Riverscape Genetics:

Insights Into the Drivers of Divergence in Coastal Brazilian Fishes

by

Andréa Tonolli Thomaz

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in The University of Michigan
2017

Doctoral Committee:

Professor L. Lacey Knowles, Chair
Professor James Diana
Associate Professor Timothy Y. James
Professor Diarmaid Ó'Foighil
Emeritus Professor Gerald Smith

Andréa T. Thomaz
thomaz@umich.edu
ORCID iD: 0000-0002-9755-2674

DEDICATION

To my mom Ana

thank you for teaching me how to fly

ACKNOWLEDGMENTS

This dissertation started to be conceived years before I enter in my Ph.D. Exactly ten years ago I was finishing my undergraduate in Brazil and, with that, came the certitude that I would like to pursue academia and, more precisely, the project presented here. During these ten years many people helped me to overcome the several steps until the end of my Ph.D., and I'm truly thankful for all of them.

Firstly, I would like to thank the **University of Michigan**, the **Rackham Graduate School** and, more specifically, the **Department of Ecology and Evolutionary Biology** (EEB) and the **Museum of Zoology** (UMMZ) for opening their doors and giving me all the necessary support for an international student pursue its research in this institution. Thank you for always believing in my studies and financially supporting them with Block Grants, Rackham Graduate Student Research Grants, Rackham International Student Award, and Rackham Predoctoral Fellowship Program and the Tinkle Award that I was awarded during my Ph.D. Also, although I was not officially a member of the **Fish Division** at the UMMZ, I'm grateful for all the opportunities this division offered me, including the Ichthyology Student Award for developing my fieldwork in Brazil, the curator assistantship position that allowed me to gain expertise by working in the fish collection and for the Hubbs, Carl L. and Laura C. Fellowship that allowed me to work full time on my research for two years. All the handling of specimens, including paper work, shipping and housing at the fish collection, wouldn't be able to be done without **Doug Nelson**'s help, support and ichthyological stories. Also, **Robbin Murrell** and **Cindy Carl** were essential for all the logistic support and answering all my questions (and I had many!).

I would like to thank the **National Science Foundation** (NSF) for awarding my advisor and I with the Doctoral Dissertation Improvement Grant (DDIG). This support was essential for developing all the sequencing and methodologies necessary for the accomplishment of this dissertation.

During my master's I had the chance to attend a seminar presented by a professor with orange nail polish on her toes (not sure why I still remember this detail) that was developing the most innovative research in phylogeography, in my opinion. Years after, she accepted me as a

student in her lab and this was the most intense learning experience I have had in my life. In reality, it was not just about methodologies, but what I will really take for my life that I have learned from her is the way of approaching a question, the reasoning, the hypotheses testing, and thinking outside the box. **Lacey**, thank you for the opportunity of being your student, for all the conversations and discussions. I'm not going to lie, you are tough sometimes :) but I'm a better scientist today because of you, and I will keep working hard to improve.

Also, this dissertation wouldn't have been possible without the support of my first advisor, Luiz R. Malabarba (Brazil). **Mala**, I'm very lucky to have you as my "fish" advisor, and you will always be. I'm extremely grateful for your generosity with me over the entire time, helping me with fieldwork, permits, being my fish guide, talking about biogeography, making barbecues every time I visited Brazil, and for always believing in my capacity, thank you!

I'm grateful for having the chance to work with our "borrowed" postdoc **Mark Christie**, who introduced me to the world of simulations and river modeling with an incredible patience and dedication. The things I learned with you will have a life long impact on my research.

The work presented here was only possible with many fieldwork trips. I would like to thank all my friends that provided their time and fishing expertise, sorting, taking pictures, and celebrating every single individual that we got in our nets, they are: **Vinícius Bertaco, Fernando Carvalho, Tiago P. Carvalho, Juliano Ferrer, Alice Hirschmann, Fernando Jerep, Giovanni Neves, Udson Santos, Priscilla C. Silva and Juliana Wingert**. This work was only possible because of you guys.

During the six years in the same department and lab, many people have come and gone, but I was very lucky to have the chance to share some time with them, each of which had a positive effect on my research, and interesting memories/impressions I have of each one: **Diego** you are the guy for GIS; **Qixin** your scripts saved my life, **Anna** your heart is humongous; **Rob** and his calm walking style; **JP** playing Frozen when I started to freak out; **Tristan** and his always positive vibe; **Carlos** it was nice meeting you; **Mariah**, I will put you in a little box and keep it in my closet**; Katie** I owe you a beer! **Raquel** thank you for all hugs and advice - I love them; **Jordan** you are not from my lab, but I like to think you are; **Pame** crazy for R. I also want to thank **Joyce**, **Lu** and **Rê** for being my Brazilian family here. Also, Ann Arbor gave me the chance to make good friends. Thank you all for the delicious brunches we had (with mimosas), the pottery fun, after work wine or beer to relax, swimming on the river, the running challenges,

bike rides to Dexter, dart games, snow fights, soccer games, dancing, numerous barbecues, etc. Thank you all for the good times!

Tiago, thank you for being my boyfriend, my best friend and my favourite ichthyologist. I know being almost seven years apart it was not easy, but, independent of the distance, you are always present in my thoughts and in my heart. Your support over all these years was fundamental to me.

Finally, I need to thank my foundation, my **family**. Even though they do not understand very well what I do, they always gave me the unconditional emotional and financial support. **Dad**, thank you for always pushing me to do my best (since my grades were never good enough to you :p). **Mom**, thank you for raising me to be an independent woman. **Kelly,** thank you for your affection. **Paty**, you explained to me the meaning of a Ph.D. when I was in the sixth grade, and since then I was dreaming of this day.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The movement of organisms in spatially structured landscapes is affected by constraints imposed by geographic and physical properties of the environment, and by the response of the organisms to this environment (i.e., ecological requirements). Freshwater environments, especially rivers, are known for imposing stronger movement constraints than terrestrial and marine environments. These constraints are associated with the isolation of different river drainages and the properties of a river itself, such as shape and water flow. Because of these unique characteristics of riverine landscapes (riverscapes), our understanding of neutral demographic processes in these environments is still lacking relative to that of other environments. This dissertation research aims to help fill this knowledge gap by advancing the understanding of the effects of riverine environments on neutral demographic processes. I combine simulated and empirical data to ask how riverine basins over spatial scales (i.e., local and regional) and temporal scales (i.e., present and past) interact with organisms to promote the observed patterns of genetic diversity in freshwater fishes. The Brazilian coastal drainages are an ideal area for this study as a series of isolated basins that were cyclically connected and disconnected because of Pleistocene sea level changes lead to a great diversity of endemic fishes. In my dissertation, I first demonstrate that paleodrainage structure during the Pleistocene is the main factor explaining population genetic differentiation in one species. Then, I give insights about how riverine landscapes and their physical properties (including during past time periods) structure genetic diversity within drainages. Finally, I used a comparative approach to elucidate whether sea level changes in coastal areas affected the freshwater community as a whole, or if responses were species-specific. The work presented here advances knowledge pertaining to the evolution of freshwater fishes, particularly those of the Neotropics. Overall, by exploring relevant hypotheses in order to identify processes that structure genetic variation within and between basins and species, my dissertation distinguishes the evolutionary mechanisms operating at different spatial and temporal scales, and provides insights into patterns of genetic diversity in freshwater fishes, especially along the coastal Brazilian basins.

CHAPTER I

Introduction

Background

Understanding the processes leading to patterns of genetic diversity and differentiation in organisms is crucial for elucidating the extent to which populations are isolated in terms of geographic and time scales. Dispersal has been identified as a mechanism for genetic differentiation among populations in both theoretical (Slatkin, 1987; Excoffier *et al.*, 2009) and empirical studies (Lester *et al.,* 2007; Row *et al.,* 2010). The movement of organisms in spatially structured landscapes is affected by constraints imposed by the environment in terms of geography and physical properties (Lee and Mitchell-Olds, 2011), and by the response of the organisms to this environment (i.e., ecological requirements). However, when correlations are made between levels of genetic differentiation and particular characteristics of the landscape, the underlying mechanisms responsible for the observed patterns are unclear (i.e., whether observed genetic differentiation reflects the environment constraints on population sizes and/or movement patterns).

Freshwater environments, especially riverine environments, are known for imposing stronger movement constraints on organisms in comparison to terrestrial and marine environments. River basins are frequently compared to island-like systems (i.e., island biogeography theory; MacArthur and Wilson, 1967) given the low levels of migration between basins, the small population sizes and rapid divergence between populations (Tedesco *et al.,* 2012). These demographic processes are a result of the physical subdivision of populations between drainages and the riverine properties, such as the dendritic shape and the water flow direction. Because of these unique features of riverine landscapes (i.e., riverscapes), a better understanding of the neutral processes in these environments over time and space are still lacking in comparison to other environments. This dissertation research aims to help fill this knowledge gap by advancing our understanding of the effects of riverine environments on demographic-neutral processes. For that, I combined simulated and empirical data to focus on how riverine basins over different spatial (i.e., local and regional) and temporal (i.e., past) scales interact with

organisms to produce observed patterns of genetic diversity. By taking this approach I provide insights to why freshwater fishes are among the most diverse vertebrate groups (~20-25% of all vertebrates and >40% of all fishes), even though they inhabit the less abundant environment on the planet, with fresh waters occupying less than 0.01% of all the water available (Lundberg *et al.,* 2000). In addition, besides advancing knowledge on phylogeographic processes shaping genetic diversity of freshwater fishes, it also provides insights into conservation of riverine ecosystems, with a direct application to a biodiversity hotspot - the Atlantic Coastal Rainforest of Brazil.

Freshwater Fish Phylogeography

The term phylogeography was coined with the study of freshwater fishes. Although the term phylogeography was introduced in 1987 (Avise *et al.,* 1987), the seminal work of Bermingham and Avise in 1986 can be considered as the first phylogeographic study (although they used the term "Molecular Zoogeography"). It used a comparative approach to characterize population structure of freshwater fish species in the southeastern United States and compare it with geologic information. Since then the field of phylogeography largely developed (e.g., Avise 2000; Knowles, 2009) by incorporating environmental properties rather than strict physical properties (e.g., ecological and climatic; Neuenschwander *et al.,* 2008; Carnaval, 2009; Bemmels *et al.*, 2016) and adding a statistical framework to test hypothesis on the top of descriptive patterns (Richards *et al,* 2007; Fagundes *et al*., 2007; Ray *et al.,* 2010; He *et al.,* 2013). However, given the constraint riverine environments impose on dispersal and migration, most of these newly developed tools and methods do not apply well to these systems, indicating the necessity to adapt these methodologies for restricted environments such as rivers (but see Neuenschwander *et al.,* 2008)

The most studied cases of genetic diversification and rapid speciation in freshwater fishes are in lake organisms and marine-freshwater transitions, in which adaptation and selection are pointed as the main process in their evolutionary history (e.g., cichlids in Africa, salmon and sticklebacks, summarized in Seehausen and Wagner, 2014). Research on riverine fishes has mostly focused on the role of allopatry and river captures in promoting speciation (Albert and Reis, 2011; Day *et al.,* 2013). For example, in the two most studied regions of the world, the phylogeographic patterns inferred for freshwater fishes are mostly related to the glacial history.

In North America, the northern area was glaciated, presenting a depauperated ichthyofauna, while the Mississippi basin has the highest diversity (Lundberg *et al.,* 2000) with a history of postglacial dispersal among drainages (e.g., darters; Bossu *et al,* 2013). In a similar way in Europe, the refugia hypothesis during glaciations and posterior dispersion by watershed crossings is largely accepted (e.g., *Cottus gobio*; Vonlanthen *et al.,* 2007; Neuenschwander *et al.,* 2008). However, in regions where species distributions and dispersal capability were not directly affected by the presence of a glacier, understanding the phylogeographic processes may be a little more challenging, especially when the region is the more diverse region for freshwater fishes in the world – the Neotropics.

Freshwater Fish Phylogeography in the Neotropics

The Neotropics is by far the region with the highest diversity of freshwater fishes in the planet. The region currently has more than 5,000 freshwater fish species described, which represents one-third of all freshwater fishes in the world (Reis *et al.,* 2016), and lots of diversity still remains to be described. Phylogeography can be an important tool to unveil the processes responsible for this high diversity.

The first paper in phylogeography of Neotropical fishes dates back to 1998, and used a comparative approach to understand the Great American Interchange (Bermingham and Martin, 1998; see also Dergam *et al.,* 1998). In the almost 20 years since the start of phylogeography of freshwater fishes in the Neotropics, the field largely focused on delineating species boundaries, given the high necessity to describe the undocumented ichthyofauna in the Neotropics (as summarized by Willis, 2017) and to use gene sequences to help in identifying lineages (e.g., DNA barcoding; Pereira *et al.,* 2013).

While historical biogeography is helping to understand the deeper history of suprageneric groups in the continent (e.g., family level; Montoya-Burgos 2003; Bloom *et al.,* 2013; Tagliacollo *et al.,* 2015; Silva *et al.,* 2016), phylogeography can give us important insights about the processes driving speciation and uncovering the recent history of lineages during the Pliocene-Pleistocene (e.g., Willis *et al.,* 2010; Hubert *et al.,* 2007; Cardoso and Montoya-Burgos 2009). Addressing broader hypotheses of riverine evolution in the Neotropics rather than focusing on species-specific questions can help us to build a better understanding of recent diversification of the ichthyofauna. For example, the primary riverine hypothesis tested at the

population level in the Neotropics is the color adaptation in the Amazonian rivers (i.e., Negro basin), which seems to be supported by an array of species (e.g., Cooke *et al.,* 2014).

While there have been several phylogeographic studies in the Neotropics, they have focused on the Amazon region, therefore, are usually restricted with large gaps in sampling of the species, given the difficulties with access to the region. Moreover, a large array of geomorphological events may contribute to patterns of genetic variation in Amazonian fish, and these are not yet fully understood (Hoorn *et al.,* 2010). These difficulties impose challenges to model evolution of the ichthyofauna in the region. Studying an area with fewer cofounding variables (e.g., outside migration) and with a better understanding of its geological and climatic history may be a good start to understand diversification in tropical areas.

Study Region

The Brazilian coastal basins are characterized by a series of small and isolated rivers that drain directly to the Atlantic Ocean (Figure I.1). The area studied here ranges from Patos drainage in southern Brazil to Caravelas drainage in the northeast, across a transect of approximately 2,000 Km. The region has one of the world's hotspot of diversity – the Atlantic Rainforest. While the terrestrial diversity has been focus of several studies because it is highly endangered (e.g., Carnaval *et al.*, 2009), the freshwater ecosystems in this region remain understudied, even though they are also endangered and face additional human disturbances like water pollution.

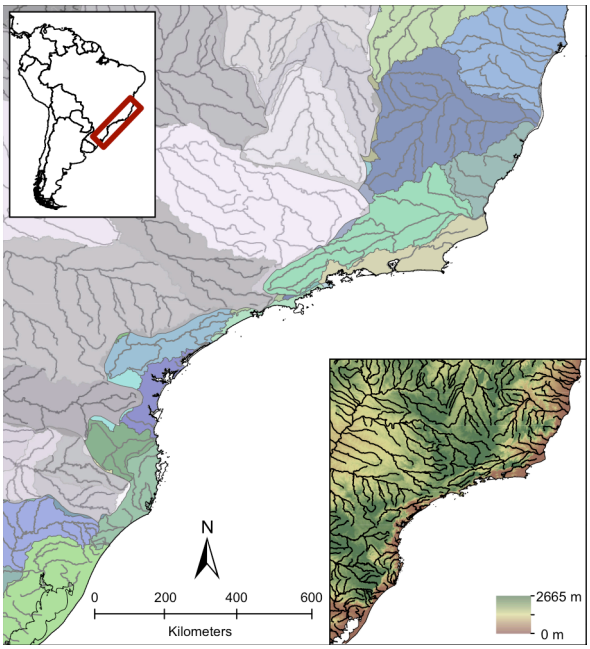


Figure I.1. Map of the Brazilian coastal drainages with each coastal drainage system shown in a different blue-green shade and inland drainages shown in shades of grey. Map in the lower-right corner shows the elevation difference between coastal and inland areas.

The area is ideal to understand demographic processes in rivers at the regional and local scale because it is formed by a series of isolated basins. The steep mountain slope of the Serra do Mar (Figure I.1) divides the coast from the inland basins of Brazil. Thus, the freshwater fishes in the coastal basins have infrequent migration from other areas, allowing a long-term isolation with a unique biogeographic history (Weitzman *et al.*, 1988; Ribeiro, 2006). Because of this

geological history, the region has 528 valid species of freshwater fishes (Figure I.2; unpublished data). Therefore, although the region is characterized by relatively low species diversity compared to the Amazon basin, the geographic isolation promoted diversification within this region, leading to high levels of endemism. For example, 78% of the species in this region are endemic (unpublished data). Given the impressive endemic diversity found in this long but narrow hotspot area, my dissertation will use this region as a model to understand evolutionary processes that led to this diversification.

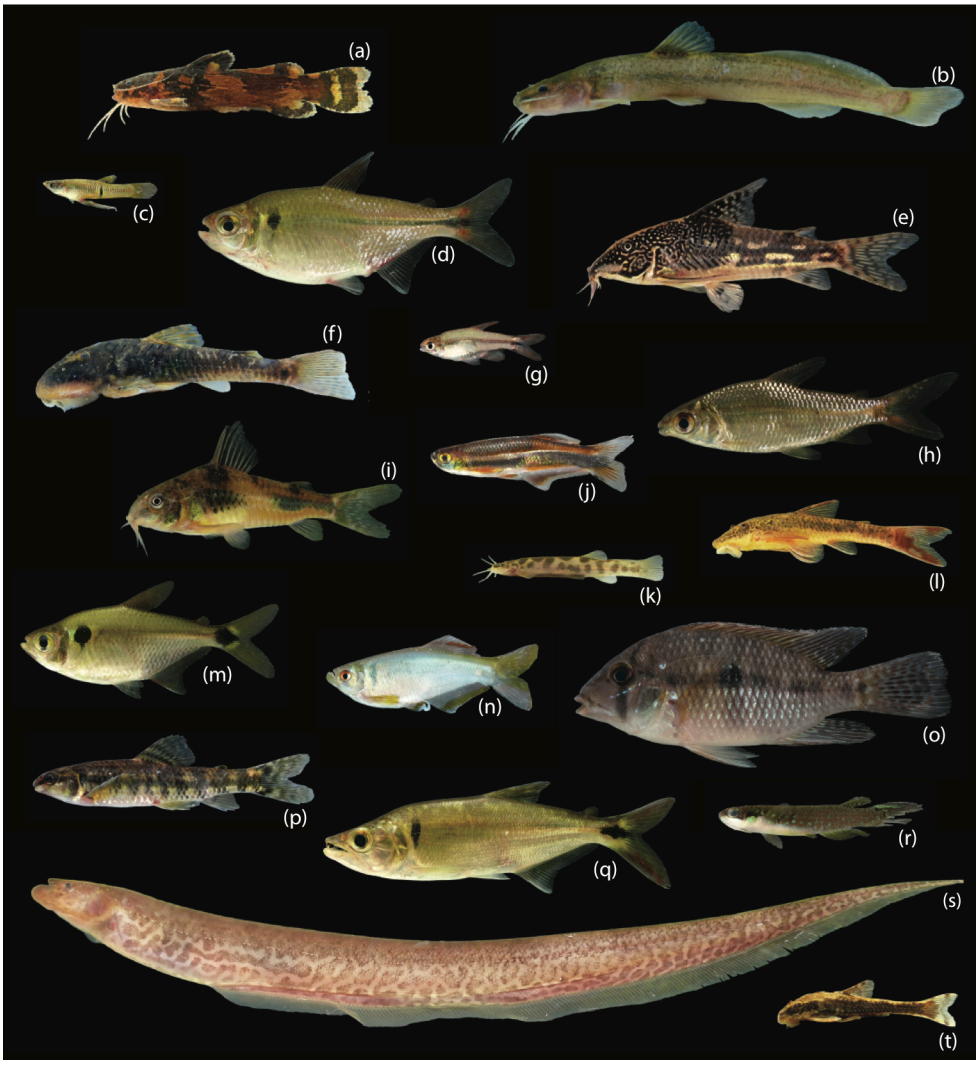


Figure I.2. Pictures demonstrating the diversity of freshwater fishes in the Brazilian coastal drainages: *Microglanis cottoides* (a; standard length = 43.2mm SL); *Heptapterus* sp. (b; 89.2mm SL); *Phalloceros* sp. (c; 22mm SL); *Deuterodon longirostris* (d; 59.6mm SL); *Scleromystax barbatus* (e; 65mm SL); *Pareiorhaphis splendens* (f; 57.4mm SL); *Spintherobolus ankoseion* (g; 23.7mm SL); *Cyphocharax sanctaecatarinae* (h; 46.9mm SL); *Corydoras erhardti* (i; 50mm SL); *Mimagoniates lateralis* (j; 30.5mm SL); *Trichomycterus* sp. (k; 30mm SL); *Parotocinclus maculicauda* (l; 44.1mmSL); *Probolodus heterostomus* (m; 45mm SL); *Mimagoniates rheocharis* (n); *Geophagus brasiliensis* (o; 72.5mm SL); *Characidium pterostictum* (p; 55mm SL); *Oligosarcus hepsetus* (q; 65mm SL); *Atlantirivulus* sp. (r; 31.5mm SL); *Gymnotus pantherinus* (s; 186mm SL); *Pseudotothyris ignota* (t; 33.2mm SL).

Chapters Overview

Historically, connectivity among coastal rivers was influenced by an increase in the severity of glacial cycles caused by pronounced sea-level retreats during the Pleistocene (e.g., Last Glacial Maximum (LGM) = -120 m; Hewitt, 2000). These fluctuations are hypothesized to have provided connections between drainages that are isolated today (Weitzman *et al.,* 1988), forming larger paleodrainages (Figure I.3). **Chapter II** tests this hypothesis and demonstrates the contribution of these paleodrainages during the Pleistocene as the main factor explaining observed genetic differentiation among populations in one freshwater fish species in the study area (Thomaz *et al.*, 2015). This work demonstrates that paleodrainage structure is key to interpreting biogeographic breaks in the fish fauna in rivers along the Brazilian coast. It also shows that genetic diversity is not evenly distributed among coastal basins, which have different geographical and physical properties. With the objective to expand from the paleodrainages proposed in chapter II and to clarify the contribution of rivers' structure and test if there is general or unique responses between species on patterns of genetic diversity, the other three chapters of my dissertation focus on local and regional patterns of diversity.

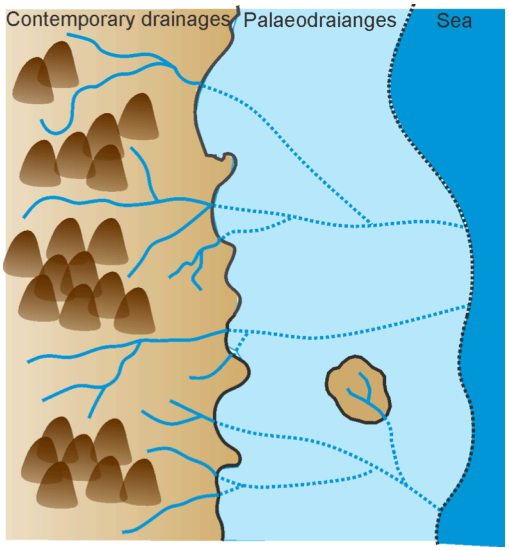


Figure I.3. Schematic draw showing current basins (brown area) and palaeo-rivers connections (light blue) during Last Glacial Maximum (LGM).

Chapters III and IV aim to understand how riverine landscapes and their physical properties structure genetic diversity of the organisms within drainages, since organismal dispersal is potentially affected by a variety of riverine properties. This research highlights the importance of incorporating riverine properties into phylogeographic studies more broadly, as has been established in terrestrial systems, e.g., the impact of habitat suitability on dispersal probabilities (He *et al.*, 2013)**.** This research has a theoretical and an empirical portion. For the theoretical portion in **chapter III** (Thomaz *et al.*, 2016), I used computer simulations to generate expected patterns of genetic variation so the impact of specific river properties (e.g., the length of river segments, the number of connecting tributaries, and the flow regime) on movement patterns of a fish can be discerned. For the empirical study presented in **chapter IV** (Thomaz *et al.*, in

review), I apply fine scale genetic data (next generation sequencing) for the same species used in chapter II to expand the paleodrainages models and understand the reasons why genetic diversity is not evenly distributed among the coastal basins. For that I expand the simulated findings from chapter III by adding a past temporal layer to test if there is a signal of paleodrainages properties (i.e., area and number of rivers) explaining patterns of genetic diversity.

Based on the findings from the local patterns of diversity, **chapter V** approaches the regional patterns of diversity using multiple species. The objective of this portion is to understand the effect of dynamic environments, such as coastal areas, that are constantly being affected by geological and climatic processes, have on different species and their impact on regional patterns of genetic diversification. By applying a comparative phylogeographic approach, this chapter addresses matches and mismatches in the regional genetic structure among four co-distributed species of tetras (Characidae) and broadly discusses timing and potential processes that culminated in the endemic and species-rich groups of freshwater fishes in the eastern coastal drainages of Brazil.

References

Albert JS, Reis RE (2011) *Historical biogeography of Neotropical freshwater fishes*, University of California Press: Berkley.

Avise JC (2000). *Phylogeography: the history and formation of species*. Harvard university press: Cambridge.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, *et al.* (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Ann Rev Ecol Evol Syst* **18**: 489-522.

Bemmels JB, Title PO, Ortego J, Knowles LL (2016). Tests of species-specific models reveal the importance of drought in postglacial range shifts of a Mediterranean-climate tree: insights from integrative distributional, demographic and coalescent modelling and ABC model selection. *Mol Ecol* **25**: 4889-4906.

Bermingham E, Avise JC (1986). Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics* **113**: 939-965.

Bermingham E, Martin AP (1998). Comparative mtDNA phylogeography of neotropical freshwater fishes: testing shared history to infer the evolutionary landscape of lower Central America. *Mol Ecol* **7**: 499-517.

Bloom DD, Weir JT, Piller KR, Lovejoy NR (2013). Do freshwater fishes diversify faster than marine fishes? A test using state-dependent diversification analyses and molecular phylogenetics of New World silversides (Atherinopsidae). *Evolution* **67**: 2040-2057.

