to build the haplotype network, and DnaSP 5.0 (Librado and Rozas, 2009) to infer the haplotypes and their frequencies.

**Simulation analyses**

The approximate Bayesian computation (ABC) approach can be briefly described as follows (for more details about this statistical technique see the reviews of Bertorelle et al., 2010, Cissléry et al., 2010, and Sunnäker et al., 2013). A large number of genealogies are simulated assuming different combinations of parameters values and demographic models. Simulations that produce genetic patterns (reduced to summary statistics) close to the observed patterns are retained for posterior analyses. Posterior probabilities of the models are estimated by the relative proportion of simulations retained by each model, while parameter values are estimated based on the selected model and the posterior distribution of its parameters. We used an ABC approach as implemented in the ABCtoolbox (Wegmann et al., 2010), using SIMCOAL 2.0 (Laval and Excoffier, 2004) to simulate the data, and a command line version of the program Arlequin 3.5 (arfsamstat) (Excoffier and Lischer, 2010) to calculate the summary statistics of each simulation. We focused on two river basins to build our models. The Rapel and the Mataquito Basins, because they (i) represent a monophyletic sample of individuals (see Fig. 2), (ii) exhibit the lowest levels of between-basin genetic structure, and (iii) allow reducing the complexity of the models being compared (see the schematic representation of the models in Fig. 3). The first model (Fig. 3A) represents the historical scenario of divergence after the last glaciation with no admixture (hereafter the “Model I”). The second model (Fig. 3B) represents the scenario of current inter-basin dispersal through artificial canals (hereafter Model II). These two models are similar, except in that in the Model II, basins are allowed to interchange individuals in recent times (since 100 years ago). Model parameters are fully described in Table 1. All model parameters being sampled were drawn from a uniform distribution to allow an unbiased exploration of the parameter space. Regardless of the prior distribution chosen, posterior distribution can usually be approximated from the retained simulations with confidence, assuming the model being tested is a fair representation of the process under study. The parameter T_DIV (i.e. time since divergence between basins) was sampled in the range of 2000–5000 generations (4000–10,000 ybp). This range seems reasonable because it is old enough to capture the end of the last glaciation (assuming a generation time of 2 years), but young enough to allow capturing potential connective events of a more recent age (up to 4000 ybp). The effective population size (i.e. N_e, the number of haploid genes in a deme) for the Rapel basin (N_e1) was set equal to twice the population size of the Mataquito basin (N_e2) to take into account the differences in size of the basins. The Mataquito basin drains an area of ~6000 km² (Source: http://www.sinia.cl/1292/articles-31018_Mataquito.pdf), whereas the Rapel basin drains an area of ~13,000 km² (Source: http://www.sinia.cl/1292/articles-31018_Rapel.pdf). Population sizes were sampled in the range of 10–500 individuals in every simulation. This range is reasonable given the small size of both basins and the biological characteristics of the species. Population sizes in the range of 500–2500 have been reported for a catfish from the Murray–Darling basin in Australia (Rourke and Gilligan, 2010) inhabiting a basin with an area of 1,000,000 km², two orders of magnitude larger than the basins studied here. Furthermore, diplomystids are known to be rare and of low fecundity (Habit et al., 2006; Vila et al., 1996) which suggests small population sizes. The mutation rate parameter was sampled in the range of 0.00002–0.00008 substitutions per locus per generation and, a generation time of 2 years was assumed based on literature (Vila et al., 1996). Regarding sequence evolution of the simulated sequences, a transition/transversion ratio was set to 0.33. A total of 200,000 data sets were simulated per model with the same characteristics of the empirical data set used in Muñoz-Ramírez et al. (2014), which consisted of 36 mtDNA sequences of 2073 base pairs in length (13 and 23 sequences from the Rapel basin and the Mataquito basin, respectively; GenBank’s accession numbers JX648663–JX648698 and JX648664–JX648899). The Tajima’s D neutrality index (a commonly used statistic to infer demographic expansion), π (the average number of pairwise differences), and S (the number of polymorphic sites) were used to summarize intra-basin genetic patterns, whereas FST (a measure of the degree of population subdivision) was used to characterize the genetic structure between the basins. These statistics were chosen because they are usually highly informative about pattern of genetic diversity at the population level. Only those simulations
Table 1
Model parameter settings and their prior distributions.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Distribution</th>
<th>Range sampled</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_1$</td>
<td>Effective population size of Mataquito Basin</td>
<td>Uniform</td>
<td>10–250</td>
<td>Units in number of individuals</td>
</tr>
<tr>
<td>$N_2$</td>
<td>Effective population size of Rapel Basin</td>
<td>Uniform</td>
<td>20–500</td>
<td>This parameter was set as a complex parameter in the .est file, $N_2 = N_1 \times 2$</td>
</tr>
<tr>
<td>$T_{DIV}$</td>
<td>Time since basins became isolated</td>
<td>Uniform</td>
<td>2000–5000</td>
<td>Units in number of generations assuming generation time = 2 years</td>
</tr>
<tr>
<td>MTDNA_RATE</td>
<td>Substitution rate for the mtDNA</td>
<td>Uniform</td>
<td>0.00002–0.00006</td>
<td>Values are expressed in substitutions per locus per generation</td>
</tr>
<tr>
<td>Population growth rate</td>
<td>Fixed</td>
<td>0</td>
<td>Population sizes were set to be constant over time</td>
<td></td>
</tr>
<tr>
<td>Ancestral population size</td>
<td>Fixed</td>
<td>1.5</td>
<td>The population size before divergence was set to be 1.5 times $N_e$</td>
<td></td>
</tr>
<tr>
<td>$N_1$</td>
<td>Effective population size of Mataquito Basin</td>
<td>Uniform</td>
<td>10–250</td>
<td>Values expressed in number of individuals</td>
</tr>
<tr>
<td>$N_2$</td>
<td>Effective population size of Rapel Basin</td>
<td>Uniform</td>
<td>20–500</td>
<td>This parameter was set as a complex parameter in the .est file, $N_2 = N_1 \times 2$</td>
</tr>
<tr>
<td>$T_{DIV}$</td>
<td>Time since basins became isolated</td>
<td>Uniform</td>
<td>2000–5000</td>
<td>Values are expressed in number of generations assuming generation time = 2 years</td>
</tr>
<tr>
<td>MTDNA_RATE</td>
<td>Substitution rate for the mtDNA</td>
<td>Uniform</td>
<td>0.00002–0.00006</td>
<td>Values are expressed in substitutions per locus per generation</td>
</tr>
<tr>
<td>$M$</td>
<td>Rate at which individuals migrate between basins</td>
<td>Fixed</td>
<td>0.1</td>
<td>10% of the population per generation (every 2 years)</td>
</tr>
<tr>
<td>Population growth rate</td>
<td>Fixed</td>
<td>0</td>
<td>Population sizes were set to be constant over time</td>
<td></td>
</tr>
<tr>
<td>Ancestral population size</td>
<td>Fixed</td>
<td>1.5</td>
<td>The population size before divergence was set to be 1.5 times $N_e$</td>
<td></td>
</tr>
</tbody>
</table>