Bossu CM, Beaulieu JM, Ceas PA, Near TJ (2013). Explicit tests of palaeodrainage connections of southeastern North America and the historical biogeography of Orangethroat Darters (Percidae: Etheostoma: Ceasia). *Mol Ecol* **22**: 5397-5417.

Cardoso YP, Montoya-Burgos JI (2009). Unexpected diversity in the catfish *Pseudancistrus brevispinis* reveals dispersal routes in a Neotropical center of endemism: the Guyanas region. *Mol Ecol* **18**: 947-964.

Carnaval AC, Hickerson MJ, Haddad CF, Rodrigues MT, Moritz C (2009). Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*. **323**: 785-789.

Cooke GM, Landguth EL, Beheregaray LB (2014). Riverscape genetics identifies replicated ecological divergence across an Amazonian ecotone. *Evolution*, **68**: 1947-1960.

Day JJ, Peart CR, Brown KJ, Friel JP, Bills R, Moritz T (2013). Continental diversification of an African catfish radiation (Mochokidae: Synodontis). *Syst Biol* **62**: 351-365.

Dergam JA, Suzuki HI, Shibatta OA, Duboc LF, Júlio Jr HF, Giuliano-Caetano L, William C (1998). Molecular biogeography of the Neotropical fish *Hoplias malabaricus* (Erythrinidae: Characiformes) in the Iguaçu, Tibagi, and Paraná rivers. *Gen Mol Biol* **21**

Excoffier L, Foll M, Petit RJ (2009). Genetic Consequences of Range Expansions. *Annu. Rev. Ecol. Evol. Syst.* **40**: 481–501.

Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L (2007). Statistical evaluation of alternative models of human evolution. *PNAS* **104**: 17614-17619.

He Q, Edwards DL, Knowles LL (2013). Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution* **67**: 3386-3402.

Hewitt G (2000). The genetic legacy of the Quaternary ice ages. *Nature*. **405**: 907–13.

Hoorn C, Wesselingh FP, Ter Steege H, Bermudez MA, Mora A, Sevink J, *et al.* (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330**: 927-931.

Hubert N, Duponchelle F, Nunez J, Garcia-Da Vila C, Paugy D, Renno JF (2007). Phylogeography of the piranha genera *Serrasalmus* and *Pygocentrus*: implications for the diversification of the Neotropical ichthyofauna. *Mol Ecol* **16**: 2115-2136.

Knowles LL (2009). Statistical phylogeography. *Ann Rev Ecol Evol Syst* **40**: 593-612.

Lee CR, Mitchell-Olds T (2011). Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Mol. Ecol.* **20**: 4631–42.

Lester SE, Ruttenberg BI, Gaines SD, Kinlan BP (2007). The relationship between dispersal ability and geographic range size. *Ecol. Lett.* **10**: 745–58.

McArthur RH, Wilson EO (1967). *The theory of island biogeography*. Princeton University Press, Princeton, NJ.

Lundberg JG, Kottela M, Smith GR, Stiassny ML, Gill AC (2000). So many fishes, so little time: an overview of recent ichthyological discovery in continental waters. *Annals of the Missouri Botanical Garden* 26-62.

Montoya-Burgos JI (2003). Historical biogeography of the catfish genus *Hypostomus* (Siluriformes: Loricariidae), with implications on the diversification of Neotropical ichthyofauna. *Mol Ecol* ***12***: 1855-1867.

Neuenschwander S, Largiadèr CR, Ray N, Currat M, Vonlanthen P, Excoffier L (2008). Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Mol Ecol* **17**: 757-772.

Pereira LH, Hanner R, Foresti F, Oliveira C (2013). Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? *BMC genetics* **14**:1-20.

Ray N, Currat M, Foll M, Excoffier L (2010). SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics* **26**: 2993-2994.

Reis RE, Albert JS, Di Dario F, Mincarone MM, Petry P, Rocha LA (2016). Fish biodiversity and conservation in South America. *J Fish Biol* **89**: 12-47.

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. *Neotr Ichthyol* **4**: 225–246.

Richards CL, Carstens BC, Knowles LL (2007). Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *J Biogeogr* **34**: 1833-1845.

Row JR, Blouin-Demers G, Lougheed SC (2010). Habitat distribution influences dispersal and fine-scale genetic population structure of eastern foxsnakes (*Mintonius gloydi*) across a fragmented landscape. *Mol Ecol* **19**: 5157–71.

Seehausen O, Wagner CE (2014). Speciation in freshwater fishes. *Annual Rev Ecol Evol Syst* **45**: 621-651.

Slatkin M (1987). Gene flow and the geographic structure of natural populations. *Science* **236**: 787–792.

Silva GS, Roxo FF, Lujan NK, Tagliacollo VA, Zawadzki CH, Oliveira C (2016). Transcontinental dispersal, ecological opportunity and origins of an adaptive radiation in the Neotropical catfish genus *Hypostomus* (Siluriformes: Loricariidae). *Mol Ecol* **25**: 1511-1529.

Tagliacollo VA, Roxo FF, Duke-Sylvester SM, Oliveira C, Albert JS (2015). Biogeographical signature of river capture events in Amazonian lowlands. *J Biogeogr* **42**: 2349-2362.

Tedesco PA, Leprieur F, Hugueny B, Brosse S, Dürr HH, Beauchard *et al*. (2012). Patterns and processes of global riverine fish endemism. *Global Ecol Biogeogr* **21**: 977-987.

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

Thomaz AT, Malabarba LR, Bonatto SL, Knowles LL (2015). Testing the effect of palaeodrainages versus habitat stability on genetic divergence in riverine systems: study of a Neotropical fish of the Brazilian coastal Atlantic Forest. *J Biogeogr* **42**: 2389-2401.

Thomaz AT, Malabarba LR, Bonatto SL, Knowles LL (in review). Genomic signatures of paleodrainages in a freshwater fish along the southeastern coast of Brazil: genetic structure reflects past riverine properties. *Heredity*.

Vonlanthen P, Excoffier L, Bittner D, Persat H, Neuenschwander S, Largiadèr CR (2007). Genetic analysis of potential postglacial watershed crossings in Central Europe by the bullhead (Cottus gobio L.). *Mol Ecol* **16**: 4572-4584.

Weitzman SH, Menezes NA, Weitzman MJ (1988). Phylogenetic biogeography of the Glandulocaudini (Teleostei: Characiformes, Characidae) with comments on the distributions of other freshwater fishes in eastern and southeastern Brazil. *Proceedings of a workshop on neotropical distribution patterns*. Academia Brasileira de Ciências, Rio de Janeiro, pp. 379–427.

Willis SC (2017). One species or four? Yes!... and, no. Or, arbitrary assignment of lineages to species obscures the diversification processes of Neotropical fishes. *PloS one* **12**.

Willis SC, Nunes M, Montana CG, Farias IP, Orti G, Lovejoy NR (2010). The Casiquiare river acts as a corridor between the Amazonas and Orinoco river basins: biogeographic analysis of the genus Cichla. *Mol Ecol* **19**: 1014-1030.

CHAPTER II

Testing the effect of paleodrainages versus habitat stability on genetic divergence in riverine systems: study of a Neotropical fish of the Brazilian coastal Atlantic Forest.

Andréa T. Thomaz, Luiz R. Malabarba, Sandro L. Bonatto, L. Lacey Knowles

Abstract

Aim: Patterns of genetic variation within freshwater fish populations may reflect the historical impact of climate change on either sea-level or environmental conditions. Sea-level changes enlarged paleodrainages and so connecting currently isolated rivers, whereas changes in environmental conditions reduced forest cover and may have constrained the movement of fish specialised to this habitat. We assayed genetic variation in *Hollandichthys multifasciatus*, a freshwater fish endemic to the Atlantic Forest of coastal Brazil, to test the relative importance of these factors in shaping current patterns of genetic divergence.

Location: River drainages along the southeastern Brazilian coast.

Methods: GIS was used to reconstruct paleodrainages during the Last Glacial Maximum (LGM). Niche modelling was used to infer areas of stability for the southern Atlantic Forest *sensu stricto* (present and LGM). The contribution of river connections inside or outside areas of stability was evaluated using a calibrated phylogeny, analyses of molecular variance, and Bayesian skyline plots from two mtDNA loci.

Results: Analyses of 182 individuals from 26 populations and 12 paleodrainages indicated that structure associated with paleodrainages explains 75% of the genetic variation among populations, with estimated divergence times occurring within the Pleistocene. The variation explained by paleodrainages and estimated population sizes was unrelated to the ecological stability of the region.

Main Conclusions: This study demonstrates the importance of Pleistocene paleodrainages in structuring genetic divergence patterns. The analyses suggest that past connections due to sea level retreat played a significant role in the diversification of the ichthyofauna along the

Brazilian coastal drainages. Moreover, the lack of a signature of habitat stability in structuring genetic variation suggests that refugia may be less important in structuring genetic diversity for freshwater species than for terrestrial species. In addition, our work highlights the utility of a GIS-based approach to recover past connections among coastal basins. Understanding these connections is crucial for studying diversification of riverine organisms and for identifying areas of conservation priority.

Introduction

Many studies try to understand how historical shifts in species distributions contribute to current patterns of genetic variation, but the factors taken into account tend to depend on the ecosystem of interest. For example, both geological and ecological processes are recognised as key factors structuring genetic variation in terrestrial organisms. This joint emphasis is apparent in the variety of methodologies used to test hypotheses, such as the coupling of ecological niche models and genetic models (e.g., He *et al*., 2013), and in comparative studies attempting to identify species-specific versus shared-community effects of past distributional shifts (e.g., Soltis *et al*., 2006; Papadopoulou *et al*., 2009; Massatti & Knowles, 2014). However, in riverine organisms, studies of genetic variation focus primarily on the role of allopatry related solely to geological events (i.e., past drainage connections; Bermingham & Martin, 1998; Carrea *et al*., 2013, Pereira *et al*., 2013). As such, these investigations tend to focus on patterns expected when geological events dictate the composition of regional species pools, as emphasised in the classic work on southeastern United States fishes (e.g., Bermingham & Avise, 1986). This emphasis is expected considering the constraints associated with the riverine lifestyle, where dispersal is restricted to past and present river connections.

In contrast, an alternative approach may consider how taxon-specific characteristics interact with geological events to produce species-specific structure of population genetic variation. For example, dispersal in riverine organisms may be constrained by ecological requirements related to the physical habitat (e.g., water depth and flow), or to the surrounding terrestrial vegetation and soil, which influences turbidity and food availability (Muneepeerajul *et al*., 2007; Cooke *et al*., 2014). In addition to ecological requirements, the inherent dispersal capabilities and behaviours of taxa may result in species-specific dispersal patterns (Whiteley *et al*., 2004; Burridge *et al*., 2008). As a consequence, the level of population connectedness

experienced by riverine organisms may differ among taxa and across geographic regions (Crispo *et al*., 2006; Neuenshwander *et al*., 2008).

Regions that experienced simultaneous shifts in hydrographic connectivity and habitat stability (e.g., Pleistocene forest habitat fragmentation) are ideal systems for studying how geological and ecological factors influence routes of population connectedness over time, and hence, the structure of genetic variation in freshwater organisms. The coastal drainages of the Atlantic Forest in southeastern Brazil are an example of such a region where both factors may be important.

The coast of Brazil was impacted by lower sea levels during Pleistocene glacial periods that exposed paleodrainages and connected currently isolated riverine basins (Weitzman *et al*., 1988; Dias *et al*., 2014). Climatic changes also resulted in shifts in the distributions of terrestrial ecosystems. In particular, the coastal Atlantic Forest is thought to have gone through pronounced shifts, with large portions transformed into grasslands; only small, isolated forest patches were stable over geological time (Behling, 2002; Carnaval & Moritz, 2008; Carnaval *et al.*, 2014). Both geological and ecological conditions may have influenced past distributions of the local endemic ichthyofauna associated with forested habitats, such as the freshwater tetra *Hollandichthys multifasciatus* (Eigenmann & Norris, 1900) (Characiformes: Characidae). This species is endemic to small coastal rivers within the Atlantic Forest *sensu stricto* of southeastern Brazil (i.e., lowland Serra do Mar forest or Dense Evergreen Forest; Olson *et al*., 2001; Fundação SOS Mata Atlântica, 2011; Figure II.1). Its distribution is restricted to rivers with dense canopy cover, where it inhabits crevices among plant roots along the river banks, and the bulk of its diet consists of terrestrial plants, spiders, and insects (Esteves & Lobon-Cervia, 2001; Abilhoa *et al*., 2009). Therefore, we hypothesise an important role for forest habitats in constraining the past distribution of *H. multifasciatus*.

However, patterns of population connectivity in *H. multifasciatus* were also potentially tied to past geological events that affected dispersal among the network of rivers dissecting the Atlantic Forest. Specifically, pronounced sea level retreat along the wide coastal plain during the Pleistocene glaciations (Fleming *et al*., 1998; Miller *et al.*, 2011) is predicted to have connected small and currently isolated rivers that rise on the eastern slope and flow directly into the ocean. Conversely, rises in the sea level of several meters above the present level during the Last Interglacial (LIG, *c.* 125 ka) and the early Holocene (*c.*12-7 ka) (Smith *et al.*, 2011; Dutton &

Lambeck, 2012) may have reduced the connectedness of the small coastal rivers, especially in the north of the species' distribution where the coastal plain is currently narrow. Therefore, sea-level changes during the Pleistocene may have had profound effects on the dynamics of these populations, potentially promoting isolation and reducing population sizes during marine transgressions, while providing connections and possibly increasing population sizes during marine regressions.

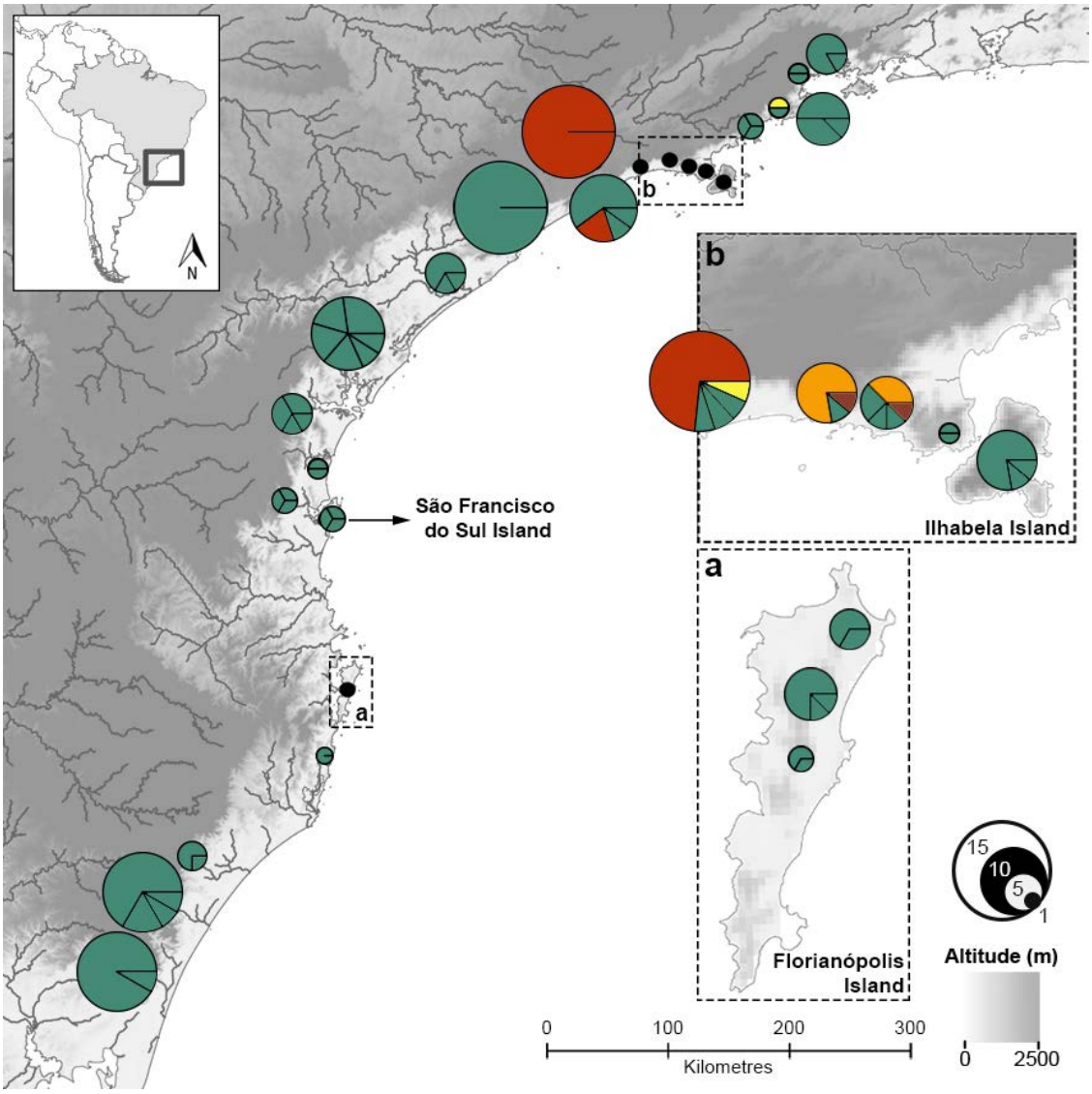


Figure II.1 – Geographic distribution of haplotypes among the 26 populations of *Hollandichthys multifasciatus* sampled along the Brazilian coast. Circle size represents the number of individuals sampled per population and circles are divided proportionally according to the number of individuals with a particular haplotype. Haplotypes restricted to the sampled population are shown in green, and haplotypes shared between two or more populations are correspondingly marked in yellow, orange, brown or red. Two regions with dense population sampling are shown in close-up (panels a and b) for better visualisation.

In this paper we test hypotheses related to how genetic diversity and structure might reflect several scenarios: (i) recent diversification of *H. multifasciatus* during the Pleistocene; (ii) opportunities for dispersal among currently isolated rivers by way of temporary routes provided by paleodrainages during glacial marine regressions in the Pleistocene; (iii) population size reductions and isolation of current basins because of interglacial marine transgressions; and, (iv) ecological constraints imposed by the association of *H. multifasciatus* with the Atlantic Forest, highlighting the importance of regions of habitat stability. To generate and test specific expectations for the structure of genetic variation during the Pleistocene, we couple analyses of genetic variation with modelling of both paleodrainage distribution based on bathymetric data analysed using GIS techniques, and modelling of Atlantic Forest stability using Environmental Niche Models (ENMs). If glacial periods provided corridors for dispersal among currently isolated basins due to sea level retreat, we predict that the structuring of genetic variation will correspond to regional patterns reminiscent of projected paleo-river basins (i.e., paleodrainages). Specifically, fish from rivers within a shared paleodrainage are expected to be genetically more similar to each other compared to fish collected from rivers from different paleodrainages. However, if genetic variation is dictated by isolation due to the current configuration of rivers, there should not be a significant contribution of paleodrainage in an analysis of the molecular variance among populations. Lastly, if the ecological association between the fish and the Brazilian Forest affected the fish's distribution over time, we expect to see a signature of habitat stability on patterns of genetic variation in a comparison of the proportion of genetic variation associated with paleodrainages from regions containing versus not containing forest refugia. If ecological stability plays an important role in structuring genetic diversity in *H. multifasciatus*, the persistence of long-term populations is more likely to be associated with ecologically stable areas (as determined by comparisons of ENMs for the present and the LGM). Hence, populations from ecologically stable regions would be more likely to show a general agreement with the paleodrainages model and the signature of larger historical population sizes than populations from regions of ecological instability.

Material and Methods

Inference of Paleodrainages

Paleodrainages during the LGM were inferred from topographic and bathymetric information extracted from the Digital Elevation Model (DEM) GEBCO_08 at 30 arc-second resolution (*c.*1km; http://www.gebco.net/) in ARCGIS10 using *Hydrological* tools. Past connections between current riverine basins and paleodrainages were identified from the inferred flow directions based on topographic relief. First, the land area exposed during the LGM, which corresponded to a maximum shift in sea level of -125m, was identified using the tool *Contour* followed by *Mask*. The *Fill* option was used to cover localised depressions in the surface. The *Flow Direction* tool was then used to identify the steepest descent from each cell, followed by the *Basin* tool to identify ridges between the basins that delineate the basins' borders. With the *Flow accumulation* tool, a raster of accumulated flow in each cell was calculated, after which putative rivers were estimated using the *Stream order* function. This method for inferring paleodrainages generally corresponds to paleochannel reconstruction methods used in other world regions (see Dias *et al*., 2014), and was used here because paleochannel data for this region of Brazil are scarce (e.g., Conti, 2009).

Atlantic Forest Niche Modelling

The current and past distribution of the Atlantic Forest *sensu stricto* (defined as the lowland Serra do Mar Forest physiognomy; Olson *et al*., 2001) was inferred for the study area using MAXENT 3.3 (Phillips *et al.*, 2006). Specifically, 1,000 points were randomly selected within the area of the present forest's distribution. These points were used to generate the present-day ENM based on 19 bioclimatic variables from the WorldClim 1.4 database (Hijmans *et al*., 2005). Then, the present-day model was projected onto the corresponding paleoclimatic data from ECHAM3 (Roeckner *et al*., 1992) for the LGM, interpolated to a 30 arc-second resolution (*c.*1km). Because over fitting is a known problem for large study areas (Anderson & Raza, 2010), we restricted our niche modelling inferences to a rectangle immediately surrounding the biome, which encompassed all localities in our study. The robustness of the modelling results was checked by subsampling 50 independent runs, with 75% of points used for training and 25% for testing in each run.

Areas of stability were characterised as regions where the Atlantic Forest persisted during both the present and the LGM (see also Carnaval & Moritz, 2008; Carnaval *et al*., 2014). The suitability threshold applied to differentiate stable from unstable areas was based on the stability score that maximised the area under the curve (AUC) values and minimised omission rates, and that better predicted the current distribution of the forest (i.e., less over or under fitting = maximum training sensitivity plus specificity logistic threshold). This threshold was applied to the LGM and present-day ENMs. Note that it was not possible to project the occurrence of the forest onto the coastal plain exposed when the sea-level dropped. However, such regions would, by definition, represent unstable areas (i.e., not continuously occupied over time), which is the primary factor being tested here given that stable forest areas have been proposed as targets of conversation and considered key in structuring genetic variation in terrestrial inhabitants of the Atlantic Forest.

Tissues and molecular genetic methods

Populations across the entire *H. multifasciatus* distribution were sampled. Specifically, 182 individuals from 26 populations were collected. Tissues were preserved in 96% ethanol and catalogued in the collections of the ichthyology laboratory in the Universidade Federal do Rio Grande do Sul (UFRGS), Museu de Ciências e Tecnologia, Pontifícia Universidade Católica do Rio Grande do Sul (MCP) and Museu de História Natural Capão da Imbuia (MHNCI) (see Table A.1). DNA was extracted following the modified salt-precipitation protocol (Medrano *et al*., 1990). Two mitochondrial genes, cytochrome oxidase I (COI) and NADH dehydrogenase 2 (ND2), were amplified and sequenced following standard protocols (for details see Thomaz *et al*., 2010), with forward and reverse chromatogram assembled and visualised using PHRED/PHRAP/CONSED (Ewing *et al.*, 1998; Gordon *et al.*, 1998). Sequences were aligned using the MUSCLE 3.6 (Edgar, 2004), checked by eye, and deposited in GenBank (see Table A.1). Standard population genetic summaries were calculated for each population and paleodrainages using DNASP (Librado & Rozas, 2009) and ARLEQUIN 3.5 (Excoffier & Lischer, 2010).

Phylogenetic inference and demographic dynamics

Haplotype sequences from the two concatenated mitochondrial genes were used to estimate a tree in BEAST 1.8.0 (Drummond *et al.*, 2012) under a coalescent constant size prior

and HKY+G model that was selected using Akaike's information criteria (AIC) in PARTITIONFINDER (Lanfear *et al.*, 2012). The tree was estimated from two independent runs; each run was performed with 200 million MCMC iterations and 10,000 trees were retained with 10% discarded as burn-in. No fossils were available for calibration to assess the timing of diversification within *Hollandichthys.* To test whether this timing occurred during the Pleistocene, we calibrated the phylogeny with a relatively conservative substitution rate for fish mtDNA (1% per Myr; e.g., Bermingham *et al*., 1997; Zardoya & Doadrio, 1999; Strecker *et al.*, 2004; Ornelas-Garcia *et al*., 2008) under a strict clock model and locus specific rates with a uniform prior [1E-7, 1E-9]. This procedure is justified within species or closely related species (Li & Drummond, 2012).

Population sizes through time were estimated with Bayesian Skyline plots (BSP – Drummond *et al.*, 2005), using the same parameters as above. These analyses were used to contrast the population dynamics over time for the stable area (i.e., the northern clade in the tree estimated with BEAST) and the unstable area (i.e., the southern clade in the tree estimated with BEAST). Note that the population from the Guaraqueçaba drainage was excluded from these analyses because it resides within an unstable area, even though it is more closely related to populations from stable areas in the northern clade (see results below). The stable area analysis included 108 individuals, whereas the analysis of the unstable region included 63 individuals. Several runs were performed to confirm the robustness of the results to different priors.

Testing the effect of paleodrainages and habitat stability

A series of nested analyses of molecular variance (AMOVA; Excoffier *et al.*, 1992) were performed in ARLEQUIN 3.5 to evaluate the contribution of paleodrainages and forest stability (inferred from the ENMs as described above) to population genetic structure. Specifically, an AMOVA was performed with populations nested within paleodrainages (= groups), as well as two additional AMOVAs with only those paleodrainages from either (a) stable areas or (b) unstable areas. Comparison of the results from the AMOVAs for stable and unstable areas provide insights into whether the stability of forest coverage has been important in structuring genetic variation in the fish, as with their terrestrial counterparts (e.g., Carnaval *et al*., 2009; Martins, 2011). Because a minimum of two populations per paleodrainage (i.e., group) is required in such hierarchical analyses of molecular variance, some populations had to be

excluded (see Table II.1 for details). For each AMOVA, significance was evaluated with 10,000 permutations. Pairwise genetic distances between paleodrainages were also calculated using ARLEQUIN 3.5 under a Jukes and Cantor substitution model.

Results

A total of 1,704 base pairs of COI (639 bp) and ND2 (1,065 bp) sequenced in 182 individuals resulted in 70 haplotypes, the majority of which were unique among populations (Figure II.1). Haplotype sharing is limited to a couple of the northern populations separated by a narrow coastline (e.g., the Bertioga, Ilhabela and Upper Tietê populations - see Figure II.1 and Table A.1). As a result, a very high partitioning of genetic variation was observed among populations (96.7%), compared with low variation represented within populations (3.3%). There was neither a substantial difference in genetic diversity among populations, nor an obvious latitudinal cline in genetic diversity (Table II.1), with the exceptions of two northern populations, Peruíbe and Upper Tietê, which are characterised by a single haplotype (Figure II.1). Divergences among *H. multifasciatus* populations date to the Pleistocene (mean = 2,161 kya and 95% HPD = 1,671-2,272 kya) based on the calibrated phylogeny (Figure II.2a and Figure A.1). Note that the estimated Pleistocene divergence is robust to other rates of molecular evolution that have been suggested in the literature (e.g., Zardoya & Doadrio, 1999; Strecker *et al.*, 2004), including some that are exceptionally low (i.e., 0.8%; Ornelas-García *et al.*, 2008), although the absolute timing of divergence would differ.

The 26 populations sampled from the current basin system correspond to 12 paleodrainages (Figure II.3) inferred using the GIS procedure described above. The ENMs inferred a forest-wide refugium region during the LGM that coincides with the northern *Hollandichthys* populations (Figure II.4), generating an area of stability (AUC = 0.87) that encompasses 14 of 26 sampled populations and 8 of 12 inferred paleodrainages. Note that this estimate of the Atlantic Forest during the LGM corresponds to the one proposed by Carnaval *et al.* (2009) and others (Amaro *et al*., 2012; Porto *et al*., 2013), but not Carnaval & Moritz (2008) and Carnaval *et al*. (2014). Estimations of population dynamics through time using the Bayesian skyline plots showed no discernable difference between populations of the stable (i.e., northern) and unstable (i.e., southern) clades (Figure II.2b and d). Instead, the analyses suggest that both regions underwent a recent population expansion that began around the end of the LGM (but see

Grant, 2015 for cautions about Bayesian Skyline plot inferences). One alternative explanation for the inferred increase in population size is a high degree of population structure. The preponderance of unique haplotypes across drainages (i.e., haplotypes that are not shared among populations; Figure II.1), would contribute to an elevated estimated population size. In other words, the recent structuring of populations within a region, even without significant population growth, may contribute to this apparent increase in the overall population size.

Table II.1 – Summary of the genetic data and the location of all 26 populations sampled for this study, as well as genetic diversity for the paleodrainages and areas of stability. Membership of sampled populations in individual paleodrainages is highlighted by groupings of either grey or white, and the black line marks the split between northern stable (above) and southern unstable (below) areas.

	Population	N	# Haplotypes	Pop. π	Paleodrainages π	Stability π
1	Paraty 1	6	2	0.33		
2	Paraty 2	2	2	4.07	2.2 (±2.64)	
3	Toca do Boi*	8	3	0.79	0.79	
4	Ubatuba 1	2	2	< 0.001		
5	Ubatuba 2	3	3	1.34	0.67 (±0.95)	
6	Ilhabela	9	3	0.67		
7	São Sebastião 1	2	2	< 0.001		
8	São Sebastião 2	8	5	1.77	0.81 (±0.89)	
9	São Sebastião 3	9	3	0.22		
10	Bertioga	15	5	2.85	1.54 (±1.86)	
11	Upper Tietê	14	1	0		
12	Santos	10	4	0.94	0.47 (±0.66)	
13	Peruíbe*	14	1	0	0.00	
14	Ribeira de Iguape*	6	3	3.42	3.42	1.2 (±1.36)
15	Guaraqueçaba	11	6	4.2		
16	Paranaguá	6	4	5.08	4.64 (±0.62)	
17	Guaratuba	2	2	1		
18	São Francisco do Sul	3	3	4.08		
19	Babitonga	3	3	2.02	2.37 (±1.57)	
20	Garopaba*	1	1	0		
21	Florianópolis 1	3	2	0.67		
22	Florianópolis 2	11	5	1.8		
23	Florianópolis 3	6	2	3.28	1.44 (±1.44)	
24	Araranguá	4	2	0.5		
25	Mampituba	12	4	0.92		
26	Maquiné	12	2	0.34	0.59 (±0.3)	2.0 (±1.74)
		182	of 70	1.55		

* Populations not considered for the AMOVAs analyses

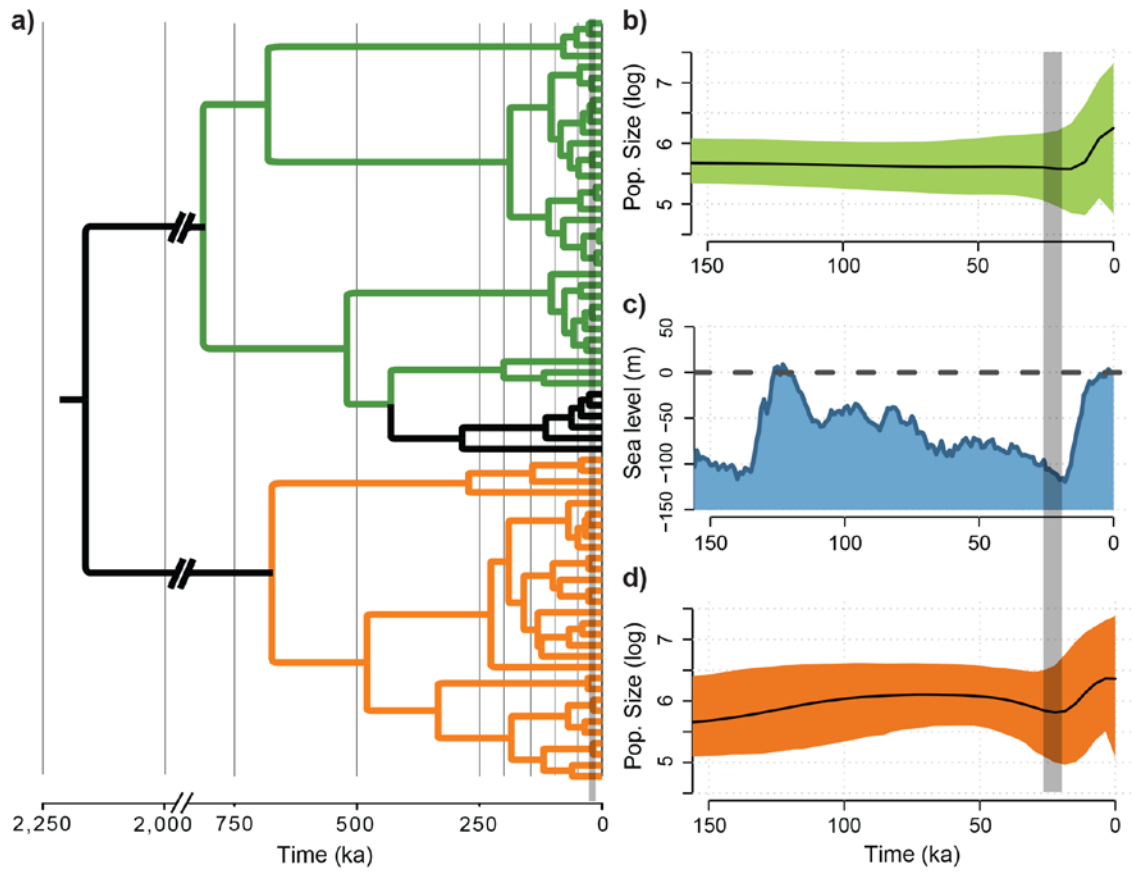


Figure II.2 – Lineage diversification and population dynamics over time estimated for individuals sampled from the stable northern area (in green) and unstable southern area (in orange) from (a) calibrated phylogenetic tree of *H. multifasciatus* haplotypes (with the excluded Guaraqueçaba population shown in black; see results for details). Population size was estimated with BSP for the (b) stable (green) and (d) unstable (orange) areas, with the corresponding (c) changes in sea level for the last 150 kya shown (Miller *et al.*, 2011), and grey shading marks the LGM period (26 – 19 kya).

Pairwise genetic differences vary among paleodrainages (Figure II.5), ranging from relatively high levels in the southern paleodrainages to fairly low levels of differentiation among some of the northernmost paleodrainages. Differentiation between paleodrainages was generally higher compared to pairwise genetic difference within paleodrainages (shown along the diagonal in Figure II.5). The only exception was an exceedingly high pairwise genetic difference between populations of the paleodrainage associated with the Paranaguá estuary (the Paranaguá and Guaraqueçaba populations: Table II.1 and Figure II.5). These populations, despite their geographic proximity and assignment to the same paleodrainage based on bathymetric data, are distantly related (i.e., they are assigned to divergent clades in the phylogenetic tree, with a divergence time of ~2 Ma; Figure II.2a and Figure A.1).

Tests of the partitioning of genetic variation showed that paleodrainages account for 75% of the genetic variation, while populations within paleodrainages explaining 22% (Table II.2). Comparing the AMOVAs for the stable region (80 individuals in 11 populations and 5 paleodrainages) and unstable region (73 individuals in 11 populations and 4 paleodrainages) shows that paleodrainage explains a substantial proportion of the genetic variation in both areas (i.e., 59% versus 47%, respectively; Table II.2 and Table A.2).

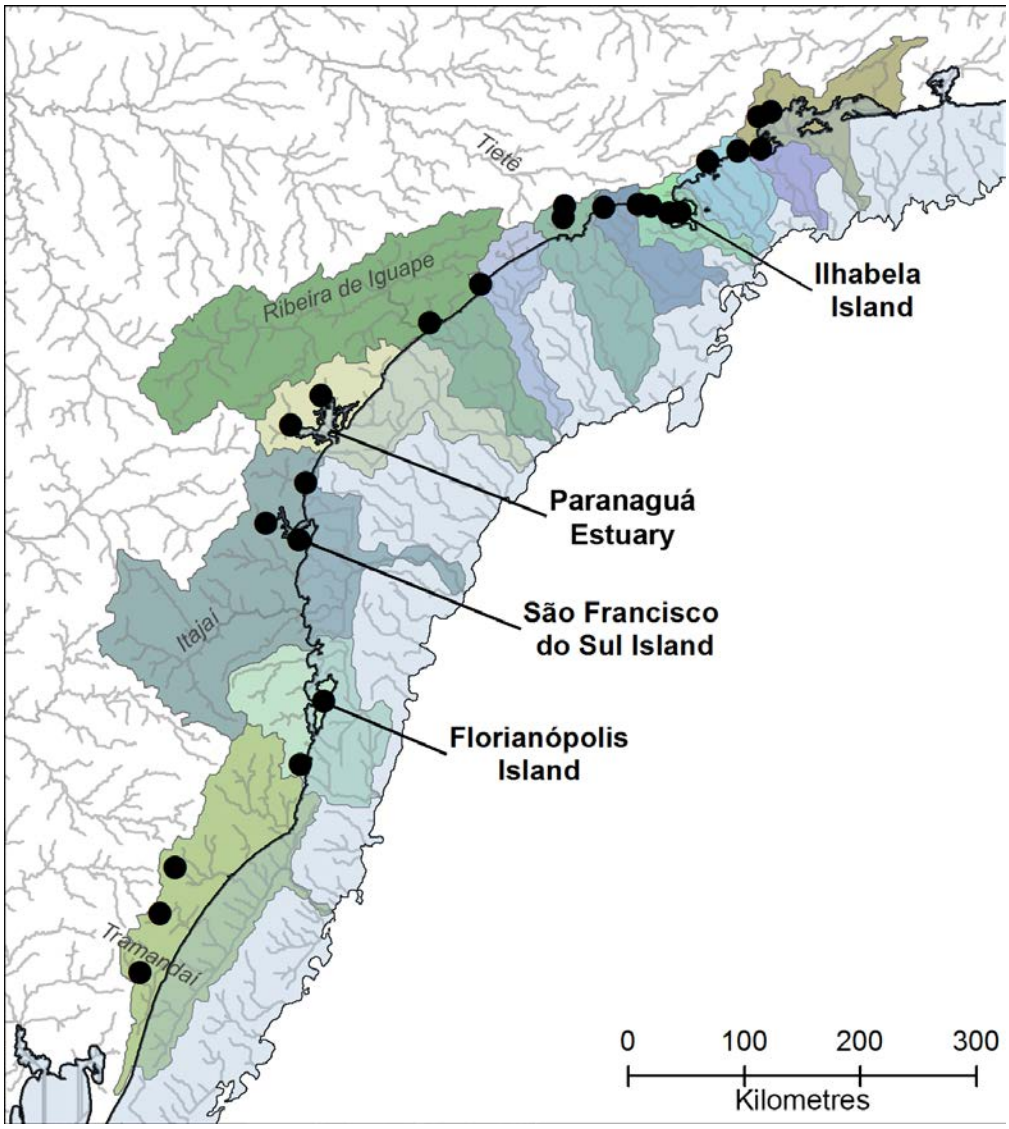


Figure II.3 – Map of the 12 paleodrainages (each shown in a different blue-green shade) identified by the area exposed with the retreat in sea level during the LGM (see methods for details) and associated populations sampled in this study (black dots). The black line indicates the current Brazilian shoreline and the three coastal islands with sampled populations: Florianópolis, São Francisco do Sul and Ilhabela, relative to what the shoreline would have been during the LGM (note that the grey areas in the exposed coast correspond to paleodrainages with no current analogs among the current configuration of sampled populations).

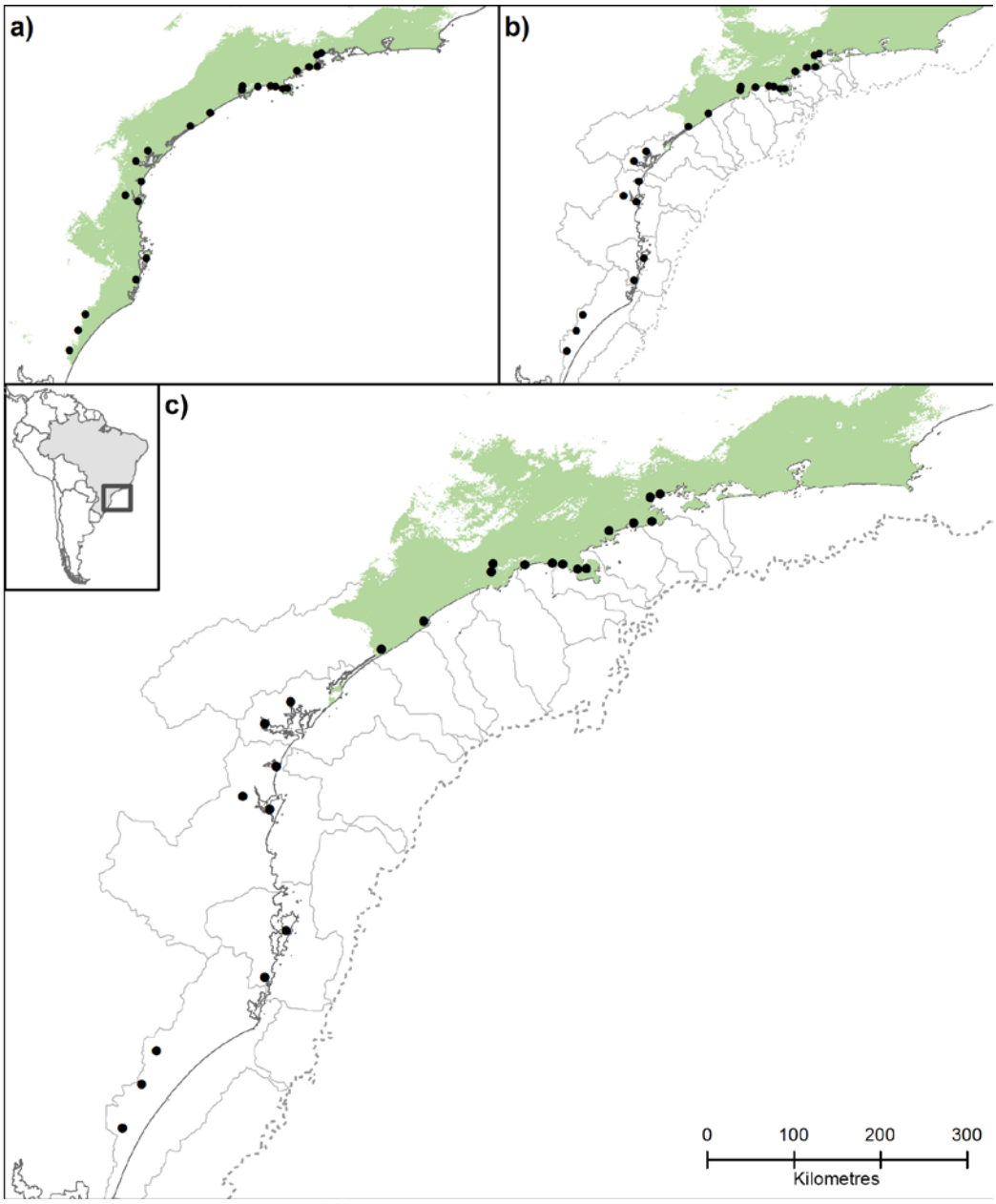


Figure II.4 – The ENM prediction for the (a) current distribution of the Atlantic Forest *sensu stricto*, (b) the LGM prediction, and (c) the stable area identified as the area of overlap (i.e., region of persistent forest cover). The distribution of the forest is shown in green. Grey lines identify the borders of paleodrainages considering a 125 m drop in sea level (the dashed line marks the coast line during LGM).

Table II.2 – Summary of AMOVAs results*.

Scenario	Source	% var.
Entire region	Among paleodrainages (Fct)	74.63 (p-value ≅ 0)
	Among pops. within paleodrainages (Fsc)	22.15 (p-value ≅ 0)
	Within pops. (Fst)	3.22 (p-value ≅ 0)
Stable region	Among paleodrainages (Fct)	59.26 (p-value < 0.05)
	Among pops. within paleodrainages (Fsc)	33.56 (p-value ≅ 0)
	Within pops. (Fst)	7.18 (p-value ≅ 0)
Unstable region	Among paleodrainages (Fct)	47.10 (p-value < 0.01)
	Among pops. within paleodrainages (Fsc)	46.81 (p-value ≅ 0)
	Within pops. (Fst)	6.08 (p-value ≅ 0)

*153 individuals from 22 populations and 9 paleodrainages were analysed (i.e., only paleodrainages with at least two populations could be analysed in the hierarchical AMOVA).

Discussion

By applying an explicit hypothesis-testing approach, we identified paleo-connections during periods of Pleistocene sea level retreat as a major factor structuring recent divergence in Brazilian coastal freshwater fishes. By contrasting the influence of paleodrainages in regions of forest stability and forest instability, we show that stable forest refugia did not leave a discernable signature on patterns of population genetic structure in *H. multifasciatus,* a fish dependent upon the forest habitat in the Atlantic Forest. This lack of difference between stable and unstable areas contrasts with work on terrestrial organisms, where habitat stability appears to be a major factor structuring genetic diversity (Carnaval *et al.*, 2009). Below we discuss the role of past riverine connections versus habitat stability in structuring genetic variation in riverine organisms, as well as the implications of our findings for identifying conservation areas for freshwater fishes.

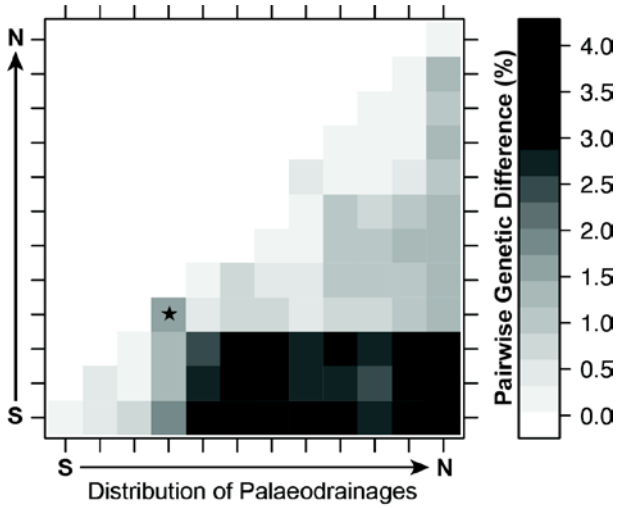


Figure II.5 – Percentage of pairwise genetic difference between palaeodrainages. Lighter colours indicate low genetic difference, while darker colours indicate larger genetic difference. Squares along the diagonal represent the average genetic distance within each palaeodrainage, with the comparison for the Paranaguá palaeodrainage (Paranaguá and Guaraqueçaba populations) marked by an asterisk.

Paleodrainages as key in structuring patterns of genetic variation

The role of paleodrainages in structuring the diversity of fishes has traditionally been inferred based on patterns of biodiversity, and in particular, on faunal boundaries (Weitzman *et al.*, 1988; Hubert & Renno, 2006; Abell *et al.*, 2008). The critical role of paleo-connections due to sea level retreat in structuring biodiversity patterns has been corroborated by recent inventories of global freshwater fish diversity and compositional similarity of coastal basins (Dias *et al.*, 2014). In contrast to such unambiguous results based on faunal surveys, results from molecular analyses have been more equivocal. For example, whether a study does or does not support the role of paleodrainages may simply be a consequence of the limitations of the methodology applied, as highlighted by differences in the apparent role of paleodrainages when studies are conducted at different geographic scales or in different regions of the world (Swartz *et al.*, 2007; Schultz *et al*., 2008; de Bruyn *et al*., 2013; Unmack *et al*., 2013).

Generally, congruence between monophyly and hypothesized historical riverine connections caused by sea level retreat has been interpreted as providing support for the role of paleodrainages in structuring diversity (de Bruyn *et al*., 2013; Unmack *et al*., 2013). However, interpreting the importance of paleodrainages relative to other factors based on visual inspection of phylogenetic trees is problematic. In particular, the lack of concordance might simply reflect the particular timing of divergence (especially for Pleistocenic events), and therefore, patterns of incomplete lineage sorting should not be directly interpreted as conflicting with hypothesized barriers (Knowles, 2009). Nevertheless, such patterns (e.g., when individuals collected from a paleodrainage do not form a clade) are often interpreted as being signatures of alternative mechanisms, such as temporary connections among populations related with drainage rearrangements (river captures) or dispersal across low-relief drainage divides (e.g., Unmack *et al*., 2013). Rather than relying on strict concordance between the geography and phylogenetic position of individuals, other approaches, such as the partitioning of genetic variation (as applied here; see also Swartz *et al*., 2007), can be used to test for genetic signature of past connections associated with paleodrainages, especially when divergences are sufficiently recent that there could still be incomplete lineage sorting. For example, a strict reliance on concordance for corroborating the role of paleodrainages in structuring variation would lead to a rejection of the role of paleodrainages in structuring genetic variation for the northern region (Figure II.2) because of shared haplotypes among paleodrainages (Figure II.1) and a lack of monophyly

(Figure A.1). However, the AMOVA indicates that the northern region, like the southern region, shows significant partitioning of genetic variation associated with the inferred paleodrainages (Table II.2).

While the results support a role for paleodrainages in structuring genetic variation across the entire distribution of *Hollandichthys multifasciatus*, additional factors might affect local population structure. For example, river captures between coastal and inland basins have been relatively well documented in the southeastern drainages of Brazil, including the river capture between Paraíba do Sul and Upper Tietê (Malabarba, 1998). However, these events predate the Pleistocene, in contrast to our genetic data, which supports population divergences that date to the Pleistocene – a period for which relatively few river captures have been documented (e.g., Upper Tietê/ Guaratuba and Iguaçú; Ribeiro 2006; Ribeiro *et al.*, 2006; Menezes *et al.*, 2008). Moreover, such river captures are limited to ichthyofaunal exchange between coastal and inland drainages, with geological evidence on their role among coastal drainages generally lacking. River captures may make contributions to patterns of divergence, as potentially indicated by local divergence patterns (e.g., the presence of a unique population of *Hollandichthys multifasciatus* in an inland drainage of the Upper Tietê). However, these events are not thought to be a general mechanism structuring population genetic variation across the species. In addition, our results demonstrate that emphasising current basin structure may be misleading, especially considering that isolation among rivers associated with current sea levels may be of less consequence given that their isolation may be relatively brief compared to the longer periods of connectivity during periods of low sea level that have dominated the Pleistocene (see Figure II.2c and Miller *et al.*, 2011).

The importance (or unimportance) of distributional stability

Ecological factors are usually invoked to explain the disagreement between genetic diversity patterns and paleodrainage configuration when the organism of interest is specialised to certain habitats (Schultz *et al*., 2008; de Bruyn *et al*., 2013). Differential effects of climatic changes on taxa are also predicted when the vagility and environmental requirements of species vary (e.g., Massatti & Knowles, 2014).

If forest instability had an effect on the persistence of forest-dependent riverine fish, we predicted that we would find an agreement between paleodrainages and genetic structure in the

stable, but not in the unstable areas; however, no significant difference in the genetic structuring was detected (Table II.1). While this result should be confirmed with multilocus data, it is unlikely that our analysis is problematic because of reliance on mtDNA. That is, there is no reason to suspect that the similarity in the structuring of genetic variation in stable versus unstable areas could be an artefact of mtDNA. Moreover, other studies based on mtDNA have supported a stability-extinction model for terrestrial forest-dependent organisms in the Atlantic Forest (for a summary see Martins, 2011). In addition, we cannot rule out the possibility that the Atlantic Forest extended into regions that were exposed by lower sea levels, because we could not include the coastal plain in our ENMs (as discussed in the methods). However, such regions would, by definition, represent unstable areas (i.e., not continuously occupied over time), and therefore would not contribute any bias to our tests on the importance of the stability of the Atlantic Forest.

What do our results imply about using the stability of the Atlantic Forest as a criterion for targeting conservation efforts, given that we do not detect any significant difference between regions of stability and instability with respect to the processes structuring genetic variation? Several aspects of our data emphasise long-term localised persistence of *H. multifasciatus*: a large percentage of genetic variation is explained by paleodrainages (Table II.2); genetic distances among populations are high (Figure II.5); population dynamics inferred from the Bayesian skyline plots are similar; and divergence times are estimated to predate the LGM (Figure II.2). In the absence of similar studies, it is not possible to conclude whether these patterns will hold across other freshwater organisms in the region. However, our results highlight the possibility that aquatic organisms, even habitat specialists such as *H. multifasciatus,* may be less reliant than their terrestrial counterparts on stable areas (e.g., Carnaval *et al*., 2009). This suggests that different local conservation efforts are needed to preserve the genetic diversity of aquatic organisms, compared to targeting only areas of habitat stability as potential sources of colonists, as is done for terrestrial organisms. It is also worth noting that the contribution of forest refugia in promoting diversification in the Atlantic Forest is still controversial and taxon-dependent, especially in the southern portion (Amaro *et al.*, 2012; Porto *et al.*, 2013; Carnaval *et al.*, 2014).

Processes structuring species diversity in coastal Brazil

Paleodrainage configuration has implications for interpreting general patterns of diversity in riverine organisms along the southeastern coast of Brazil. Although the region is characterised by markedly high levels of endemism, with up to 95% of freshwater fish species representing local endemics (Bizerril, 1994), few population studies have been conducted (Torres & Ribeiro, 2009; Pereira *et al.*, 2013) that provide evidence for the processes contributing to divergence in the region. Nevertheless, links between the cyclical changes in sea level during the Pleistocene and the high frequency of endemism have been proposed (Weitzman *et al.*, 1988). Our results lend strong support to this hypothesis. A general correspondence between some of the breaks observed in patterns of population genetic differentiation in *H. multifasciatus* with breaks in species distributions across the region suggest that processes structuring genetic variation might also parallel those structuring species diversity. For example, the southernmost paleodrainage is congruent with one proposed freshwater ecoregion, the Tramandaí-Mampituba ecoregion (Abell *et al.*, 2008), suggesting the barrier between paleodrainages drives local endemism. Indeed, the *H. multifasciatus* populations in this southernmost paleodrainage (Maquiné, Mampituba and Araranguá populations) have recently been identified as a putative new species of *Hollandichthys* (*H. taramandahy* Bertaco & Malabarba, 2013). Future investigations looking for correspondences between population and species-level divergence will be key to determining the extent to which the phenomena we document within *H. multifasciatus* also structure patterns of species diversification more broadly.

Summary

Our findings demonstrate the importance of contrasting alternative hypotheses to determine the relative contribution of past connections among currently isolated populations compared to that of habitat persistence over time (as recently emphasised by Thomé *et al.*, 2014). Our results demonstrate that habitat stability is less important in structuring genetic variation in the freshwater fish *H. multifasciatus*, and perhaps other aquatic organisms, than in terrestrial species. Proposals to conserve hotspots of diversity based on refugia models of the Atlantic Forest (Carnaval *et al*., 2009) are not appropriate for freshwater fishes. As an alternative, we propose the use of paleodrainages in conserving species and genetic diversity for freshwater organisms in coastal regions, especially given that the structuring of genetic variation in our

study is corroborated by faunal surveys supporting a role for paleodrainages in structuring freshwater fish biodiversity (Dias *et al.*, 2014). Our study also highlights the general utility of using detailed bathymetric maps (i.e., GEBCO) with GIS techniques (i.e., *Hydrological* tools in ARCGIS 10) to make inferences about past connections among coastal basins in regions, when geological information about paleochannels is unavailable.

References

Abell, R., Thieme, M.L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., Coad, B., Mandrak, N., Balderas, S.C., Bussing, W., Stiassny, M.L.J., Skelton, P., Allen, G.R., Unmack, P., Naseka, A., Ng, R., Sindorf, N., Robertson, J., Armijo, E., Higgins, J.V., Heibel, T.J., Wikramanayake, E., Olson, D., López, H.L., Reis, R.E., Lundberg, J.G., Sabaj, M.H., Petry, P. (2008) Freshwater ecoregions of the world: a new map of biogeographic units for freshwater biodiversity conservation. *BioScience*, **58**, 403-414.

Abilhoa, V., Bornatowski, H. & Otto, G. (2009) Temporal and ontogenetic variations in feeding habits of *Hollandichthys multifasciatus* (Teleostei: Characidae) in coastal Atlantic rainforest streams, southern Brazil. *Neotropical Ichthyology*, **7**, 415-420.

Amaro, R.C., Rodrigues, M. T., Yonenaga-Yassuda, Y. & Carnaval, A.C. (2012) Demographic processes in the montane Atlantic Rainforest: molecular and cytogenetic evidence from the endemic frog *Proceratophrys boiei*. *Molecular Phylogenetics and Evolution*, **62**, 880-888.

Anderson R.P. & Raza A. (2010) The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution : preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *Journal of Biogeography*, **37**, 1378–1393.

Behling, H. (2002). South and southeast Brazilian grasslands during Late Quaternary times: a synthesis. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **177**, 19-27.

Bermingham, E. & Avise, J.C. (1986) Molecular zoogeography of freshwater fishes in the Southeastern United States. *Genetics*, **113**, 939-965.

Bermingham E. & Martin A.P. (1998) Comparative mtDNA phylogeography of Neotropical freshwater fishes: testing shared history to infer the evolutionary landscape of lower Central America. *Molecular Ecology*, **7**, 499–517.

Bermingham, E., McCafferty, S.S. & Martin, A.P. (1997) Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus. *Molecular Systematics of Fishes*. (ed. by T.D. Kocher and C.A. Stepien), pp. 113-128. Academic Press, San Diego.

Bertaco, V.A., & Malabarba, L.R. (2013) A new species of the characid genus *Hollandichthys* Eigenmann from coastal rivers of southern Brazil (Teleostei: Characiformes) with a discussion on the diagnosis of the genus. *Neotropical Ichthyology*, **11**, 767-778.

Bizerril, C.R.S.F. (1994) Análise taxonômica e biogeográfica da ictiofauna de água doce do leste brasileiro. *Acta Biologica Leopoldensia*, **16**, 51-80.

de Bruyn, M., Rüber, L., Nylinder, S., Stelbrink, B., Lovejoy, N.R., Lavoué, S., Tab, H.H., Nugroho, E., Wowor, D., Ng, P.K.L., Azizah, M.N.S., Von Rintelen, T., Hall., R. & Carvalho, G. R. (2013) Palaeo-drainage basin connectivity predicts evolutionary relationships across three southeast Asian biodiversity hotspots. *Systematic Biology*, **62**, 398-410.

Burridge, C.P., Craw, D., Jack, D. C., King, T. M., & Waters, J. M. (2008). Does fish ecology predict dispersal across a river drainage divide? *Evolution*, **62**, 1484-1499.

Carnaval, A.C. & Moritz, C. (2008) Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography*, **35**, 1187–1201.

Carnaval, A.C., Hickerson, M.J., Haddad, C.F.B., Rodrigues, M.T. & Moritz, C. (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, **323**, 785–789.

Carnaval, A.C., Waltari, E., Rodrigues, M.T., Rosauer, D., VanDerWal, J., Damasceno, R., Prates, I., Strangas, M., Spanos, Z., Rivera, D., Pie, M.R., Firkowski, C.R., Bornschein, M.R., Ribeiro, L.F., Moritz, C. (2014) Prediction of phylogeographic endemism in an environmentally complex biome. *Proceeding of Royal Society B: Biological Sciences*, **281**, 2014.1461.

Carrea, C., Anderson, L.V., Craw, D., Waters J.M. & Burridge C.P. (2013) The significance of past interdrainage connectivity for studies of diversity, distribution and movement of freshwater-limited taxa within a catchment. *Journal of Biogeography*, **41**, 536–547.

Conti, L.A. (2009) Evidências da evolução dos sistemas de paleodrenagens na plataforma continental da região de São Sebastião (litoral norte do estado de São Paulo). *Revista Brasileira de Geomorfologia*, **10**, 45-55.

Cooke, G.M., Landguth, E.L. & Beheregaray, L.B. (2014) Riverscape genetics identifies replicated ecological divergence across an Amazonian ecotone. *Evolution*, **68**, 1947-1960.

Crispo, E., Bentzen, P., Reznick, D.N., Kinnison, M.T. & Hendry, A.P. (2006) The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology*, **15**, 49-62.

Dias, M.S., Oberdorff, T., Hugueny, B., Leprieur, F., Jézéquel, C., Cornu, J.F., Brosse, S., Grenouillet, G. & Tedesco, P.A. (2014) Global imprint of historical connectivity on freshwater fish biodiversity. *Ecology Letters*, **17**, 1130-1140.

Drummond, A.J., Rambaut, A., Shapiro, B. & Pybus, O.G. (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185-1192.

Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969-1973.

Dutton, A. & Lambeck, K. (2012) Ice volume and sea level during the last interglacial. *Science*, **337**, 216-219.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Esteves, K.E. & Lobón-Cerviá, J. (2001) Composition and trophic structure of a fish community of a clear water Atlantic rainforest stream in southeastern Brazil. *Environmental Biology of fishes*, **62**, 429-440.

Ewing, B., Hillier, L., Wendl, M.C. & Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, **8**, 175–185.

Excoffier, L. & Lischer, H.E.L. (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.

Excoffier, L., Smouse, P.E. & Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479-491.

Fleming, K., Johnston, P., Zwartz, D., Yokoyama, Y., Lambeck, K. & Chappell, J. (1998) Refining the eustatic sea-level curve since the Last Glacial Maximum using far- and intermediate-field sites. *Earth and Planetary Science Letters*, **163**, 327–342.

Fundação SOS Mata Atlântica (2011) *Atlas dos remanescentes florestais da Mata Atlântica - período 2008-2010.* Fundação SOS Mata Atlântica, São Paulo, Brasil.

Gordon, D., Abajian, C. & Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Research*, **8**, 195–202.

Grant, W.S. (2015) Problems and cautions with sequence mismatch analysis and Bayesian skyline plots to infer historical demography. *Journal of Heredity*, **106**, 333-346.

He, Q., Edwards, D.L. & Knowles, L.L. (2013) Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, **67**, 3386-3402.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Hubert, N. & Renno, J.F. (2006). Historical biogeography of South American freshwater fishes. *Journal of Biogeography*, **33**, 1414-1436.

Knowles, L.L. (2009) Statistical phylogeography. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 593-612.

Lanfear, R., Calcott, B., Ho, S.Y. & Guindon, S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695-1701.

Li, W.L.S. & Drummond, A.J. (2012) Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution*, **29**, 751-61.

Librado, P. & Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451-1452.

Martins, F.M. (2011) Historical biogeography of the Brazilian Atlantic forest and the Carnaval-Moritz model of Pleistocene refugia : what do phylogeographical studies tell us? *Biological Journal of the Linnean Society*, **104**, 499-509.

Massatti, R. & Knowles, L.L. (2014) Microhabitat differences impact phylogeographic concordance of codistributed species: genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. *Evolution*, **68**, 2833-2846.

Medrano, J.F., Aasen, E. & Sharrow, L. (1990) DNA extraction from nucleated red blood cells. *Biotechniques*, **8**, 43.

Menezes, N.A., Ribeiro, A.C., Weitzman, S. & Torres, R.A. (2008) Biogeography of Glandulocaudinae (Teleostei: Characiformes: Characidae) revisited: phylogenetic patterns, historical geology and genetic connectivity. *Zootaxa*, **1726**, 33-48.

Miller, K.G., Mountain, G.S., Wright, J.D. & Browning, J.V. (2011) A 180-million record of sea level and ice volume variations from continental margin and deep-sea isotopic records. *Oceanography*, **24**, 40-53.

Muneepeerakul, R., Weitz, J.S., Levin, S.A., Rinaldo, A. & Rodriguez-Iturbe, I. (2007) A neutral metapopulation model of biodiversity in river networks. *Journal of Theoretical Biology*, **245**, 351-363.

Neuenschwander S., Largiadèr C.R., Ray N., Currat M., Vonlanthen P. & Excoffier L. (2008) Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Molecular ecology*, **17**, 757–72.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., & Kassem, K.R. (2001) Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, **51**, 933-938.

Ornelas-García, C.P., Domínguez-Domínguez, O. & Doadrio, I. (2008) Evolutionary history of the fish genus *Astyanax* Baird & Girard (1854) (Actinopterygii, Characidae) in Mesoamerica reveals multiple morphological homoplasies. *BMC Evolutionary Biology*, **8**, 340-357.

Papadopoulou, A., Anastasiou, I., Keskin, B., & Vogler, A. P. (2009) Comparative phylogeography of tenebrionid beetles in the Aegean archipelago: the effect of dispersal ability and habitat preference. *Molecular Ecology*, **18**, 2503-2517.

Pereira, T.L., Santos, U., Schaefer, C.E., Souza, G.O., Paiva, S.R., Malabarba, L.R., Schmidt, E.E. & Dergam, J.A. (2013) Dispersal and vicariance of *Hoplias malabaricus* (Bloch, 1794) (Teleostei, Erythrinidae) populations of the Brazilian continental margin. *Journal of Biogeography*, **40**, 905-914.

Phillips S.J., Anderson R.P. & Schapire R.E. (2006) Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Porto, T.J., Carnaval, A.C. & da Rocha, P.L.B. (2013) Evaluating forest refugial models using species distribution models, model filling and inclusion: a case study with 14 Brazilian species. *Diversity and Distributions*, **19**, 330-340.

Ribeiro, A.C. (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.

Ribeiro, A.C., Lima, F.C., Riccomini, C. & Menezes, N.A. (2006). Fishes of the Atlantic Rainforest of Boracéia: testimonies of the Quaternary fault reactivation within a Neoproterozoic tectonic province in Southeastern Brazil. *Ichthyological Exploration of Freshwaters*, **17**, 157-164.

Roeckner, E., Arpe, K., Bengtsson, L., Brinkop, S., Duemenil, L., Esch, M., Kirk, E., Lunkeit, F., Ponater, M., Rockel, B., Sausen, R., Schlese, U., Schubert, S. & Windelband, M. (1992) *Simulation of the present-day climate with the ECHAM model: impact of model physics and resolution*. pp. 171. Max-Planck-Institut fuer Meteorologie, Hamburg, Germany.

Schultz, M.B., Ierodiaconou, D.A., Smith, S. A., Horwitz, P., Richardson, A. M., Crandall, K.A. & Austin, C.M. (2008) Sea-level changes and palaeo-ranges: reconstruction of ancient shorelines and river drainages and the phylogeography of the Australian land crayfish *Engaeus sericatus* Clark (Decapoda: Parastacidae). *Molecular Ecology*, **17**, 5291-5314.

Smith, D.E., Harrison, S., Firth, C.R. & Jordan, J.T. (2011) The early Holocene sea level rise. *Quaternary Science Reviews*, **30**, 1846-1860.

Soltis, D.E., Morris, A.B., McLachlan, J.S., Manos, P.S. & Soltis, P.S. (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261-4293.

Strecker, U., Faúndez, V.H. & Wilkens, H. (2004) Phylogeography of surface and cave *Astyanax* (Teleostei) from Central and North America based on cytochrome b sequence data. *Molecular Phylogenetics and Evolution*, **33**, 469-481.

Swartz, E.R., Skelton, P.H. & Bloomer, P. (2007) Sea-level changes, river capture and the evolution of populations of the Eastern Cape and fiery redfins (*Pseudobarbus afer* and *Pseudobarbus phlegethon*, Cyprinidae) across multiple river systems in South Africa. *Journal of Biogeography*, **34**, 2086-2099.

Thomaz, A.T., Malabarba, L.R. & Bonatto, S.L. (2010) The phylogenetic placement of *Hollandichthys* Eigenmann 1909 (Teleostei: Characidae) and related genera. *Molecular Phylogenetics and Evolution*, **57**, 1347-1352.

Thomé, M.T.C., Zamudio, K.R., Haddad, C.F. & Alexandrino, J. (2014) Barriers, rather than refugia, underlie the origin of diversity in toads endemic to the Brazilian Atlantic Forest. *Molecular Ecology*, **23**, 6152-6164.

Torres, R.A. & Ribeiro, J. (2009) The remarkable species complex *Mimagoniates microlepis* (Characiformes: Glandulocaudinae) from the Southern Atlantic Rain forest (Brazil) as revealed by molecular systematic and population genetic analyses. *Hydrobiologia*, **617**, 157-170.

Unmack P.J., Hammer M.P., Adams M. & Johnson J.B. (2013) The role of continental shelf width in determining freshwater phylogeographic patterns in south-eastern Australian pygmy perches (Teleostei: Percichthyidae). *Molecular Ecology*, **22**, 1683–1699.

Weitzman S.H., Menezes N.A. & Weitzman M.J. (1988) Phylogenetic biogeography of the Glandulocaudini (Teleostei: Characiformes, Characidae) with comments on the distributions of other freshwater fishes in eastern and southeastern Brazil. *Proceedings of a workshop on neotropical distribution patterns. Academia Brasileira de Ciências.* pp. 379–427. Academia Brasileira de Ciências, Rio de Janeiro.

Whiteley, A.R., Spruell, P. & Allendorf, F.W. (2004) Ecological and life history characteristics predict population genetic divergence of two salmonids in the same landscape. *Molecular Ecology*, **13**, 3675-3688.

Zardoya, R. & Doadrio, I. (1999) Molecular evidence on the evolutionary and biogeographical patterns of European Cyprinids. *Jounal of Molecular Evolution*, **49**, 227-237.

APPENDIX A

Supplementary material from Chapter II

Table A.1 - Table with detailed sample information (haplotype, locality, description), including the exact coordinates of the localities where fish were sampled per population

Population (Drainage)	Voucher	Lat.	Long.	Hap.	GenBank Accession	
					COI	ND2
Paraty 1	UFRGS 11776/TEC909A	-23.04	-44.60	69	JF836527	JF836701
	UFRGS 11776/TEC909B	-23.04	-44.60	70	JF836528	JF836702
	UFRGS 11776/TEC909C	-23.04	-44.60	70	JF836529	JF836703
	UFRGS 11776/TEC909D	-23.04	-44.60	70	HM562850	HM562881
	UFRGS 11776/TEC909F	-23.04	-44.60	70	JF836531	JF836704
	UFRGS 11776/TEC909G	-23.04	-44.60	70	JF836532	JF836705
Paraty 2	MCP 30665/174	-23.08	-44.70	67	JF836408	JF836591
	MCP 30666/175	-23.04	-44.69	68	JF836409	JF836592
Toca do Boi	MCP 30664/387	-23.33	-44.68	64	JF836414	JF836597
	MCP 30664/388	-23.33	-44.68	64	JF836415	JF836598
	MCP 30664/389	-23.33	-44.68	65	JF836416	JF836599
	UFRGS 11775/TEC908A	-23.33	-44.68	66	JF836519	JF836696
	UFRGS 11775/TEC908B	-23.33	-44.68	64	JF836520	JF836697
	UFRGS 11775/TEC908C	-23.33	-44.68	66	JF836521	JF836698
	UFRGS 11775/TEC908D	-23.33	-44.68	66	JF836522	JF836699
	UFRGS 11775/TEC908E	-23.33	-44.68	64	JF836523	JF836700
Ubatuba 1	UFRGS 11789/TEC866A	-23.35	-44.87	63	JF836517	-
	UFRGS 11789/TEC866B	-23.35	-44.87	45	JF836518	JF836695
Ubatuba 2	MCP 30663/170	-23.43	-45.13	60	JF836407	JF836590
	UFRGS 11774/TEC900A	-23.41	-45.11	61	JF836515	JF836693
	UFRGS 11774/TEC900B	-23.41	-45.11	62	JF836516	JF836694
Ilhabela	MCP 30661/358	-23.83	-45.36	57	JF836410	JF836593
	MCP 30661/359	-23.83	-45.36	57	JF836411	JF836594
	MCP 30662/383	-23.82	-45.36	57	JF836412	JF836595
	MCP 30662/384	-23.82	-45.36	57	JF836413	JF836596
	UFRGS 11773/TEC897A	-23.82	-45.35	58	JF836511	JF836689
	UFRGS 11773/TEC897B	-23.82	-45.35	59	JF836512	JF836690
	UFRGS 11773/TEC897C	-23.82	-45.35	57	HM562849	HM562880
	UFRGS 11773/TEC897D	-23.82	-45.35	57	JF836513	JF836691
	UFRGS 11773/TEC897E	-23.82	-45.35	57	JF836514	JF836692
São Sebastião 1	UFRGS 11788/TEC859A	-23.82	-45.45	55	HM562848	HM562879
	UFRGS 11788/TEC859B	-23.82	-45.45	56	JF836510	-
São Sebastião 2	UFRGS 11795/TEC1229A	-23.77	-45.61	49	JF836502	JF836681
	UFRGS 11795/TEC1229B	-23.77	-45.61	52	JF836503	JF836682
	UFRGS 11795/TEC1229C	-23.77	-45.61	49	JF836504	JF836683
	UFRGS 11795/TEC1229D	-23.77	-45.61	53	JF836505	JF836684
	UFRGS 11795/TEC1229E	-23.77	-45.61	49	JF836506	JF836685
	UFRGS 11795/TEC1229F	-23.77	-45.61	54	JF836507	JF836686
	UFRGS 11795/TEC1229G	-23.77	-45.61	52	JF836508	JF836687
	UFRGS 11795/TEC1229H	-23.77	-45.61	51	JF836509	JF836688
São Sebastião 3	MCP 30658/164	-23.76	-45.71	49	JF836405	JF836587
	UFRGS 11772/TEC891A	-23.76	-45.72	49	JF836494	JF836673
	UFRGS 11772/TEC891B	-23.76	-45.72	49	JF836495	JF836674

	UFRGS 11772/TEC891C	-23.76	-45.72	49	JF836496	JF836675
	UFRGS 11772/TEC891D	-23.76	-45.72	50	JF836497	JF836676
	UFRGS 11772/TEC891E	-23.76	-45.72	49	JF836498	JF836677
	UFRGS 11772/TEC891F	-23.76	-45.72	49	JF836499	JF836678
	UFRGS 11772/TEC891G	-23.76	-45.72	49	JF836500	JF836679
	UFRGS 11772/TEC891H	-23.76	-45.72	51	JF836501	JF836680
Bertioga	UFRGS 11771/TEC888A	-23.78	-46.00	45	HM562847	HM562878
	UFRGS 11771/TEC888B	-23.78	-46.00	44	JF836480	JF836658
	UFRGS 11771/TEC888C	-23.78	-46.00	44	JF836481	JF836659
	UFRGS 11771/TEC888D	-23.78	-46.00	44	JF836482	JF836660
	UFRGS 11771/TEC888E	-23.78	-46.00	44	JF836483	JF836661
	UFRGS 11771/TEC888F	-23.78	-46.00	44	JF836484	JF836662
	UFRGS 11771/TEC888G	-23.78	-46.00	46	JF836485	JF836663
	UFRGS 11771/TEC888H	-23.78	-46.00	44	JF836486	JF836664
	UFRGS 11777/TEC980A	-23.78	-46.00	44	JF836487	JF836665
	UFRGS 11777/TEC980B	-23.78	-46.00	47	JF836488	JF836666
	UFRGS 11777/TEC980C	-23.78	-46.00	44	JF836489	JF836667
	UFRGS 11777/TEC980D	-23.78	-46.00	44	JF836490	JF836668
	UFRGS 11777/TEC980E	-23.78	-46.00	44	JF836491	JF836669
	UFRGS 11777/TEC980F	-23.78	-46.00	48	JF836492	JF836670
	UFRGS 11777/TEC980G	-23.78	-46.00	44	JF836493	JF836671
Upper Tietê *	UFRGS 11786/TEC826A	-23.77	-46.33	44	JF836465	JF836645
	UFRGS 11786/TEC826B	-23.77	-46.33	44	JF836466	JF836646
	UFRGS 11786/TEC826C	-23.77	-46.33	44	JF836467	JF836647
	UFRGS 11786/TEC826D	-23.77	-46.33	44	JF836468	JF836648
	UFRGS 11786/TEC826E	-23.77	-46.33	44	JF836469	JF836649
	UFRGS 11786/TEC826F	-23.77	-46.33	44	JF836470	JF836650
	UFRGS 11786/TEC826G	-23.77	-46.33	44	JF836471	JF836651
	UFRGS 11786/TEC826H	-23.77	-46.33	44	JF836472	JF836652
	UFRGS 11787/TEC827A	-23.77	-46.31	44	JF836473	JF836653
	UFRGS 11787/TEC827B	-23.77	-46.31	44	JF836474	JF836654
	UFRGS 11787/TEC827C	-23.77	-46.31	44	HM562846	HM562877
	UFRGS 11787/TEC827E	-23.77	-46.31	44	JF836476	JF836655
	UFRGS 11787/TEC827F	-23.77	-46.31	44	JF836477	JF836656
	UFRGS 11787/TEC827H	-23.77	-46.31	44	JF836479	JF836657
Santos	MCP 30560/290	-23.84	-46.33	41	JF836399	JF836581
	MCP 30560/291	-23.84	-46.33	42	JF836400	JF836582
	MCP 30559/341	-23.86	-46.35	42	JF836401	JF836583
	MCP 30559/342	-23.86	-46.35	42	JF836402	JF836584
	MCP 30559/343	-23.86	-46.35	42	JF836403	JF836585
	MCP 30559/344	-23.86	-46.35	43	JF836404	JF836586
	UFRGS 11785/TEC820A	-23.86	-46.35	44	JF836459	JF836642
	UFRGS 11785/TEC820C	-23.86	-46.35	44	JF836461	JF836643
	UFRGS 11785/TEC820E	-23.86	-46.35	42	HM562845	HM562876
	UFRGS 11785/TEC820G	-23.86	-46.35	42	JF836464	JF836644
Peruíbe	MCP 30554/276	-24.36	-47.04	40	JF836396	JF836578
	MCP 30555/294	-24.37	-47.05	40	JF836397	JF836579
	MCP 30561/316	-24.36	-47.04	40	JF836398	JF836580
	UFRGS 11783/TEC817B	-24.36	-47.04	40	JF836448	JF836632
	UFRGS 11783/TEC817C	-24.36	-47.04	40	JF836449	JF836633
	UFRGS 11783/TEC817D	-24.36	-47.04	40	JF836450	JF836634
	UFRGS 11783/TEC817E	-24.36	-47.04	40	JF836451	JF836635
	UFRGS 11784/TEC 722A	-24.37	-47.06	40	JF836452	JF836636
	UFRGS 11784/TEC 722B	-24.37	-47.06	40	JF836453	JF836637
	UFRGS 11784/TEC 722C	-24.37	-47.06	40	JF836454	JF836638
	UFRGS 11784/TEC 722D	-24.37	-47.06	40	JF836455	JF836639
	UFRGS 11784/TEC 722E	-24.37	-47.06	40	JF836456	JF836640
	UFRGS 11784/TEC 722G	-24.37	-47.06	40	JF836458	JF836641

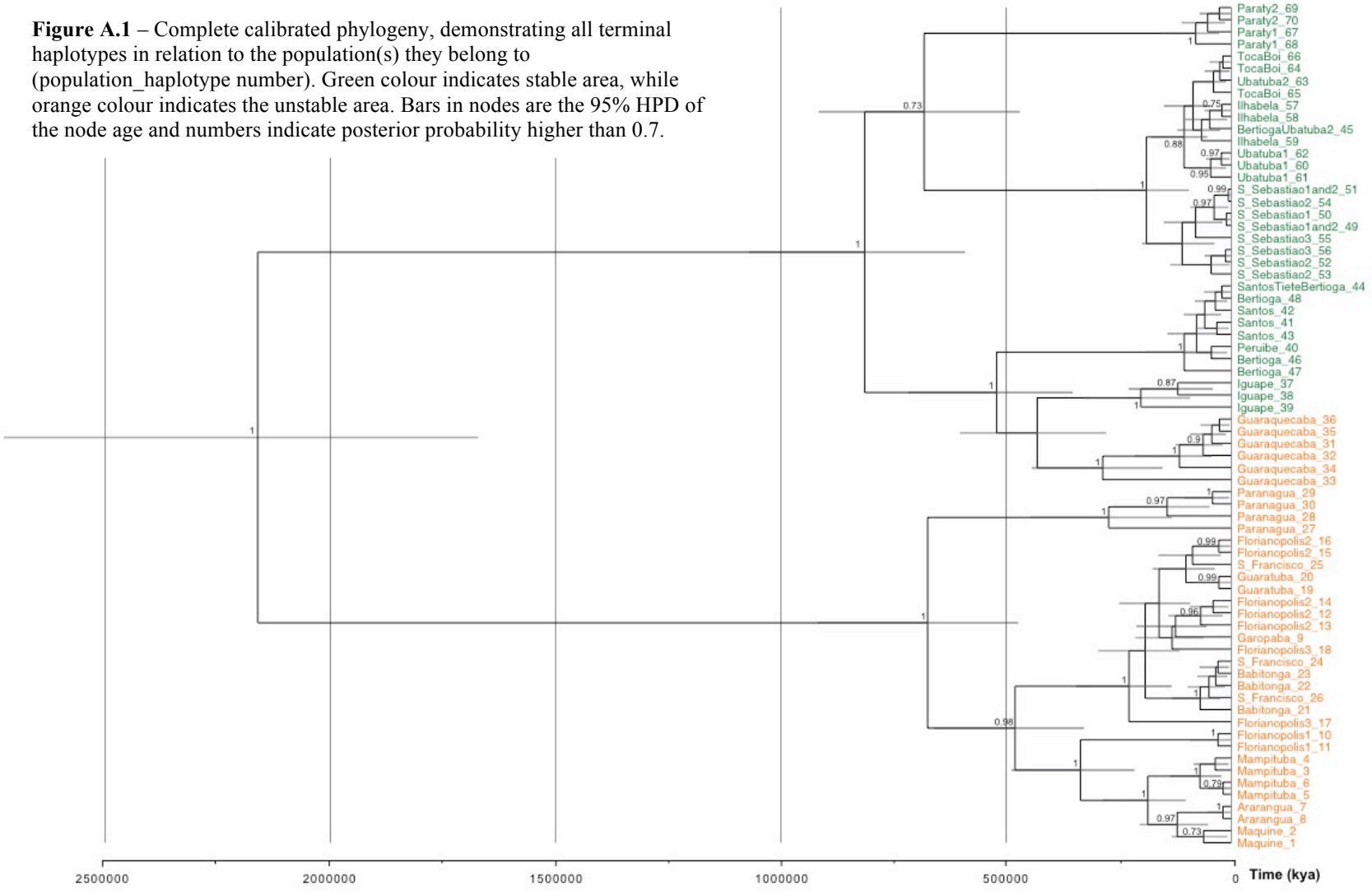
	UFRGS 11784/TEC 722H	-24.37	-47.06	40	HM562844	HM562875
Ribeira de Iguape	UFRGS 11781/TEC806A	-24.66	-47.49	37	JF836443	JF836626
	UFRGS 11781/TEC806B	-24.66	-47.49	38	JF836444	JF836627
	UFRGS 11781/TEC806C	-24.66	-47.49	37	JF836445	JF836628
	UFRGS 11781/TEC806D	-24.66	-47.49	39	JF836446	JF836629
	UFRGS 11782/TEC739A	-24.65	-47.48	37	HM562843	HM562874
	UFRGS 11782/TEC739B	-24.65	-47.48	37	JF836447	JF836630
Guaraqueçaba	MCP 30557/162	-25.17	-48.42	31	HM562840	HM562871
	MCP 30558/163	-25.21	-48.43	32	JF836395	JF836577
	UFRGS 11778/TEC755A	-25.22	-48.45	33	JF836433	JF836616
	UFRGS 11778/TEC755B	-25.22	-48.45	34	JF836434	JF836617
	UFRGS 11778/TEC755C	-25.22	-48.45	33	JF836435	JF836618
	UFRGS 11778/TEC755D	-25.22	-48.45	35	JF836436	JF836619
	UFRGS 11778/TEC755E	-25.22	-48.45	35	JF836437	JF836620
	UFRGS 11779/TEC763A	-25.17	-48.42	31	JF836438	JF836621
	UFRGS 11779/TEC763B	-25.17	-48.42	32	JF836439	JF836622
	UFRGS 11779/TEC763C	-25.17	-48.42	32	JF836440	JF836623
	UFRGS 11779/TEC763D	-25.17	-48.42	36	JF836441	JF836624
Paranaguá	MCP 30556/161	-25.40	-48.87	27	JF836388	JF836569
	MHNCI/1	-25.43	-48.70	28	JF836391	JF836572
	MHNCI/3	-25.43	-48.70	29	JF836392	JF836574
	MHNCI/4	-25.43	-48.70	28	JF836393	JF836575
	MHNCI/5	-25.43	-48.70	30	JF836394	JF836576
	MHNCI/6	-25.43	-48.70	30	HM562839	HM562870
São Francisco do Sul	MCP 30553/159	-26.33	-48.65	24	JF836387	JF836568
	UFRGS 10579/TEC105	-26.33	-48.65	25	HM562842	HM562873
	UFRGS 9359/TEC345	-26.33	-48.65	26	JF836430	JF836613
Babitonga	MCP 30552/158	-26.38	-48.73	21	JF836386	JF836567
	MCP 30667/176	-26.19	-48.93	22	JF836389	JF836570
	MCP 31486/177	-26.19	-48.93	23	JF836390	JF836571
Guaratuba	UFRGS 10578/TEC109	-25.88	-48.58	19	JF836431	JF836614
	UFRGS 9358/TEC346	-25.88	-48.58	20	JF836432	JF836615
Florianópolis 1	MCP 28737/100	-27.59	-48.48	10	JF836379	JF836560
	MCP 28737/101	-27.59	-48.48	11	JF836380	JF836561
	MCP 28737/102	-27.59	-48.48	10	JF836381	JF836562
Florianópolis 2	MCP 28747/104	-27.51	-48.49	12	JF836382	JF836563
	MCP 28732/107	-27.51	-48.49	12	JF836383	JF836564
	MCP 28732/125	-27.51	-48.49	13	JF836385	JF836566
	MCP 38333/A	-27.52	-48.47	12	JF836419	JF836602
	MCP 38333/B	-27.52	-48.47	12	JF836420	JF836603
	MCP 38333/C	-27.52	-48.47	12	HM562841	HM562872
	MCP 38333/D	-27.52	-48.47	12	JF836421	JF836604
	MCP 38333/E	-27.52	-48.47	14	JF836422	JF836605
	MCP 38317/A	-27.48	-48.44	15	JF836423	JF836606
	MCP 38317/B	-27.48	-48.44	15	JF836424	JF836607
	MCP 38317/C	-27.48	-48.44	16	JF836425	JF836608
Florianópolis 3	MCP 37654/A	-27.46	-48.42	17	JF836417	JF836600
	MCP 37654/B	-27.46	-48.42	18	JF836418	JF836601
	MCP 37635/A	-27.48	-48.42	17	JF836426	JF836609
	MCP 37635/B	-27.48	-48.42	17	JF836427	JF836610
	MCP 37635/C	-27.48	-48.42	17	JF836428	JF836611
	MCP 37635/D	-27.48	-48.42	18	JF836429	JF836612
Garopaba	MCP 28734/108	-28.07	-48.70	9	JF836384	JF836565
Araranguá	UFRGS 11791/TEC840A	-28.79	-49.93	7	JF836534	JF836708
	UFRGS 11791/TEC840B	-28.79	-49.93	8	JF836535	JF836709
	UFRGS 11791/TEC840C	-28.79	-49.93	8	JF836536	JF836710
	UFRGS 11791/TEC840D	-28.79	-49.93	8	JF836537	JF836711
Mampituba	MCP 29241	-29.17	-49.99	3	JF836370	JF836551

	MCP 29242	-29.17	-50.00	3	JF836371	JF836552
	MCP 23625	-29.17	-50.00	4	JF836378	JF836559
	UFRGS 11790/TEC839	-29.17	-49.98	3	JF836533	JF836707
	UFRGS 11792/TEC841A	-29.25	-50.12	5	JF836538	JF836712
	UFRGS 11792/TEC841B	-29.25	-50.12	3	HM562851	HM562882
	UFRGS 11792/TEC841C	-29.25	-50.12	3	JF836539	JF836713
	UFRGS 11792/TEC841D	-29.25	-50.12	3	JF836540	JF836714
	UFRGS 11792/TEC841E	-29.25	-50.12	3	JF836541	JF836715
	UFRGS 11792/TEC841F	-29.25	-50.12	6	JF836542	JF836716
	UFRGS 11792/TEC841G	-29.25	-50.12	6	JF836543	JF836717
	UFRGS 11792/TEC841H	-29.25	-50.12	3	JF836544	JF836718
Maquiné	MCP 29244/153	-29.43	-50.18	1	JF836372	JF836553
	MCP 26969/38	-29.58	-50.28	2	JF836373	JF836554
	MCP 26969/39	-29.58	-50.28	2	JF836374	JF836555
	MCP 26969/40	-29.58	-50.28	2	JF836375	JF836556
	MCP 26969/41	-29.58	-50.28	2	JF836376	JF836557
	MCP 26969/42	-29.58	-50.28	2	JF836377	JF836558
	UFRGS 11793/TEC842B	-29.59	-50.29	2	JF836545	JF836720
	UFRGS 11793/TEC842C	-29.59	-50.29	2	JF836546	JF836721
	UFRGS 11793/TEC842D	-29.59	-50.29	2	JF836547	JF836722
	UFRGS 11793/TEC842E	-29.59	-50.29	2	HM562852	HM562883
	UFRGS 11793/TEC842F	-29.59	-50.29	2	JF836548	JF836723
	UFRGS 11793/TEC842G	-29.59	-50.29	2	JF836549	JF836724

Table A.2 – Summary AMOVA result for unstable region after excluding the paleodrainage with Paranaguá and Guaraqueçaba populations due to the big phylogenetic break observed between these populations.

Scenario	**Source**	**% var.**
Unstable region	Among groups (Fct)	53.68 (p-value < 0.05)
	Among pops. within groups (Fsc)	36.69 (p-value ≅ 0)
	Within pops. (Fst)	9.63 (p-value ≅ 0)

Figure A.1 – Complete calibrated phylogeny, demonstrating all terminal haplotypes in relation to the population(s) they belong to (population_haplotype number). Green colour indicates stable area, while orange colour indicates the unstable area. Bars in nodes are the 95% HPD of the node age and numbers indicate posterior probability higher than 0.7.



CHAPTER III

The architecture of river networks can drive the evolutionary dynamics of aquatic populations.

Andréa T. Thomaz, Mark R. Christie, L. Lacey Knowles

Abstract

It is widely recognized that physical landscapes can shape genetic variation within and between populations. However, it is not well understood how riverscapes, with their complex architectures, affect patterns of neutral genetic diversity. Using a spatially explicit agent-based modeling (ABM) approach, we evaluate the genetic consequences of dendritic river shapes on local population structure. We disentangle the relative contribution of specific river properties to observed patterns of genetic variation by evaluating how different branching architectures and downstream flow regimes affect the genetic structure of populations situated within river networks. Irrespective of the river length, our results illustrate that the extent of river branching, confluence position, and levels of asymmetric downstream migration dictate patterns of genetic variation in riverine populations. Indeed, comparisons between simple and highly branched rivers show a 20-fold increase in the overall genetic diversity and a 7-fold increase in the genetic differentiation between local populations. Given that most rivers have complex architectures, these results highlight the importance of incorporating riverscape information into evolutionary models of aquatic species and could help explain why riverine fishes represent a disproportionately large amount of global vertebrate diversity per unit of habitable area.

Introduction

Understanding how the physical characteristics of an ecosystem affect patterns of genetic variation can enhance our ability to address a variety of evolutionary questions. River networks, for example, can have complex architectures that may drive patterns of neutral genetic variation

among or within river basins (e.g., Burridge et al. 2008; Hopken et al. 2013; Paz-Vinas and Blanchet 2015; Thomaz et al. 2015). Yet, without knowing which factors contribute to population structure (e.g., the effects of the number of tributaries, river length, or differences in flow regime), it isn't clear why genetic structure may differ across regions and/or taxa. Likewise, misleading inferences about putatively selected loci or purported genomic regions associated with local adaptation can result when an inappropriate null model is used to generate expected patterns of variation for neutral loci for tests of selection (Cruickshank and Hahn 2014; Lotterhos and Whitlock 2015; He and Knowles, 2016). For example, outlier approaches for detecting selected loci are prone to errors if the differences between riverine systems and their terrestrial counterparts are not taken into account (Fourcade et al. 2013).

Here we take an agent-based modeling (ABM) approach to disentangle the relative contributions of different river properties to observed patterns of genetic variation. Specifically, we evaluate how the dendritic nature of river networks (e.g., different levels of branching and confluence position) and constraints on dispersal associated with downstream flow regimes affect the genetic diversity of localized populations within rivers, as well as genetic differentiation along river segments. The contribution of asymmetric migration in rivers (Morrissey and Kerckhove 2009; Paz-Vinas et al. 2013) and the role of connectivity (Labonne et al. 2008) at the metapopulation level (i.e., in the whole river) has previously been demonstrated to be an important factor for the maintenance of genetic variation, but how the interaction between these variables and other important river properties generates local population structure remains poorly understand. We also contrast riverscapes with open landscapes (e.g., terrestrial landscapes) from a local population perspective in order to highlight key river parameters for modeling expected patterns of genetic variation and differentiation. With the insights the simulations provide about the causal factors structuring genetic variation, we suggest particular mechanisms that might be worth pursuing as possible agents for explaining patterns of species diversity itself in these aquatic habitats. Specifically, we discuss how our findings could help explain why riverine freshwater fishes represent a disproportionate amount of freshwater fish diversity, and total fish diversity more generally, given that rivers represent a minute fraction of habitable aquatic area globally (Lundberg et al. 2000; Guinot and Cavin 2015).

Material and Methods

Agent-based modeling in rivers

To understand how different river shapes affect patterns of genetic diversity within local populations and genetic variation between local populations, we employed a spatial explicit forward-time agent-based model (ABM) (see Christie and Knowles 2015). To test the effect of riverscape properties on population genetic structure, we measured both local genetic diversity within populations and genetic differentiation among populations along a riverscape after varying three explanatory variables: 1. the proportion of downstream migrants (from bidirectional to entirely asymmetrical migration from upstream to downstream), 2. river complexity (from linear to highly dendritic), and 3. the position of local populations within the riverscape (upstream and downstream populations, or confluence populations). We refer to upstream and downstream populations with respect to the direction of water flow, as well as those in the confluence (i.e., populations where two river segments come together; Figure III.1). We also use river order as a measure of network complexity (also called Horton-Strahler number; Horton 1945; Strahler 1957), where river order is defined as the number of upstream tributaries.

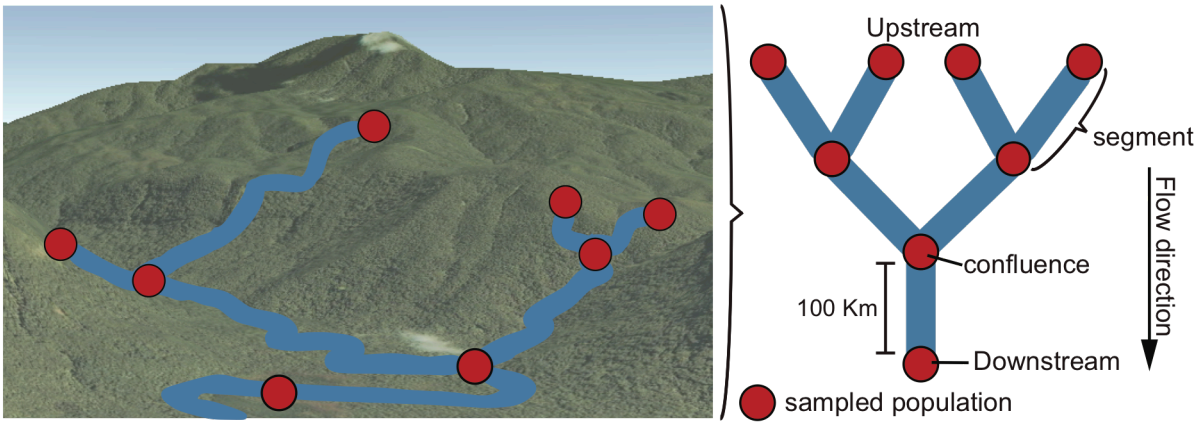


Figure III.1- Schema illustrating a hypothetical river network and the corresponding theoretical representation used in our model. In all simulations performed here, segments (blue lines) denote connection routes between local populations (red circles), with populations equally spaced along the river. Migration was varied according flow direction from symmetric to completely asymmetric downstream (black arrow).

In the ABM, individuals and their genotypes are tracked through a river network in both space and time (i.e., per generation), where each generation is characterized by reproduction, migration, mortality (as a function of user-defined carrying capacities), and mutation. Local

population sizes were constrained by a carrying capacity (K) such that the total number of individuals in the entire network depended on the total number of local populations distributed throughout the network. Simulations were conducted with three different carrying capacities (K = 100, 200, and 1000) and numbers of offspring per individual (N = 2, 10, and 20). To consider the generalizability of the results on the contribution of river architecture when other factors vary (e.g., when fish are not evenly distributed throughout a river network), we also conducted simulations in which the population size varied systematically from large upstream populations to small downstream populations. We modeled asexual reproduction because sexual reproduction would introduce a species-specific behavioral component to the model (i.e., mate choice, maximum distance between mates, etc.) that would make it difficult to identify generalizations about how the properties of rivers contribute to genetic structure. Dispersal was modeled using a leptokurtic dispersal kernel with the package Fishmove in R (Radinger and Wolter 2013), which provides a taxon-specific probability distribution of dispersed individuals as a function of distance from the source population for a given set of environmental characteristics. The taxon characteristics chosen for these simulations corresponded to a small generalist fish (size = 60mm standard length, caudal-fin aspect ratio = 1.5), with a dispersal probability of 0.095 under a lognormal distribution (i.e., dispersal in the Fishmove R package was set to river order three), where the dispersal probability of individuals was integrated over a one-year period. We did not vary the dispersal kernel based on the location of the populations (i.e., different migration rates for different river sections; Figure III.1) because we did not want to confound the interpretation of network shape with migration rate. The total number of migrants per generation was calculated as a function of K, the dispersal kernel, and the distance between populations. Following the dispersal step in each generation, individuals were randomly removed (introducing an expected variance in reproductive success) from each local population according to the local population carrying capacity K. Lastly, a per locus mutation rate (10^{-6}) was applied under an infinite alleles model, prior to the start of a new generation.

For all simulations reported here, models were run for 500 generations with 30 independent replicates for each combination of river properties (i.e., direction of migration and level of river complexity). Models were run for 500 generations with a relatively high mutation rate in order to strike a balance between relevant timescales and a computational processing time, throughout which we tracked changes in the allele frequencies (see Supplementary Text in

Appendix B.1 for a detailed discussion on parameter settings). At the beginning of each simulation, 50 bi-allelic codominant SNPs were created in Hardy-Weinberg equilibrium and genotype frequencies were distributed equally throughout the populations.

The output from a single simulation consisted of multi-locus genotypes for all individuals from every local population. We present the number of distinct multi-locus genotypes per population (averaged across replicates) for the entire river, as well as the pairwise population F_{ST}'s calculated with the package hierfstat in R (Goudet 2005). Patterns of genetic differentiation were visualized with a Principal Coordinate Analysis, PCoA, of the pairwise F_{ST} matrix using the function *pcoa* in the vegan package (Oksanen et al. 2013). For graphical illustration of patterns of genetic variation across populations, the first two axes of the PCoA were rescaled to range between 0 and 1 with a corresponding color score (Red Green Blue = RGB) that also ranged between 0 and 1. The first axis of the PCoA was represented by Red color variation and the second axis by the Green (Blue color variation was held constant at 0.8) such that the more dissimilar the colors are between two populations, the higher the F_{ST} value is between these populations. All models and analyses were performed with R version 3.1.2 (R Core Team, 2014; scripts are available at Dryad doi: 10.5061/dryad.7rs2d).

Landscapes vs. Riverscapes

To assess how river networks differ from open landscapes (e.g., some terrestrial and marine environments) a hypothetical landscape scenario was generated to contrast with the results from the riverscapes. For this comparison, river networks and open landscapes each had 16 populations with identical numbers of individuals, with populations being separated by a constant segment length (67 Km). While the open landscape was organized as a 4x4 matrix, connected vertically and horizontally (Figure III.2A), the river network resembled a medium complexity river (branching = 4^{th} order river; Figure III.2B). For the river network, we used asymmetric migration because it is a more realistic migration in this environment (9:1 ratio for downstream:upstream migrants; see Morrissey and Kerckhove 2009; Paz-Vinas et al. 2015). For the open landscape, we included two different migration scenarios: 1. asymmetric migration rate (9:1 ratio; see black arrows in Figure III.2A), and 2. symmetric migration where the same number of individuals migrated between two populations. Thus, in the first set of comparisons the open landscape differed from the river network only by the number of connections between

local populations while in the second set of comparisons we varied both the number of connections between local populations and the symmetry of migration (two factors that likely differ between terrestrial and riverine populations). Across all comparisons, the same total number of migrants was exchanged in each generation.

Riverscape architectures

To assess the effect of river architecture on the evolutionary dynamics of local populations we focused on three spatially important variables: 1. asymmetric migration, 2. the position of a confluence, and 3. the degree of branching (see Figure III.1). To assess the effect of asymmetric migration caused by downstream water flow on different river architectures, we compared two river networks: a low complexity river with a strictly linear shape, and a higher complexity river with a dendritic shape (branching = 4^{th} order river; Figure III.3). Both networks had 16 populations equally spaced, with constant total length of 1,000 km (67 km per segment). For each river network, we varied the frequency of downstream migration from completely symmetric (0.5) to completely asymmetric downstream (1.0) in increments of 0.1. Even though the proportion of downstream migrants was varied, the total number of migrants was held constant. For example, if there were 100 total migrants and the migration was completely symmetric, then 50 individuals dispersed upstream and 50 individuals dispersed downstream. Conversely, if migration were completely asymmetric, then all 100 individuals would disperse downstream.

To evaluate how the location of the confluence with respect to other local populations influences the genetic pattern in river networks, we compared four networks with different confluence positions: no confluence, upstream confluence (e.g., those close to headwaters), confluence in the middle of the network, and downstream confluence populations (e.g., those close to river mouths) (Figure III.4A-H). The total number of populations was kept constant to eight; total network size was set to 700 Km (100 Km per segment), and the frequency of asymmetric migration downstream to 0.9 among all scenarios.

Lastly, we created four river networks with differing levels of branching complexity that varied from very simple (i.e., a single confluence; second-order river) to highly branched (i.e., 15 confluences; fifth-order river). In this case the total number of populations (and thus the total number of individuals) varied as a function of the river complexity, ranging from four

populations to 32 populations (Figure III.4I-P). Although the number of local populations varied, the results from this design were highly conservative because increasing the number of populations could only reduce the effects of genetic drift (see discussion). The total length of the network and the frequency of asymmetric migration downstream were kept constant (1000 Km and 0.9, respectively).

Results

For all comparisons performed across different networks, the per population carrying capacity (K= 100, 200 and 1,000) and number of offspring (N = 2, 10 and 20) did not qualitatively affect the genetic patterns observed (see Figure B.1). Varying the carrying capacity within the network (i.e., introducing heterogeneity in the system to represent a fish specialized to upstream environments) did not change the qualitative results (i.e., only the absolute standardized genetic diversity and amount of genetic differentiation differed; see Figure B.2). For these reasons, we report only the results for simulations with constant K = 200 and 10 offspring per individual per generation.

Landscapes vs. Riverscapes

We compared an open landscape with a branched river network to assess how river systems differ from less confined systems, such as certain terrestrial landscapes and marine environments. Based on levels of genetic diversity and differentiation among the populations, the genetic patterns associated with river networks differ quantitatively and qualitatively from open landscapes. Remarkably, genetic differentiation was an order of magnitude greater in the river network compared to the open landscape (river mean F_{ST} = 0.06 and open landscape mean F_{ST} = 0.002; Figure III.2C,D). The average genetic diversity per population across replicate simulation runs was the same in both systems (mean no. of genotypes = 21; Figure III.2A,C). However, at the population level, river networks have greater variance in genetic diversity among populations in comparison to the open landscapes (sd = 0.03 and 0.01, respectively). Note that the differences reported here are due solely to differences in connectivity between rivers and open landscapes because all other parameters were held constant.

At the population level, river networks accumulate higher genetic diversity at populations downstream of the confluence, with genetic diversity being on average three times smaller at the

upstream populations (e.g., headwaters) in comparison to downstream populations (e.g., those at river mouth) (Figure III.2B). In terms of genetic differentiation, river networks have the highest genetic differentiation between upstream populations than populations in open landscapes. For example, pairwise F_{ST} values between upstream populations are one order of magnitude larger in river networks (mean $F_{ST} = 0.1$) in comparison to two populations separated by the same distance in open landscapes (mean $F_{ST} = 0.01$) with asymmetric migration (see Figure III.2C,D). When we modify the migration directionality to be symmetric in the open landscape scenario (which is more realistic for landscapes where individuals can move freely between two populations), there is even greater homogenization of the genetic diversity in the open landscape (Figure B.3).

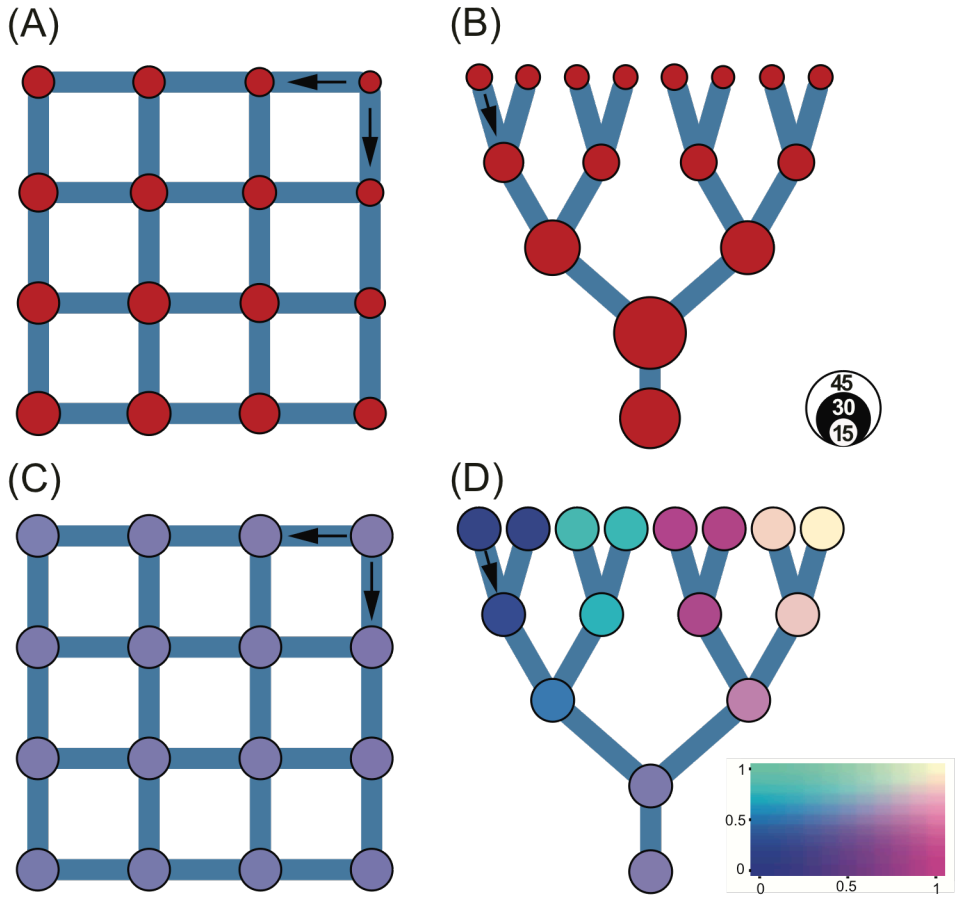


Figure III.2: Genetic diversity (A and B) and genetic differentiation (C and D) for open landscapes (i.e., terrestrial) and river networks with asymmetric migration (black arrows indicate the direction of migration); all parameters were held constant between both scenarios (i.e., segment length, *K*, and number of offspring). Genetic diversity is demonstrated by red circle sizes indicating the number of distinct genotypes present in each population (see legend in B), while genetic differentiation is demonstrated by color where more divergent differences in colors between two populations indicate higher pairwise F_{ST} values (legend in D illustrates position of local population in multivariate space). River networks have a distinctive distribution of genetic variation, with much greater genetic diversity found downstream of confluences and higher genetic differentiation between headwater populations.

Riverscape architectures

To evaluate the effect of asymmetric gene flow, we compared different migration patterns for both a linear network (i.e., low complexity; 1st order river) with a heavily branched network (i.e, more complex; fourth-order river). All variables were kept constant except the proportion of downstream migration, which varied from completely symmetric (0.5) to completely asymmetric (1; downstream only). In the strictly linear network, increasing downstream migration did not affect mean genetic differentiation, while in the branched network, mean F_{ST} increased with increases in asymmetric migration (i.e. downstream migration), being more than eight times larger in completely asymmetric scenarios than in symmetric migration scenarios (symmetric mean $F_{ST} = 0.03$; asymmetric mean $F_{ST} = 0.25$; Figure III.3A). It is important to keep in mind that these differences were found after varying the proportion of downstream vs. upstream migrants, but the total number of migrants was held constant across all tested scenarios.

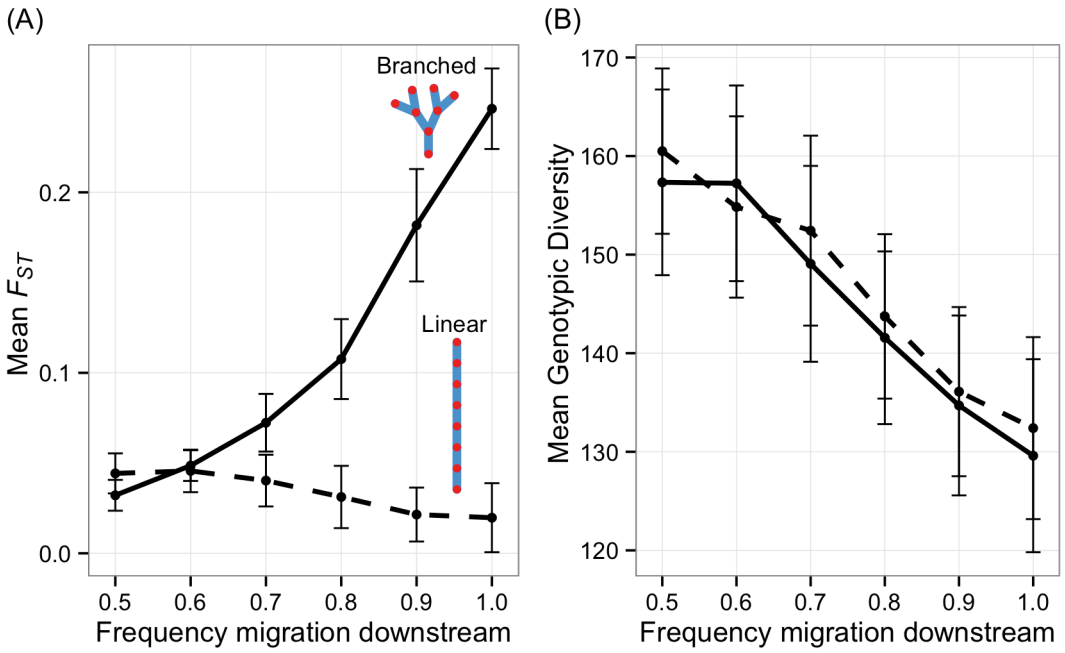


Figure III.3: Mean genetic differentiation (as measured by F_{ST}; A) and genetic diversity (i.e., number of genotypes; B) between a branched, 4th order river network (solid line), and a strictly linear network (dashed line). The total number of migrants between two populations was kept constant, however, we varied the proportion of downstream migrants from symmetric (i.e., 0.5 translates to an equal percentage of migrants that dispersed upstream and downstream) to completely asymmetric downstream (i.e., 1.0 translates to 100 percent of migrants moved downstream). Genetic differentiation in networks with branched architecture increases with increases in the proportion of downstream migrants. However, there are no substantial changes in genetic differentiation in linear networks. Genetic diversity is lost in both networks when asymmetric migration increases towards downstream migration. This pattern is due to the more rapid loss of alleles associated with genetic drift when migration is asymmetric. Error bars represent one standard deviation.

In contrast, a similar pattern in genetic diversity is observed for both simple and complex river networks, independent of the river network complexity. Increasing asymmetric migration downstream reduces overall genetic diversity in both systems where genetic diversity was calculated as the total number of unique genotypes in the entire river network (averaged across replicates; Figure III.3B). This effect is caused by the increasing rate at which genetic diversity is lost in the upstream populations because, with higher asymmetric rates of migration, the effects of genetic drift are not mitigated by gene flow.

Varying the position of a confluence along a river network, while all other parameters were kept constant, demonstrates that the location of a river confluence with respect to other local populations influences both genetic diversity and genetic differentiation (Figure III.4A-H). In general, headwaters have a lower accumulation of genotypes than downstream populations (Figure III.4A-D). Importantly, populations located along the tributary rivers have lower genetic diversity, compared with populations located at a confluence. These patterns are observed because unique genotypes from different tributaries are mixed in populations positioned downstream of a confluence. For the results presented here, this pattern drives confluence populations to have an average of two times greater genetic diversity than upstream populations.

Genetic differentiation increases with the increase in length of the tributary rivers (confluence close to upstream populations: mean F_{ST} = 0.003 (sd = 0.005) and confluence close to downstream populations: mean F_{ST} = 0.055 (sd = 0.05); Figure III.4E,H, respectively). For a confluence positioned further downstream, higher genetic differentiation accumulates in the upstream populations, indicating that rivers with longer tributaries (Figure III.4G-H) will have higher genetic differentiation than rivers with linear shape and/or shorter tributaries (Figure III.4E-F).

To assess how increasing complexity can affect genetic patterns in river networks, we varied the number of branches in a network following a fractal shape (river order 2 to 5, Figure III.4I-P). Increasing the complexity of the river networks increases the overall genetic diversity and differentiation in these systems (second-order river: mean F_{ST} = 0.008 (sd = 0.007) and mean of 10 genotypes (sd = 2.2); fifth-order river: mean F_{ST} = 0.06 (sd = 0.03) and mean of 278 genotypes (sd = 12.2); Figure III.4I-P). Likewise, while the absolute level of genetic differentiation was affected by varying the population size in the river network (Figure B.2), the same qualitative patterns of differentiation compared to a linear network (Figure III.3A) persists.

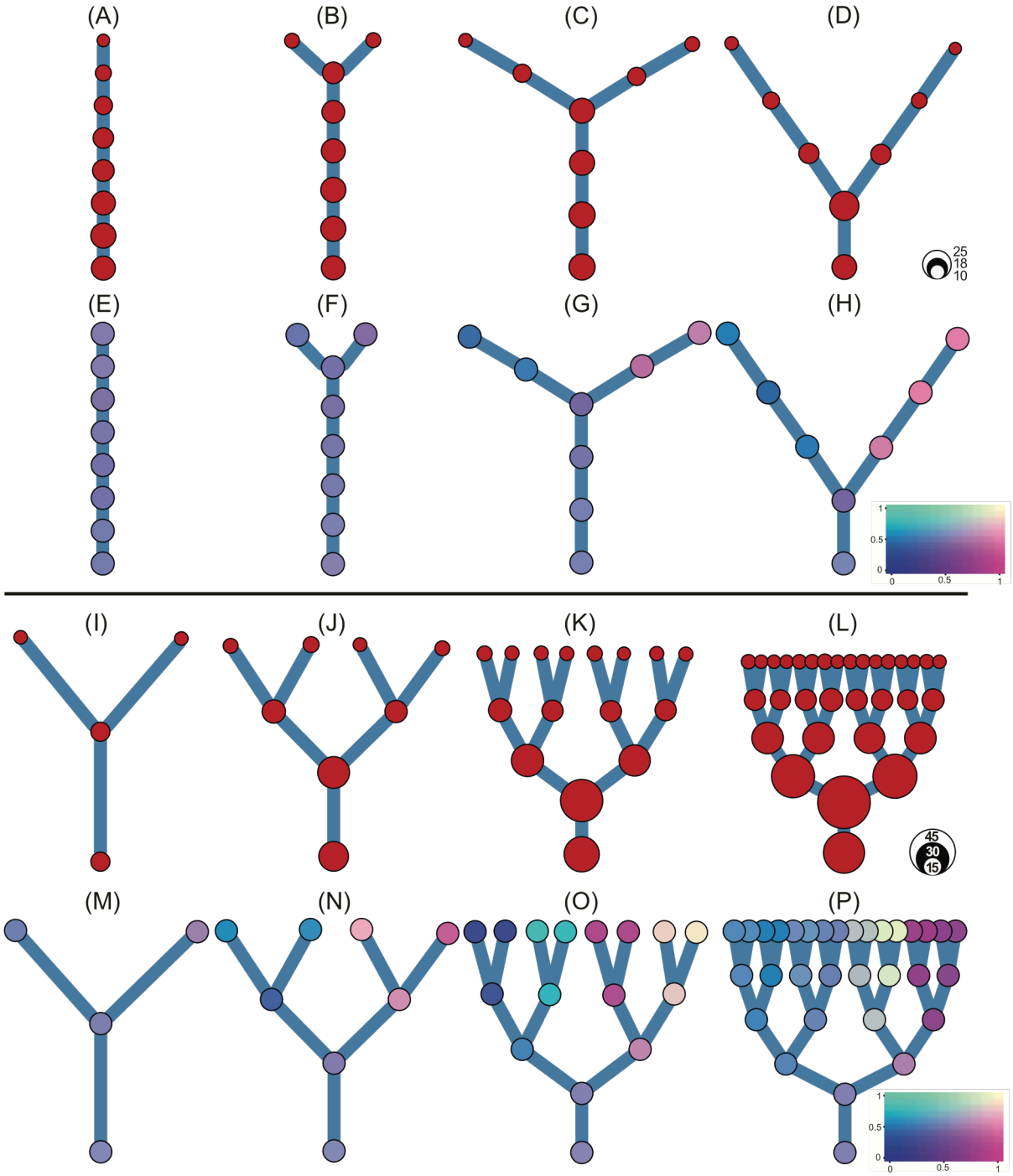


Figure III.4 - Comparison between different levels of river network complexity (i.e., confluence position and branching) in relation to genetic diversity (A – D; I - L) and genetic differentiation (E – H; M - P). Besides the network shape, all other variables were held constant (i.e., river total length, *K*, number of offspring, asymmetric migration rate, and dispersal kernel). Genetic diversity is demonstrated by red circle sizes indicating the number of unique genotypes present in each population (see legend in D and L), while genetic differentiation is demonstrated by color where more divergent differences in colors between two populations indicate higher pairwise F_{ST} values (legend in H and P illustrates position of local population in multivariate space). The confluence (A - H) accumulates higher genetic diversity downstream its position, in comparison to tributaries that has less diversity. As the location of a confluence moves further downstream, greater genetic differences are observed between populations in the

Genetic diversity increases substantially with increases in network complexity, but the diversity of upstream populations are uniformly low, regardless of the network complexity (Figure III.4I-L). On the other hand, downstream populations genetic diversity increases with each increase in river complexity (from a mean number of genotypes of 17 for 2nd order river to 35 in a 5th order river). Interestingly, the downstream populations do not have the maximum genetic diversity; the local population with highest genetic diversity is the population positioned at the main confluence position (Figure III.4I-L). In contrast to genetic diversity, genetic differentiation is higher among upstream populations when complexity increases (mean F_{ST} among most distant headwaters: second-order river = 0.02 and fifth-order river = 0.1; Figure III.4M-P) and, in general, differentiation is diminished between downstream populations. For all simulations reported here, the total length of the river network was kept constant to 1,000 km. As such, the distances between local populations were much greater in the less branched network (333 Km; Figure III.4I,M) versus the more branched network (32 Km; Figure III.4L,P).

Discussion

Half of all vertebrate species are fishes (33,629 of ~66,000; Eschmeyer and Fong 2015). Of this total, more than 40% of fishes live in fresh water, yet fresh water only accounts for 0.01% of all the water on earth. The high species diversity found among freshwater fishes is usually attributed to the comparatively high isolation of fishes found in different riverine basins (Lundberg et al. 2000; Guinot and Cavin 2015). Although this isolation leads to allopatric speciation without gene flow, the observed patterns of genetic differentiation presented by our ABM suggest that river properties may also be an important factor contributing to this high diversity of freshwater fishes, potentially leading to reproductive isolation within a river basin. For example, substantial isolation may be attained in the headwater populations, which can have little to no gene flow imposed by high levels of river branching and asymmetric migration downstream. Thus, the architecture of river networks alone may be an important factor in explaining the high species diversity observed in freshwater fishes.

In dendritic river systems, our results illustrate that greater levels of genetic differentiation occur with increases in network complexity (note that this effect is observed when total length is kept constant across river shapes). We also document that higher genetic diversity is observed in the downstream populations in comparison to the upstream populations. This

result has been previously documented in observational studies. For example, in a comparative study in Great Plain fishes, higher haplotypic diversity was observed in branched systems in comparison to linear rivers across three different species (Osborne et al. 2014). Other empirical studies corroborate the accumulation of genetic diversity in populations located further towards the river mouth (for a summary see Figure 6 in Morrissey and Kerckhove 2009; Paz-Vinas et al. 2015). Furthermore, our results suggest that the population with highest genetic diversity is not necessarily the most downstream populations (e.g., those closest to a river mouth), but rather the population located in the main confluence of rivers (Crispo et al. 2006), suggesting that genetic diversity can decline between the confluence and populations located further downstream. From a conservation and management perspective, these reservoir populations may warrant additional protection, especially given that they may be subject to increased anthropogenic stresses (e.g., dams, upstream expansion by invasive species).

It is important to point out that the simulations presented here were modeled after a small, non-migratory fish and the results would quantitatively and qualitatively vary depending on the vagility of the species (Labonne et al. 2008). For example, fishes that often migrate long distances could homogenize the genetic diversity within a river basin. Likewise, moderate to high genetic differentiation may still be maintained among local populations in species that exhibit high natal philopatry (e.g., anadromous salmonids; Whiteley et al. 2004), although river architecture per se may not be a factor of primary importance if reproduction is restricted to a particular river segment. Other potential caveats of our modelling approach were the equal population sizes and use of a constant dispersal kernel throughout the river network (i.e. dispersal patterns may vary along the length of a river network). Keeping these variables constant allowed us to isolate the effect of the river complexity. Adding additional parameters, such as decreased dispersal and smaller carrying capacities in populations located downstream, would only serve to increase genetic differentiation and decrease levels of gene flow in the upstream populations (e.g., in the headwaters of a river; see Figure B.2).

River architecture is just one factor that contributes to patterns of genetic variation and other geographic and/or ecological barriers (not considered in our simulations) within river networks could be important (e.g., waterfalls, different geomorphology along a river, habitat heterogeneity). Moreover, such barriers have been demonstrated to increase genetic differentiation between populations (Crispo et al. 2006; Pearse et al. 2009; Waters et al. 2015).

Our work doesn't discount the contribution of these other factors, but instead provides an important null model for expected patterns of differentiation that would arise from strictly neutral processes based on river architecture. This has important implications for partitioning the effects of geographic and ecological barriers, especially for comparisons among rivers or regions. For example, apparent differences in the effectiveness of a barrier due to differences in the degree of genetic differentiation among rivers might instead reflect differences in underlying river architectures. Likewise, differentiation that arises from river architecture should be taken into account for any inference about the role of adaptive divergence in isolating populations. Specifically, if genetic differentiation exceeds expectations based on genetic distance (one of the tests used to infer isolation by adaptation or ecology; Sexton et al. 2014), such differentiation might reflect the contribution due to river architecture rather than isolation arising from local adaption. For example, patterns ascribed to ecological divergence as inferred by F_{ST} outliers (i.e., adaptive divergence related to water color; Cooke et al. 2014) could be an artifact of the dendritic shape of the river. Here we show that by taking the river shape into consideration, a more strict and realistic assessment could be performed to test for ecological divergence, avoiding the false classification of neutral loci as having been under selection (Fourcade et al. 2013).

It is also important to note that the simulations performed here for identifying the impact of river shape are conservative by keeping the total river length constant across different levels of network complexity. For example, because the populations were positioned in each confluence, more complex networks contained local populations that were closer together in two-dimensional space (Figure III.4I-P). Yet, even with populations that were closer together, branched rivers had higher genetic differentiation than less branched rivers despite higher amounts of gene flow between local populations. If we had kept segment length constant (and varied the total length of the river network; a less conservative approach), then the genetic differentiation would be even higher in complex river networks as there would be a concomitant decrease in gene flow between local populations. This has important implications for considering what factors might contribute to the low levels of genetic differentiation observed in marine fish populations distributed along coastline environments (Riginos and Nachman 2001; Bernardi et al. 2003; Knutsen et al. 2003; Palumbi 2003). In particular, our results imply that it is not simply the distance separating coastal marine fish populations, but also the linear environment imposed

by coastlines itself, that contributes to patterns of genetic differentiation, based on the levels of gene flow between local populations observed in comparisons between linear and dendritic models in our simulations.

The genetic patterns obtained from our spatially explicit model demonstrate the importance of taking riverscape properties into consideration in population genetic and phylogeographic studies of aquatic populations. Our findings give clear evidence that the precise position of each local population can be an important variable to take into consideration when interpreting patterns of genetic variation within and between river basins (as opposed to averaging across populations within a river basin). Consequently, fully accounting for riverscape properties can generate realistic expectations of the genetic variation observed in these populations, which can be substantially different than open, more-connected landscapes (see also Paz-Vinas and Blanchet 2015). In general, our model results demonstrate that the architecture of river networks can explain the distribution and extent of genetic diversity found within aquatic populations (see also Della Croce et al. 2014 for how river architecture impacts in introgressive hybridization). Comparing these findings with empirical data and reconciling adaptive and neutral processes of diversity remain the next challenges for riverscape genetics.

References

Bernardi, G., L. Findley, and A. Rocha Olivares. 2003. Vicariance and dispersal across Baja California in disjunct marine fish populations. Evolution 57:1599-1609.

Burridge, C. P., D. Craw, D. C. Jack, T. M. King, and J. M. Waters. 2008. Does fish ecology predict dispersal across a river drainage divide? Evolution 62:1484-1499.

Christie, M. R., and L. L Knowles. 2015. Habitat corridors facilitate genetic resilience irrespective of species dispersal abilities or population sizes. Evol. Applic. 8:454-463.

Cooke, G. M., E. L. Landguth, and L. B. Beheregaray. 2014. Riverscape genetics identifies replicated ecological divergence across an Amazonian ecotone. Evolution 68:1947-1960.

Crispo, E., P. Bentzen, D. N. Reznick, M. T. Kinnison, and A. P. Hendry. 2006. The relative influence of natural selection and geography on gene flow in guppies. Mol. Ecol. 15:49-62.

Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23:3133-3157.

Della Croce, P., G. C. Poole, R. A. Payn, and C. Izurieta. 2014. Simulating the effects of stream network topology on the spread of introgressive hybridization across fish populations. Ecol Model. 279:68-77.

Eschmeyer, W. N., and J. D. Fong. 2015. Species by family/subfamily. Available at http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp. Accessed September 02, 2015.

Fourcade, Y., A. Chaput‐Bardy, J. Secondi, C. Fleurant, and C. Lemaire. 2013. Is local selection so widespread in river organisms? Fractal geometry of river networks leads to high bias in outlier detection. Mol. Ecol., 22: 2065-2073.

Goudet, J. 2005. Hierfstat, a package for R to compute and test hierarchical F‐statistics. Mol. Ecol. Notes 5:184-186.

Guinot, G., and L. Cavin. 2015. Contrasting "Fish" Diversity Dynamics between Marine and Freshwater Environments. Curr. Biol. 25:2314-2318.

He, Q., and L. L. Knowles. 2016. Identifying targets of selection in mosaic genomes with machine learning: applications in *Anopheles gambiae* for detecting sites within locally adapted chromosomal inversions. Mol. Ecol. 25:2226-2243.

Hopken, M. W., M. R. Douglas, and M. E. Douglas. 2013. Stream hierarchy defines riverscape genetics of a North American desert fish. Mol. Ecol. 22:956-971.

Horton, R. E. 1945. Erosional development of streams and their drainage basins: hydrophysical approach to quantitative morphology. Geol. Soc. Amer. Bull. 56:275-370.

Knutsen, H., P. E. Jorde, C. André, and N. CHR. Stenseth. 2003. Fine‐scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. Mol. Ecol. 12:385-394.

Labonne, J., V. Ravigné, B. Parisi, and C. Gaucherel. 2008. Linking dendritic network structures to population demogenetics: the downside of connectivity. Oikos 117:1479-1490.

Lotterhos, K. E., and M. C. Whitlock. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical methos. Mol. Ecol. 24:1031-1046.

Lundberg, J. G., M. Kottelat, G. R. Smith, M. L. Stiassny, and A. C. Gill. 2000. So many fishes, so little time: an overview of recent ichthyological discovery in continental waters. Annals Missouri Bot. Gard. 87:26-62.

Morrissey, M. B., and D. T. de Kerckhove. 2009. The maintenance of genetic variation due to asymmetric gene flow in dendritic metapopulations. Amer. Natur. 174:875-889.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2013. Vegan: community ecology package. R Package Version 2.0–9.

Osborne, M. J., J. S. Perkin,K. B. Gido, and T. F. Turner. 2014. Comparative riverscape genetics reveals reservoirs of genetic diversity for conservation and restoration of Great Plains fishes. Mol. Ecol. 23:5663-5679.

Palumbi, S. R. 2003. Population genetics, demographic connectivity, and the design of marine reserves. Ecol. Appl. 13:146-158.

Paz-Vinas, I., and S. Blanchet. 2015. Dendritic connectivity shapes spatial patterns of genetic diversity: a simulation-based study. J. Evol. Biol. 28:986-994.

Paz-Vinas, I., G. Loot, V. M. Stevens, and S. Blanchet. 2015. Evolutionary processes driving spatial patterns of intra-specific genetic diversity in river ecosystems. Mol. Ecol. 24:4586-4604.

Paz-Vinas, I., E. Quéméré, L. Chikhi, G. Loot, and S. Blanchet. 2013. The demographic history of populations experiencing asymmetric gene flow: combining simulated and empirical data. Mol. Ecol. 22:3279-3291.

Pearse, D. E., S. A. Hayes, M. H. Bond, C. V. Hanson, E. C. Anderson, R. B. Macfarlane, and J. C. Garza. 2009. Over the falls? Rapid evolution of ecotypic differentiation in steelhead/rainbow trout (*Oncorhynchus mykiss*). J. Heredity 100:515-525.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Radinger, J., and C. Wolter. 2013. Patterns and predictors of fish dispersal in rivers. Fish Fish. 15: 456-473.

Riginos, C., and M. W. Nachman. 2001. Population subdivision in marine environments: the contributions of biogeography, geographical distance and discontinuous habitat to genetic differentiation in a blennioid fish, Axoclinus nigricaudus. Mol. Ecol. 10:1439-1453.

Sexton, J. P., S. B. Hangartner, and A. A. Hoffmann. 2014. Genetic isolation by environment or distance: which pattern of gene flow is most common? Evolution, 68:1-15.

Strahler, A. N. 1957. Quantitative analysis of watershed geomorphology. Civ. Eng. 101:1258-1262.

Thomaz, A. T., L. R. Malabarba, S. L. Bonatto, and L. L. Knowles. 2015. Testing the effect of palaeodrainages versus habitat stability on genetic divergence in riverine systems: study of a Neotropical fish of the Brazilian coastal Atlantic Forest. J. Biogeogr. 42:2389-2401.

Waters, J. M., D. Craw, C. P. Burridge, M. Kennedy, T. M. King, and G. P. Wallis. 2015. Within-river genetic connectivity patterns reflect contrasting geomorphology. J. Biogeogr. 42:2452-2460.

Whiteley, A. R., P. Spruell, and F. W. Allendorf. 2004. Ecological and life history characteristics predict population genetic divergence of two salmonids in the same landscape. Mol. Ecol. 13:3675-3688.

APPENDIX B

Supplementary material from Chapter III

Supplementary Text

Discussion on Parameter Choice and, Scalability and Computation

Parameter choices: To test the contribution of river architecture under the model, we made a number of parameter choices aimed at characterizing a large number of stream and river dwelling fishes. For example, at a local (demic) scale, most stream fish populations are rarely larger than several hundred individuals and most are considerably smaller (Ward et al. 1994). In our model, changing the carrying capacity influences the amount of genetic drift and the absolute number of migrants per generation, but not the migration rate. As illustrated in Figure S1 (changing *K* across river networks) and Figure S2 (changing *K* within river networks), varying the carrying capacity resulted in changes to the absolute values of genetic differentiation and genetic diversity, but did not change the interpretation of our results. Likewise changing the number of offspring produced did not substantially change the interpretation of our results (Figure S1).

Regarding dispersal, our main assumption is that dispersal occurs between neighboring populations. For many freshwater species we believe this assumption is valid, although there may be occasional long-distance dispersal events that were not captured by this model. It is important to note that we did not model anadromous or highly migratory fishes because they represent a disproportionately small amount of riverine fish diversity and because their unique life-histories (e.g., high natal philopatry in salmon) may mitigate some of the effects of network complexity.

Regarding mutation rate, we selected a value of 10^{-6} to help balance the effects of drift (*i.e.,* with low mutation drift would eliminate much (or most) of the genetic diversity when *K* was set to 100 individuals per population). In reality, few natural populations are in drift-mutation equilibrium (Whitlock and McCauley 1999), and most are balanced by immigration,

but for computational reasons (see below), we chose not to simultaneously model multiple river networks.

Scalability and Computation: Agent-based models can be computationally intensive and thus some of our parameter and model-construction choices were designed to minimize the amount of time each run took in order to maximize the number of replicates we could run for each combination of parameters. The model runs more slowly as the number of individuals per population and the number of populations increases. That being said, we believe that our results can scale to a large number of systems. For example, the carrying capacity, dispersal kernel, distance between populations, and number of offspring produced during migration all dictate the number of migrants that move between populations. However, if we varied the parameters to run at much larger spatial or much longer temporal scales, the neutral interactions between the individuals and their landscape would remain the same. Thus, varying our parameter values to fit a particular species could certainly increase the real-world application of this model (and that certainly can be done), but the overall finding that neutral processes can be greatly affected by the architecture of river networks remains the same. Lastly, it is important to keep in mind that our results can scale to an entire river network (from headwaters to river mouth) or just a smaller section of a river network (e.g., connected upstream tributaries) depending on the particular habitats where species are found.

References

Ward, R. D., M. Woodwark, and D. O. F. Skibinski. 1994. A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. J. Fish Bio. 44:213-232.

Whitlock, M. C., and D. E McCauley. 1999. Indirect measures of gene flow and migration: FST≠ 1/(4Nm+ 1). Heredity 82:117-125.

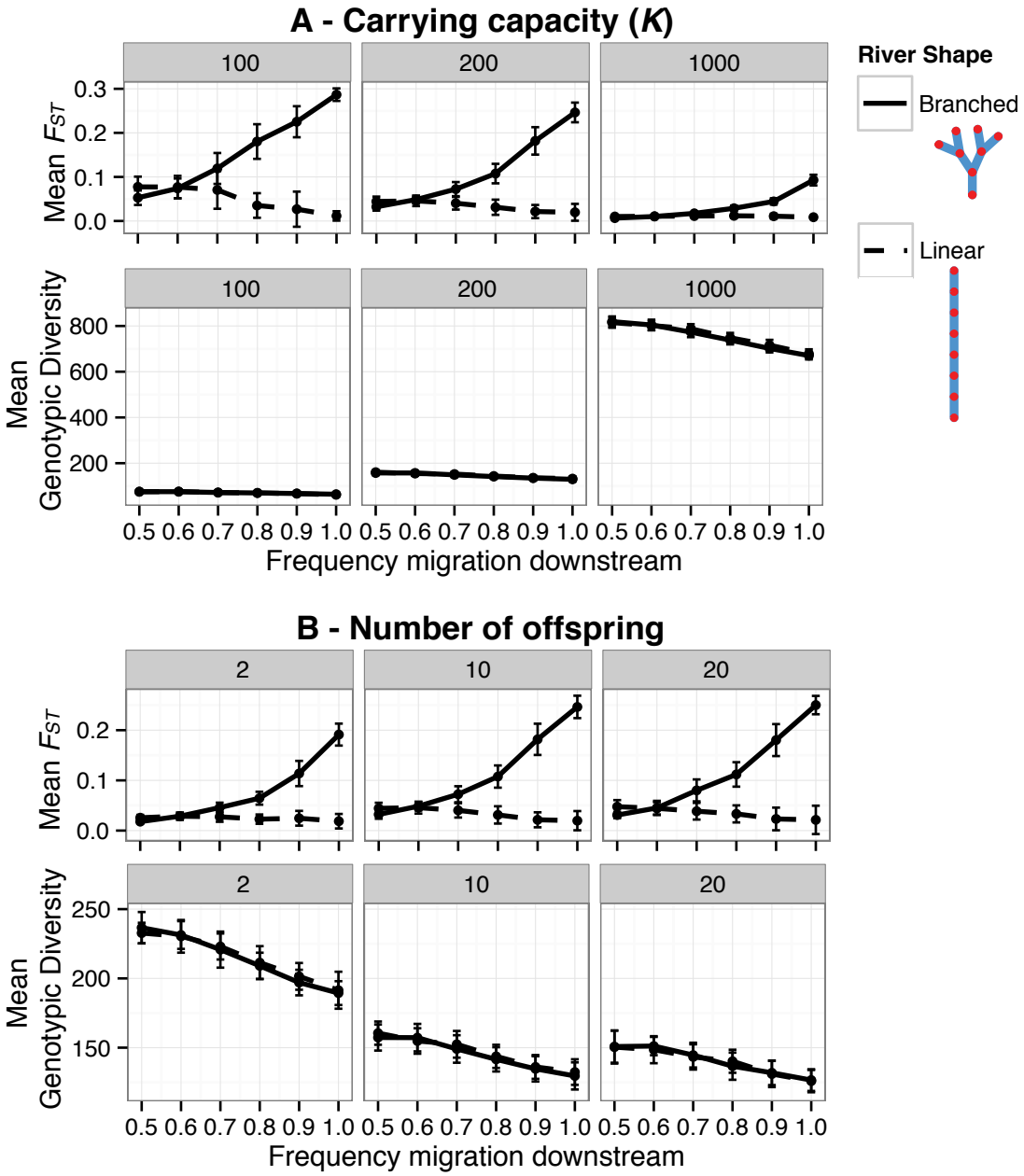


Figure B.1. Comparison based on mean genetic differentiation (as measured by F_{ST}) and genetic diversity (i.e., number of genotypes) between a branched, 4th order river (solid line), and a strictly linear network (dashed line) for: (A) different values of per population carrying capacity (K) and (B) number of offspring produced per individual per generation. The total number of migrants between two populations was kept constant, however, we varied the proportion of downstream migrants from symmetric (i.e., 0.5 translates to an equal percentage of migrants that dispersed upstream and downstream) to completely asymmetric downstream (i.e., 1.0 translates to 100 percent of migrants moved downstream). These comparisons demonstrate that while the absolute values differ for different values of K and offspring, they do not affect the qualitatively patterns of genetic variation. Error bars represent one standard deviation.

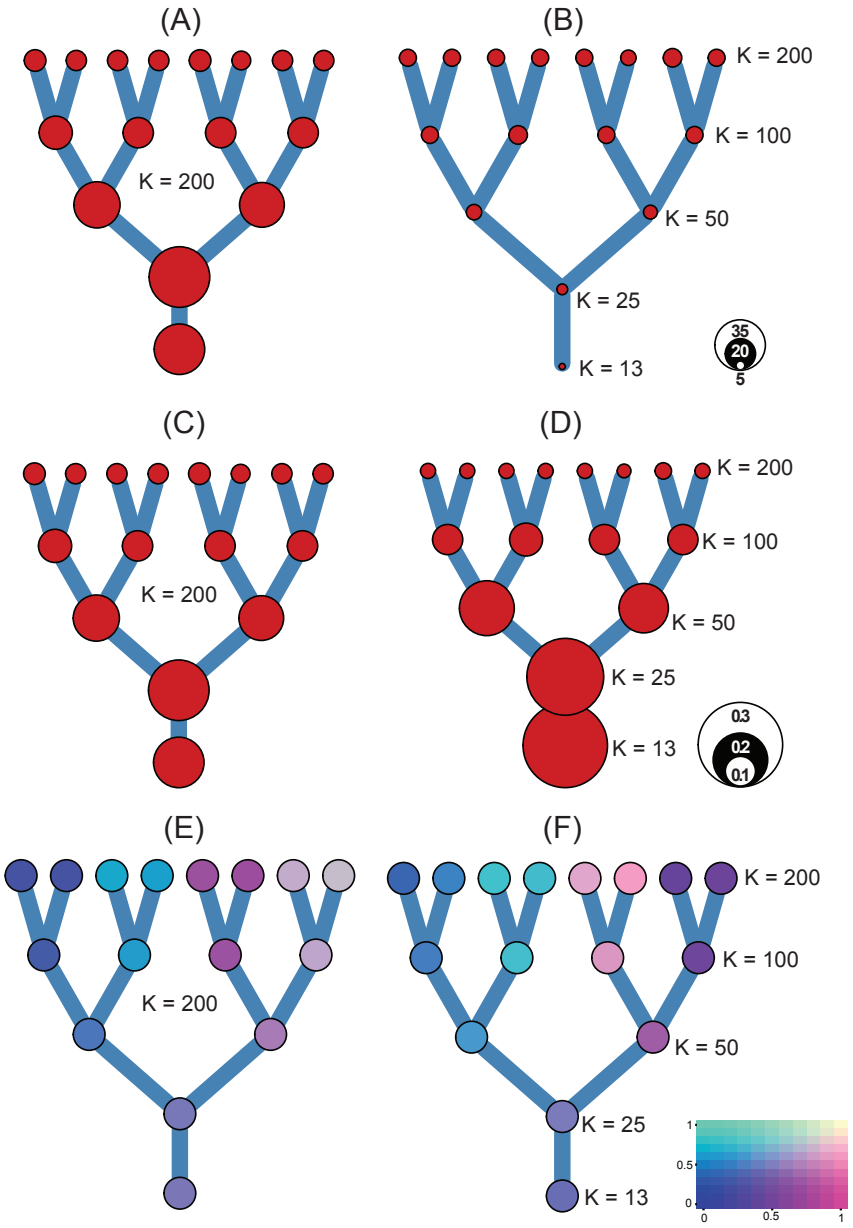


Figure B.2. Absolute genetic diversity (A and B), genetic diversity standardized by carrying capacity (genetic diversity divided by *K*; C and D) and genetic differentiation (E and F) for a habitat-generalist fish where *K* is identical between populations (A, C and E) and for a fish specialized to upstream habitats, where *K* decreases downstream (B, D and F). All additional parameters were held constant between both scenarios (i.e., segment length, number of offspring, dispersal Kernel, asymmetric migration). The standardization of genetic diversity was not needed in the main text of the manuscript because *K* was held constant for all presented simulations. Genetic differentiation is demonstrated by color where more divergent differences in colors between two populations indicate higher pairwise F_{ST} values (legend in F illustrates position of local population in multivariate space). Upstream habitat specialists have lower absolute genetic diversity than habitat-generalists (*cf.* A and B), but this pattern is driven by the overall smaller population sizes. When we take into account the per population *K,* we observe that the habitat-specialists have an even greater accumulation of genetic diversity downstream in comparison to generalists (*cf.* C and D). In terms of genetic differentiation, upstream habitat specialists have increased F_{ST} among headwaters, ranging from 0 to 0.1 for the habitat generalists in comparison to 0.02 to 0.14 for the upstream habitat specialists.

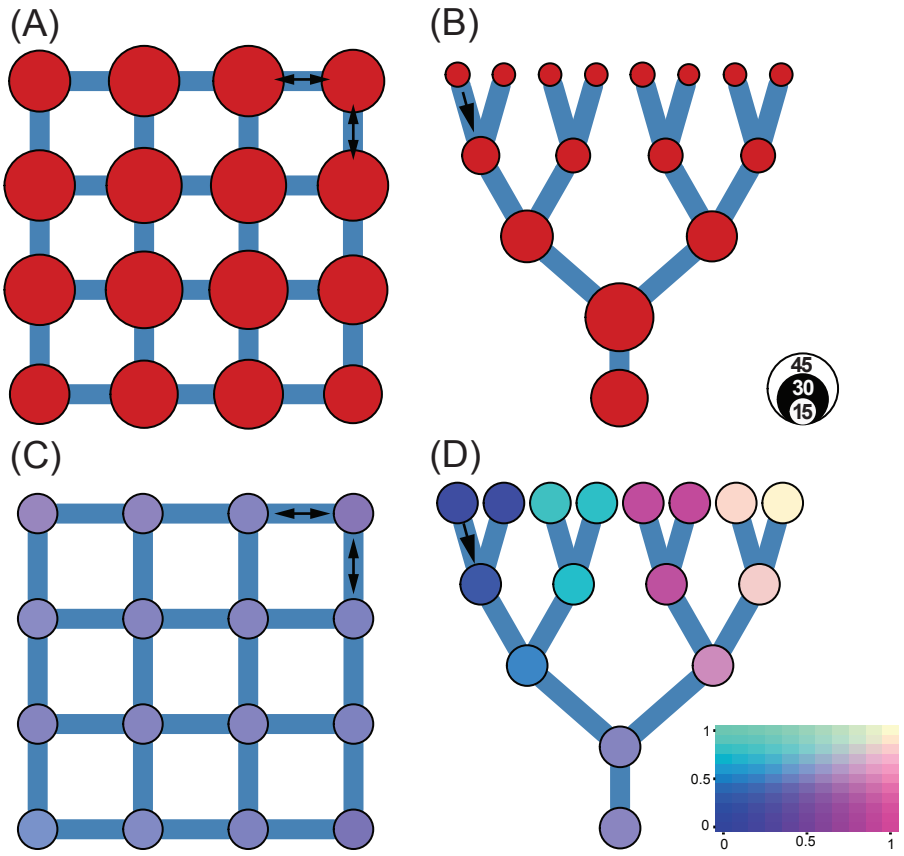


Figure B.3. Genetic diversity (A and B) and genetic differentiation (C and D) for open landscapes (i.e., terrestrial) and river networks with symmetric migration on the open landscape (compare to Fig. 2 with asymmetric migration on the open landscape; black arrows indicate the direction of migration). Genetic diversity is demonstrated by red circle sizes indicating the number of distinct genotypes present in each population (see legend in B), while genetic differentiation is demonstrated by color where more divergent differences in colors between two populations indicate higher pairwise F_{ST} values (legend in D illustrates position of local population in multivariate space). All parameters were held constant between both scenarios (i.e., segment length, K, number of offspring and dispersal Kernel), except the migration direction that was symmetric in the open landscape and asymmetric (9:1 ratio) for the river network. River networks have a distinctive distribution of genetic variation, with much greater genetic diversity found downstream of confluences and higher genetic differentiation between headwater populations

CHAPTER IV

Genomic signatures of paleodrainages in a freshwater fish along the southeastern coast of Brazil: genetic structure reflects past riverine properties

Andréa T. Thomaz, Luiz R. Malabarba, L. Lacey Knowles

Abstract

Past shifts in connectivity in riverine environments (e.g., sea level changes) and the properties of current drainages can act as drivers of genetic structure and demographic processes in riverine population of fishes. However, it's unclear whether the same river properties that structure variation on recent time scales will also leave similar genomic signatures that reflect paleodrainage properties. By characterizing genetic structure in a freshwater fish species (*Hollandichthys multifasciatus*) from a system of basins along the Atlantic coast of Brazil we test for the effects of paleodrainages caused by sea-level changes during the Pleistocene. Given that the paleodrainage properties differ along the Brazilian coast, we also evaluate if estimated genetic diversity within paleodrainages can be explained by past riverine properties (i.e., area and number of rivers in a paleodrainage). Our results demonstrate that genetic structure between populations is not just highly concordant with paleodrainages, but that differences in the genetic diversity among paleodrainages correspond to the joint effect of differences in the area encompassed by, and the number of rivers, within a paleodrainage. Our findings extend the influence of current riverine properties on genetic diversity to those associated with past paleodrainage properties. We discuss how these findings may explain the inconsistent support for paleodrainages in structuring divergence from different global regions and the importance of taking into account past conditions for understanding the high species diversity of freshwater fish that we currently observe in the world, and especially in the Neotropics.

Introduction

The properties of a riverine drainage are known to structure genetic variation among fish populations because of the constraints this habitat imposes on movement patterns. For example, theoretical models reveal how specific attributes of a river's architecture act as a driver of genetic divergence (e.g., Morrissey and de Kerckhove, 2009; Thomaz *et al*., 2016). Likewise, empirical studies identify genetic structure associated with shifts in species distribution in the past (e.g., Neuenschwander *et al*., 2008), especially for coastal fishes where Pleistocene sea level changes provided connections among rivers that are not present today (Thomaz *et al*., 2015). However, it's unclear whether the same properties of river architecture that structure variation on recent time scales will also leave similar genomic signatures (i.e., patterns of genetic variation among individuals/populations) that reflect paleodrainage architecture. In particular, although regional structuring of genetic variation reflective of the isolation among different paleodrainages due to changes in sea level have been documented in some cases (e.g., Chakona *et al*., 2013; Unmack *et al*., 2013, Thomaz *et al*., 2015), the impact of the properties of the paleodrainages themselves on patterns of genetic variation has not yet been tested. Specifically, because of the connections paleodrainages provide among currently isolated rivers during periods of sea level retreat, the properties of paleodrainages themselves may be reflected in regional measures of genetic diversity.

We address this question using genomic analyses in the freshwater fish *Hollandichthys multifasciatus* (Characiformes: Characidae), which is endemic to drainages along the southeastern Atlantic coast of Brazil. Specifically, we test the extent to which (i) structuring of genetic variation reflects past riverine connections (i.e., connections among currently isolated rivers within a paleodrainage) during the most extreme sea level retreat on the Pleistocene, the Last Glacial Maximum (LGM, 24-18ka), and given that the architecture of paleodrainages differs along the Brazilian coast (Figure IV.1), we (ii) test whether there are corresponding differences in the genetic diversity across paleodrainages that reflect the properties of paleodrainages themselves. We examine these questions using an approach that does not presuppose that genetic structure will be partitioned by paleodrainage boundaries. That is, we don't a priori define paleodrainages to ask whether there is a significant effect on genetic structure (as with a F_{ST}-analysis; see Thomaz *et al.*, 2015). Because multiple drainages are sampled within paleodrainages (except for 4 northern paleodrainages; see Figure IV.1), the genetic divergence

associated with paleodrainages and their respective properties is not reducible to a single drainage (or properties of a single drainage; see Thomaz *et al.*, 2015). Moreover, we do not assume that paleodrainages are the only potential factors influencing patterns of genetic variation. Instead, we apply a series of hierarchical analyses to infer genetic clusters that can accommodate a complex history in which multiple factors may be operating at different temporal and spatial scales (i.e., recent versus deeper past, and local versus regional barriers; see Massatti and Knowles, 2014). As such, our study provides not only the first analysis (that we are aware of) of the effects of paleodrainage properties on patterns of genetic diversity, but also our approach highlights potential methodological issues that might bias or contribute to some of the inconsistencies in past studies on the role of paleodrainages in structuring divergence among fish populations. Moreover, this historical perspective provides a complement to investigations of the effects of contemporary river architecture on genetic variation (Morrissey and de Kerckhove, 2009; Paz-Vinas *et al.,* 2015; Thomaz et al., 2016), although our specific study does not address the effects of contemporary river architecture.

Material and Methods

Sampling and RADseq genomic data generation and processing

Genomic data was generated for 182 individuals across the entire distribution of *Hollandichthys multifasciatus*. Sampled individuals were collected from 28 rivers (hereafter, referred to as populations; Figure IV.1) that span 12 paleodrainages; however, only individuals from 23 populations and 11 paleodrainages were analyzed (see below); for a brief description of how paleodrainages were identified from bathymetric data see supplementary text in Appendix C. Ethanol preserved tissues used in the study are catalogued in the ichthyology collections at the Universidade Federal do Rio Grande do Sul (UFRGS), Museu the Ciências e Tecnologia, Pontifícia Universidade Católica do Rio Grande do Sul (MCP) and Museu the História Natural Capão da Imbuia (MHNCI) (see complete list in Table C.1).

Genomic DNA was extracted from body muscle using Qiagen DNeasy Blood and Tissue Kit or modified salt-precipitation protocol (Medrano *et al*., 1990) and two double digest reduced representation libraries (ddRAD) were constructed following the protocol of Parchman *et al.* (2012). Briefly, the DNA was double digested with two restriction enzymes (*EcoRI* and *MseI*), followed by a ligation step and amplification by PCR, where unique barcodes (10bp) and

Illumina adaptors were added to the digested DNA. PCR products were cleaned to select fragments between 350-450bp by gel extraction (QIAquick Gel Extraction Kit – Qiagen). The two libraries were sequenced for 100bp in two lanes of Illumina *HiSeq2000* at the University of Michigan DNA core facility, producing 325 million reads in total (146 and 179 million in each library).

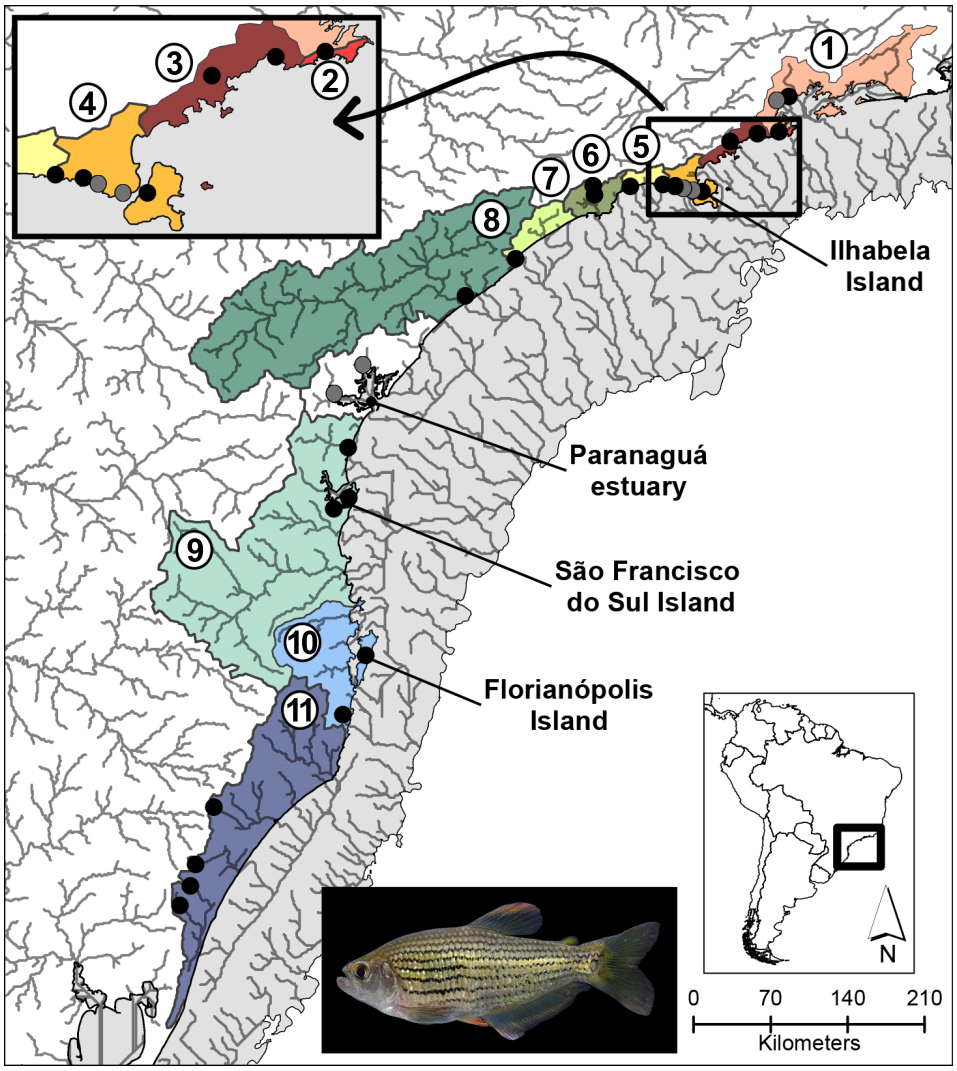


Figure IV.1. Map of the 11 studied paleodrainages that formed during sea level retreats of the LGM along the southeastern coast of Brazil, with an image of *H. multifasciatus* (99.5mm SL). The paleodrainage area is shown in different colors and populations sampled for genomic analyses are marked by black dots. Note that one dot in paleodrainage 10 represents three populations on Florianópolis Island. The grey shaded area marks the exposed area during the sea-level retreat in the LGM. The grey dots identify populations excluded from analyses (see methods for detail; also Thomaz *et al.*, 2015).