that produced summary statistics very close to the observed patterns (Euclidean distance of 1 or smaller between the observed and the simulated summary statistics) were retained for further analyses. A post-sampling regression adjustment was performed under the ABC–GLM approach to improve the estimation of parameters (Leuenberger and Wegmann, 2010). Based on the results from the model selection procedure, the best supported model was then used to estimate some additional parameters, such as $N_1$ and $N_2$ as independent free parameters and the migration rate between basins ($m$). Here, both $N_1$ and $N_2$ were sampled from a uniform distribution and range 10–250. Migration rate was sampled from a uniform distribution with range 0.01–0.3 (proportion of the population per generation).

Results

The summary statistics produced the following results when applied to the empirical dataset: $S_{Mataquito} = 1$, $S_{Rapel} = 0.9359$, $D_{Mataquito} = −0.9180$, $D_{Rapel} = −0.7748$, $\pi_{Mataquito} = 1.7628$, $\pi_{Rapel} = 4.2051$, $F_{ST} = 0.0614$. The results from the simulations showed that Model I (no admixture) was unable to reproduce the observed patterns described above (see Fig. 4 for an example with two summary statistics). For example, Fig. 4 shows that although Model I was able to reproduce PI values close to the observed values in a small proportion of simulations, $F_{ST}$ values produced by this model were far from those observed in the empirical data. On the contrary, Model II (recent admixture) produced simulated patterns that were very close to the observed patterns (Fig. 4). This pattern was consistent across all summary statistics studied, but more accentuated for $F_{ST}$ (Fig. 5). Based on the model selection procedure, the posterior probability of Model II was 1.0 versus a value of 0.0 for the Model I. This means that all the retained simulations (1237 simulations or 0.3%) were obtained from Model II and none was retained from Model I since all simulations from model I produced genetic patterns that were over the minimum Euclidean distance to the observed summary statistics. In addition, the marginal density of Model II was 10,341.8 resulting in a $p$-value of 1, while the marginal density of Model I was 4736E−241, resulting in a $p$-value of 0. Based on these results, we then used Model II to estimate some demographic parameters of interest. Although with generally wide posterior densities, parameter values were as follow. $N_e$ were 49.1 and 18.6 for the Rapel and the Mataquito basins, respectively (Fig. 6A and B). The migration rate ($m$) between basins was 0.23, indicating that about a 23% of each population is migrating between basins per generation (Fig. 6C). The mutation rate was estimated in 0.000027 substitutions per locus per generation (Fig. 6D). The simulated data was not informative for estimating the time of divergence ($T_{DIV}$; Fig. 6E).
Fig. 5. Summary statistic values obtained from the simulated data under the Model I (no admixture) and model II (recent admixture). $F_t$ was calculated between basins. The red dotted line in each plot represents the value obtained from the empirical data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 6. Posterior densities for parameters in Model II. This model is a modification of model II (best supported model) used for hypotheses testing, and includes the population size of Rapel basin ($N_{e2}$) as a free parameter (i.e. not constrained to be twice the population size of the Mataquito basin), and the migration rate parameter ($m$) not fixed. (A) Population size of Mataquito basin ($N_{e1}$), (B) population size of Rapel basin ($N_{e2}$), (C) migration rate ($m$), (D) divergence time ($T_{DIV}$), and (E) mutation rate (MTDNA_RATE).