The pipeline STACKS version 1.35 (Catchen *et al.*, 2013) was used to demultiplex and process the genomic sequences. One mismatch in the adapter sequence (--adapter_mm) and a barcode distance of two was used in *process radtags* to allow barcode rescue (--barcode_dist); adaptor sites were removed using Seqtk (Heng Li, https://github.com/lh3/seqtk) by deleting 5bp in the 5' end (-b 5). Individuals from the two libraries were then pooled together and individuals with less than 500K sequences were excluded. The resulting 239 million retained reads from 166 individuals (average of 1,422,655 ± 615,385 sequences per individual) was run in USTACKS with the following settings: a minimum depth coverage of five (-m 5), the *Removal algorithm* (-r), and the *Deleveraging algorithm* (-d), with model type equal bounded (--model_type), and an error bound for ε of 0.1 (--bound_high), which generated data with a mean coverage of 13.7 (±5.7). A catalog of genomic sequences was built in CSTACKS, allowing for two mismatches between sample tags (-n 2), and loci for each individual were identified using SSTACKS under default options.

From SSTACKS output we directly run the POPULATIONS module (with parameters: -r 0 -p 2 -m 5 --min_maf 0 --max_obs_het 0.5). The resulting output was processed in R version 3.3.1 (R Core Team, 2016) to eliminate SNP's from the five last base pairs in the 3' end of each locus, as well as loci with exceedingly high genetic diversity since such high values are suggestive of sequencing and assembly errors (i.e., $\theta > 0.024$, representing loci in the upper 95% quantile of the distribution of genetic diversity; Figure C.1). In addition, five populations were excluded because of limited data (i.e., three populations with less than two individuals after data processing) or ambiguities with paleodrainage assignment (i.e., two populations associated with the Paranaguá estuary; see Thomaz *et al.*, 2015). The resulting dataset contained a total of 517,874 SNPs in 196,845 loci (maximum of 10 SNP's per locus), with a genotyping rate of 0.29, for 149 individuals sampled in 23 populations from 11 paleodrainages (see Table C.1 for number of reads per individual). All STACKS modules were run under parallel execution with 8 threads in the University of Michigan flux.

Because the robustness of different methods of analysis to missing data varies, we generated two datasets with different levels of missing data. Specifically, one dataset included loci present in at least 10 populations and 75% of individuals within a population (i.e., 149 individuals and 62,549 SNPs in 18,407 loci, with a genotyping rate of 0.55) and was used to calculate genetic diversity summary statistics for each paleodrainage in STACKS (i.e., π and

H_{EXP} averaged across populations within a given paleodrainage; see Table C.2). F_{ST}-values and its significances were calculated in Arlequin 3.5.2.2 (Excoffier and Lisher, 2010) with 10,000 replicates with a Bonferroni correction for multiple comparisons. The other dataset included loci with a maximum of 25% missing data per unlinked SNP (hereafter referred to simply as SNPs) per individual (i.e., 116 individuals and 6,574 SNPs, with a genotyping rate of 0.89) and was filtered using the toolset PLINK v.1.07 (Purcell *et al*., 2007).

Because the degree of divergence among individuals affects the proportion of shared loci in RADseq data (see Huang and Knowles, 2014), in addition to the two aforementioned datasets with individuals from the entire geographic range (hereafter referred to as the full datasets), we also processed the genetic data using two subsets of individuals to minimize the effect of missing data. Specifically, we processed individuals from the northern and southern regions separately (hereafter referred to as the northern and southern datasets, respectively), thereby increasing the amount of SNPs retained in each subset because of a fairly deep divergence separating northern and southern groups (Thomaz *et al*., 2015).

Tests of genetic structure associated with paleodrainages

To evaluate whether there was a correspondence between population genomic structure and paleodrainages without conditioning upon paleodrainage membership, inferences of genetic structure were made using STRUCTURE 2.3.4 (Pritchard *et al*., 2000). The full dataset was analyzed with *K*-values ranging from 1 to 12 (maximum number of paleodrainages + 1). An iterative approach was then used to explore potential hierarchical genetic structure (i.e., genetic structure that might be present within initial clusters identified by STRUCTURE; see Ryan *et al*., 2007; Massatti and Knowles, 2014). Specifically, STRUCTURE analyses were run for a subset of individuals contained within genetic clusters and individuals were assigned probabilistically to genetic clusters, where the number of *K*-values analyzed ranged from $K = 1$ to the number of paleodrainages + 1, depending the data subset. These analyses were conducted using the northern and southern datasets to take advantage of inclusive loci to each of the two geographic areas (as described above). Ten independent runs were performed for each STRUCTURE analyses using the "Admixture model" and "Allele Frequencies Correlated" model for 300,000 MCMC iterations and 100,000 of burn-in, except for a few cases in which 500,000 MCMC and 200,000 of burn-in were performed to ensure convergence. The ΔK of Evanno *et al*. (2005) implemented in

STRUCTURE HARVESTER (Earl and vonHoldt, 2012) was used to identify the *K* number of genetic clusters that best fit the data, with the assignment of individuals in proportion to their putative ancestral history presented graphically using the CLUMPAK pipeline (Kopelman *et al*., 2015).

Estimates of divergence times

Divergence times between neighboring paleodrainages were estimated using a composite-likelihood method based on the site frequency spectrum (SFS) as implemented in FASTSIMCOAL2 (Excoffier and Foll, 2011; Excoffier *et al*., 2013) to evaluate whether they were consistent with a Pleistocene divergence, and specifically, a divergence time during the LGM. We used a python script to remove all missing data to calculate the joint SFS between each neighboring paleodrainage pair (available from Papadopoulou and Knowles, 2015), based on the vcf file from STACKS with a single SNP per locus. Five individuals from each paleodrainage were used to calculate the SFS, except for two paleodrainages (paleodrainage 3 and 8; see Figure IV.1) where only three individuals were available. Divergence times were estimated assuming no migration between paleodrainages from polymorphic loci (i.e., using the *"removeZeroSFS"* option in FASTSIMCOAL2). This assumption of no migration might result in underestimates of divergence times, however we note that the STRUCTURE analyses do not provide strong evidence of substantial admixture. Moreover, it is the relative similarity in the estimated divergence, not the absolute timing of divergence per se, that is particularly relevant to interpreting the relationship between paleodrainage properties and genetic diversity (i.e., general similarities in divergence times controls for the potential confounding effect of different genetic diversities that could have resulted if the times to accumulate genetic diversity differed among paleodrainages).

To improve the accuracy of parameter estimates from the SFS (following the recommendations of the program; see Excoffier and Foll, 2011), we calculated the effective population size of one paleodrainage (N_1) directly from the empirical data (i.e., specifically, from the nucleotide diversity (π) of fixed and variable sites). The other parameters of the divergence model (N_2, ancestral population size N_{ANC} and divergence time T_{DIV}) were estimated based on the SFS, with a mutation rate, μ, of 2.24×10^{-8}. This mutation rate was estimated from the regression formula for cellular organisms (Lynch, 2010) based on a genome size of 1500mb for

Hollandichthys (which is based on the average genome size of Characidae "*clade C*", where *Hollandichthys* is currently positioned; Thomaz *et al.*, 2010; www.genomesize.com), with one generation per year. A total of 40 FASTSIMCOAL2 runs were conducted for each paleodrainage pair with 100,000 to 250,000 simulations per likelihood estimation based upon a stopping criterion of 0.001, and 10 to 40 expectation-conditional cycles (ECM). A parametric bootstrap was used to estimate 95% confidence intervals on the model parameters. Specifically 100 simulated SFS with the same number of individuals, loci and parameters from the maximum composite likelihood estimate were used to re-estimate demographic parameters (as with the estimates of the empirical data, 40 FASTSIMCOAL2 runs was performed per simulated dataset with the same criteria for likelihood estimation).

Tests of relationship between genetic diversity and paleodrainage properties

To test whether patterns of genetic diversity (i.e., π and H_{EXP}) correspond to paleodrainage properties, we estimated two properties: land area and number of isolated rivers within a paleodrainage. The relationship between genetic diversity and these paleodrainage properties were evaluated using generalized linear models (i.e., linear regression and covariance analyses) with the function *lm* in the basic stats package in R. For the four models (i.e., area, number of rivers, area + number of rivers, area * number of rivers), the corrected Akaike information criterion (AICc) was used for model comparison using the function *aictab* in the R package *AICcmodavg* (Mazerolle, 2016).

The paleodrainage property of land area was characterized based on the current exposed land area (i.e., excluding the submerged area) in ArcGIS 10 based on the paleodrainages map (see text in Appendix C for a brief summary of details regarding the identification of paleodrainages based on topographic relief contours; see also Thomaz *et al*., 2015). Note that total paleodrainage area was also calculated. However, because it was highly correlated with current exposed area ($R^2 = 0.97$; p-value <0.001; Figure C.2A), and since all inferences about genetic diversity are based on sampled populations from the exposed area, we only present results on the current exposed area (and hereafter is referred for simplification as area).

The number of isolated rivers in a paleodrainage (i.e., those that are not currently connected) was used as a measure of complexity, in the sense that more rivers translate into more opportunities for the retention of genetic differences. The number of rivers in a paleodrainage

was calculated using **Hydro**logical data and maps based on **S**huttle **E**levation **D**erivatives at multiple **S**cale (HydroSHEDS - USGS) maps. Grids with an upstream catchment area of ≥ 1,000 cells were defined as rivers, which for the region is ~8 km^2.

Results

Tests of genetic structure associated with paleodrainages

STRUCTURE analyses identified genetic clusters that corresponded to paleodrainage membership without using prior information about the geographic location of individuals (i.e., without conditioning on paleodrainage; Table IV.1 and Figure IV.2). At each level of the analysis for each subset of data an additional paleodrainage break was identified given the hierarchical structure of genetic variation. Moreover, probabilistic assignment of individuals to the respective genetic clusters revealed little evidence of admixture; admixture was inferred between two of seven sampled populations from paleodrainages 9 and 10 (Figure IV.2).

There was one exception in which the genetic break did not correspond to a paleodrainage boundary, in addition to the previously documented pronounced biogeographic division between northern and southern populations (Figure C.3; see also Thomaz *et al.*, 2015, based on mtDNA). Specifically, there was an unexpected genetic break between IB and SS populations in the paleodrainage 4 (Figure IV.2). Note that since there was not significant structuring associated with paleodrainage 4 it was not included in the subsequent STRUCTURE analyses aimed at detecting additional structure within regional groups.

Table IV.1. Summary of STRUCTURE results for a series of sequential analyses to account for the hierarchical nature of divergence (see Figure IV.2 for detailed plots of the probable ancestry of each individual). For each analysis (i.e., row), the first and second most probable *K*-values identified using Evanno method are reported along with the correspondent *ΔK*. The total number of loci and individuals analyzed are given, as well as the total individual genotyping rate.

Paleodrainages analyzed	Loci	Individuals	Genotyping rate	1st *K*	*ΔK*	2nd *K*	*ΔK*
All (1-11)	6574	117	0.89	2	7218.7	4	1799.0
North (1-8)	8638	70	0.91	2	7270.5	6	118.4
1, 2, 3	8126	22	0.94	3	1120.0	2	2.23
5, 6, 7, 8	8204	36	0.91	3	910.03	4	697.42
7, 8	7459	12	0.91	2	509.4	-	-
South (9-11)	9105	51	0.89	2	5053.6	3	396.1
9, 10	7387	23	0.9	2	2651.3	4	6.4

Pairwise genetic differentiation (F_{ST}) varied almost one order of magnitude (0-0.95 mean = 0.67 ± 0.21; Table C.3). This broad range reflected the hierarchical structuring of genetic variation (Figure IV.2). Specifically, there is a pronounced differentiation between comparisons of populations between the southern and northern regional groups (mean = 0.77 ± 0.12) relative to lower levels of differentiation between paleodrainages (mean = 0.56 ± 0.23) within the respective northern and southern regions, or among populations from the same paleodrainage (mean = 0.41 ± 0.32).

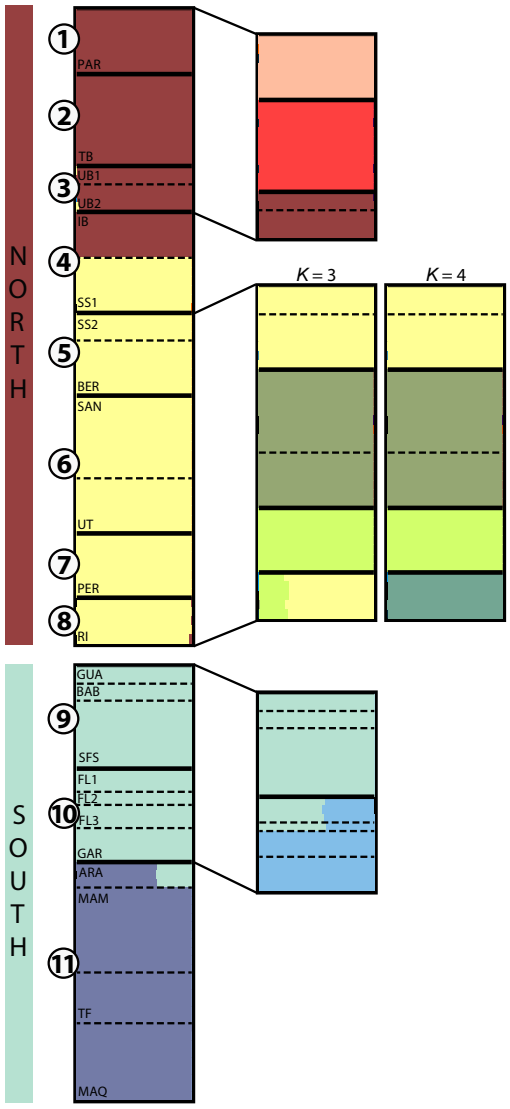


Figure IV.2. Results from hierarchical STRUCTURE analyses depicting the hierarchical nature of genetic structure (i.e., each block corresponds to separate analyses, with different colors identifying the different numbers of inferred *K* genetic clusters). Thick black lines and numbers in circles demarcate paleodrainages and dashed lines the populations within a paleodrainage, whose names are listed on the left, arranged from north (PAR) to south (MAQ). Color pattern in the hierarchical runs corresponds to the individual paleodrainages on the map in Figure IV.1.The posterior probabilities of the ancestry of each individual are shown (i.e., the relative proportion of different colors).

Estimates of divergence times

Divergence time estimates corroborate the hierarchical structure of genetic variation with an older regional divergence between the northern and southern regions versus relatively recent divergence times among geographically adjacent paleodrainages not separated by this geographic split (Figure IV.3). Specifically, the divergence between the northern and southern regions was estimated around 80ka, whereas divergence time estimates between paleodrainages pairs are generally centered on the LGM, ranging between 12ka to 44ka (Figure IV.3). In most cases, estimates of the ancestral population sizes were larger than the current populations, except for the paleodrainage pairs 2 – 3 and 7 - 8 (see Table C.4). Also, note that the most recent divergence time is estimated between paleodrainages 4 and 5 and one of the largest ancestral population sizes is estimated between paleodrainages 3 and 4 (Table C.4). These parameter estimates are likely biased because paleodrainage 4 violates the assumption that divergence times between paleodrainages predate divergence among populations within a paleodrainage (see Figure IV.2).

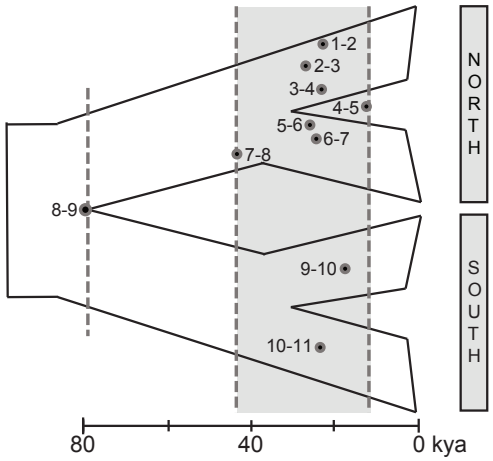


Figure IV.3. Schematic representation of divergence time point estimates between geographically adjacent pairs of paleodrainages, (represented schematically as a general population split in either the northern or southern areas), along with the divergence time estimate between the northern and southern regions (i.e., between paleodrainages 8 and 9). The scale bar at the bottom of the figures shows estimates of the absolute divergence times (see methods for details regarding the mutation rate used for dating).

Tests of relationship between genetic diversity and paleodrainage properties

Irrespective of which measure of genetic diversity was used (i.e., π or $H_{EXP;}$ Table IV.2), the linear models identified a significant association between genetic diversity and the joint effect of paleodrainage area and number of rivers within a paleodrainage (Figure IV.4; Figure C.4). Specifically, despite the additional model complexity, when both paleodrainages properties are analyzed together (i.e., considering the covariance between area and the number of rivers within a paleodrainage), model comparison based on AICc scores suggests a significantly better fit compared to analyses based on each riverine property separately, or when considering a possible interaction between the paleodrainage properties (Table IV.2); area is correlated with

the number of rivers in each paleodrainage (R^2 = 0.39; p-value = 0.04; Figure C.2B). Among the models tested, the number of rivers within a paleodrainage was the worst fit, and by itself was not significant; however, this may reflect reduced statistical power given the restricted range of differences in this variable across paleodrainages (see Table C.5).

Table IV.2. Comparison of the relative effect of area and number of rivers per paleodrainage on patterns of genetic variation based on the corrected Akaike Information Criterion (AICc); models are listed in order of their predictive value for analysis based either of the population genetic summary statistics, π or H_{EXP}.

Sum. stat.	Model	# parameters	R^2	R^2-adj	p-value	AICc	ΔAICc	Model prob.
π	Area + River	4	0.81	0.76	<0.01	-57.25	0.00	0.94
	Area	3	0.43	0.37	0.03	-50.62	6.63	0.03
	Area*River	5	0.81	0.72	0.01	-49.96	7.28	0.02
	River	3	0.01	-0.11	0.84	-44.50	12.75	0.00
H_{EXP}	Area + River	4	0.79	0.74	<0.01	-60.09	0.00	0.93
	Area	3	0.40	0.33	0.04	-53.77	6.32	0.04
	Area*River	5	0.79	0.70	0.01	-52.79	7.31	0.02
	River	3	0.01	-0.10	0.78	-48.25	11.85	0.00

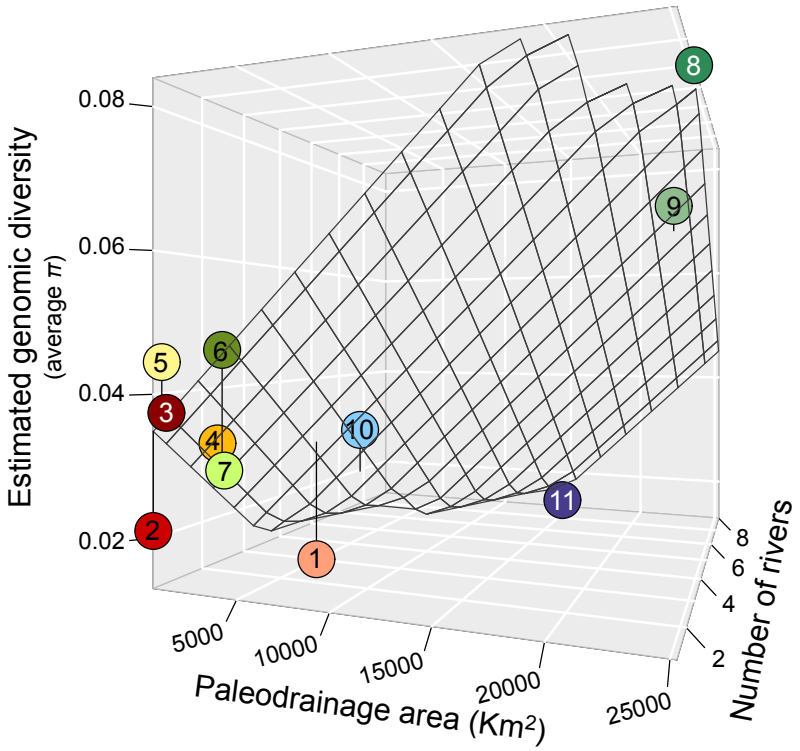


Figure IV.4. General linear model fit between paleodrainage properties (i.e., area and number of rivers) and nucleotide diversity (π; R^2 = 0.81, p-value <0.01; see Figure C.4 for results based on H_{EXP}). The colored dots and numbers correspond to the individual paleodrainages on the map in Figure 1.

Discussion

Studies have clearly demonstrated the role of paleodrainages in structuring patterns of genetic variation, where genetic divergence accumulates due to the relative isolation of rivers from different paleodrainages compared with the past connections forged among rivers within a paleodrainage, although these effects appear to vary (e.g., Chakona *et al*., 2013; Unmack *et al*., 2013, Thomaz *et al*., 2015). Our work adds another empirical example, and it extends this influence to dimensions that have not yet been studied. Specifically, inferences about the structuring of genetic variation by paleodrainage are (i) detected without a prior classification of paleodrainage membership of populations, in contrast to tests like F_{ST} analyses in which the groups are defined a priori, and (ii) we show that the paleodrainage properties themselves affect genetic diversity (i.e., the presumed connections among currently isolated rivers during periods of sea level influence regional patterns of genetic diversity). Below we discuss why considering potential contributors of processes at different spatial and temporal scales (i.e., regional versus local, and current versus past history) might explain some of the enigmatic results about the relative importance of specific factors in structuring populations of fish, as demonstrated in terrestrial environments (e.g., Papadopoulou and Knowles, 2016), as well as processes of fish diversification that might underlie regional and/or taxonomic differences in richness patterns (e.g., Tedesco *et al*., 2012; Dias *et al*., 2014).

Paleodrainage effects on genetic variation

With regards to the methodologies used to detect the contribution of paleodrainages, our results highlight how the criteria applied for such inferences may influence the conclusions (see also Papadopoulou and Knowles 2016). For example, our results show a strong correspondence of genomic structure in *Hollandichthys* with paleodrainages boundaries (i.e., in 10 of the 11 paleodrainages, with the only exception in paleodrainage 4), without a priori classification of populations to paleodrainage (i.e., the genetic clustering of populations sampled within a paleodrainage reflects shared ancestry under the presumed genetic equilibrium being modeled here). However, the detected genetic structure above the level of individual populations is not limited to paleodrainage boundaries (Figure IV.2; Table IV.1). For example, the northern-southern split between the Paranaguá estuary populations (Figure C.3; see also Thomaz *et al.*, 2015) predates the divergences reflective of paleodrainage structure (Figure IV.3). By applying a

series of hierarchical independent STRUCTURE analyses to accommodate this complex history of divergence, the genetic structure associated with paleodrainages becomes clear (Figure IV.2). In other words, the effects of paleodrainages are clear when accounting for the complexity of the history of *Hollandichthys*, but could have been overlooked by framing the question about structuring of genetic variation by paleodrainages as a binary "yes" or "no" question.

Similar arguments about potentially misleading conclusions might be made based on how DNA sequences are analyzed. For example, for recent divergence histories, a correspondence between clades in a gene tree and paleodrainage boundaries or the distribution of haplotypes across populations within paleodrainages (e.g., Chakona *et al*., 2013; Unmack *et al*., 2013) are very conservative criteria for inferences about the role of paleodrainages in structuring genetic variation. The lack of monophyly may simply reflect that there has not been sufficient time for the sorting of ancestral polymorphism (see Knowles, 2009; Hudson and Coyne, 2002). Likewise, the lack of shared haplotypes among rivers within a paleodrainage does not discount the possible role of paleodrainages; it simply identifies structure associated with current isolated rivers (as do our analyses; Figure IV.2, Table C.3). Because of overestimation of divergence times when based directly on pairwise sequence differences (see Edwards and Beerli, 2000), such estimates are also unlikely to coincide with Pleistocene driven sea-level shifts that define paleodrainage boundaries. In other words, conclusions about the role of paleodrainages associated with Pleistocene sea level changes might be sensitive to how tests are conducted and interpreted given the time frame of these historical events (Knowles, 2009). With relatively larger ancestral population sizes than current effective population sizes estimated for paleodrainages in *Hollandichthys* (Table C.4), the much more recent divergences estimated here (Figure IV.3) compared to previous estimates based on mtDNA (Thomaz *et al*., 2015) are not unexpected given these divergence estimates reported here take into account gene divergences that predate population divergence (see Carstens and Knowles 2007). Migration (which was not modeled here) would result in underestimates of divergence times; however, there is little to no evidence of admixture among paleodrainages (see Figure IV.2).

Besides methodological biases, differences in the detected effects of paleodrainages on genetic variation across studies might also reflect differences in specific properties of a local region. For example, the availability and stability of environments overtime are known to affect the current genetic diversity of species in terrestrial organisms (e.g., Pleistocene refugia theory;

He *et al*., 2013; Massatti and Knowles, 2016). In a similar way, our findings demonstrate that genetic diversity within paleodrainages is a function of its properties, with higher genetic diversities observed in larger and more branched paleodrainages (i.e., more constituent rivers). Note the similarity in divergence times among paleodrainages (except the north-south break; Figure IV.3) means that this pattern cannot be explained by differing times of accumulation of genetic diversity among paleodrainages.

This strong genomic signature urges the incorporation of information about past river structure (see also Neuenschwander *et al.*, 2008), rather than just considering the properties of current rivers. As with the detected effects of paleodrainage area and river number demonstrated here (Figure IV.4), additional properties of rivers in the past (which were assumed to be constant in space and time here) might also contribute to differences in genetic diversity among paleodrainages. For example, the effect of water flow intensity, river shape and environment (i.e., geomorphology) are known to differ regionally and affect the distribution of genetic diversity (Morrissey and de Kerckhove 2009; Albert *et al.*, 2011; Paz-Vinas *et al.,* 2015; Thomaz *et al.*, 2016), which make them potentially interesting to explore with respect to paleodrainages. However, this would require new developments, as with recent advances for incorporating environmental variables to study the effects of current river properties (e.g., Domisch *et al.*, 2015). The impacts of such methodological developments are likely to extend beyond, deepening our knowledge of the effect of shifts in riverine properties over time.

Insights into species diversification of freshwater fish

Vicariance plays a clear role in structuring species diversity patterns of riverine fish, reflecting a life history constrained to the rivers that predisposes fish in particular to becoming geographic isolated (