Discussion

Our analyses showed that obtaining the observed genetic patterns of low genetic structure between basins is unlikely from a model without admixture, even for divergences as recent as 2000 generations ago (or 4000 ybp). On the other hand, the observed patterns were easily obtained with a model that takes into account current admixture. These results strongly support the scenario of dispersal through irrigation canals as the most likely explanation for the observed genetic patterns, and reject the possibility that they can represent a case of incomplete lineage sorting due to recent divergence. The generality of this finding needs to be investigated with other taxa, but similarly low genetic structure in another catfish (*Trichomycterus areolatus*; Unmack et al., 2009) suggests that dispersal between basins might be a more general phenomenon involving several species. Furthermore, the presence of these canals connecting other basins further south (e.g. the Itata and Biobio basins; Habit, 1994) suggests that the impact on
biodiversity might be even more geographically widespread than currently recognized, involving distinct biogeographic areas, and unique freshwater fish assemblages (Dyer, 2000).

Migration between basins may impact biodiversity at different scales (Olden, 2006). At a population level, inter-basin migrations have the potential to erode genetic differentiation that has been accumulated for thousands of years resulting in a lack of overall regional genetic diversity (Olden et al., 2004) and the disruption of local adaptations (Storfer, 1999). At the species level, homogenization may result in the introgressive hybridization between closely related species with the potential result of species merging, and consequently, a decrease in species richness (Rhymer and Simberloff, 1996). In addition, inter-basin migrations may cause extinctions by competitive exclusion. With scarce knowledge on the ecology of species, it is hard to predict whether ecologically similar species from different basins may coexist if they migrate into adjacent basins or the exclusion of the less competitive species will take place. In the case of diplomystids, the scarce literature about their ecology suggest that species and populations from different basins occupy the same trophic niche (Beltrán-Concha et al., 2012) and similar habitats (Arratia, 1983), which would suggest that either strong negative interactions or complete genetic homogenization may occur, depending on the degree of reproductive isolation between lineages. Finally, regional diversity may still decrease without extinctions by the simple fact of homogenization of the unique biotas of each basin. In other words, the distinctiveness of different biogeographic provinces can be reduced as endemics of each basin spread to adjacent areas, in detriment of the β diversity of the region (Olden, 2006).

Our results exemplify the power of ABC methods to distinguish between different demographic scenarios in general, and demonstrate how they can be successfully applied to test hypotheses in the context analyzed here in particular. Additional multifocus data may add statistical power to estimate multiple demographic parameters more accurately. However, the approach used here proved to be powerful enough for model testing.

Although our results are robust and provide strong evidence of inter-basin dispersal for D. cf. chilensis, confirmation and quantification with more direct types of evidence (e.g. mark and recapture techniques) would help to better estimate the magnitude of organismal movement between basins and the degree of impact on the aquatic community. Are all species equally likely to migrate between basins? When and under what conditions they migrate? What will be the consequences for the biodiversity of the region? What species will be more affected? What can be done to reduce the potential impact of this new threat? All these questions remain to be investigated and further research needs to be conducted to fully understand the impact of this poorly known, but probably common, environmental alteration on the conservation of freshwater biodiversity. Answers to these questions will be key to not only better assess the impacts of these alterations, but also to better plan and manage future developments in a region that face increasing water demands.

Acknowledgments

We thank to the Knowles's Lab, University of Michigan, for insightful discussions and three anonymous reviewers for helpful comments and suggestions. CMR thanks the “Becas Chile” scholarship program funded by CONICYT – Chile. This study was funded by Fondo Nacional de Desarrollo Científico y Tecnológico, Chile (FONDECYT # 1120872 to PV and FONDECYT # 1150154 to EH). E. Habitat is also supported by Red Doctoral REDOC.CTA, MINEDUC project UCO12102 at Universidad de Concepción.

References


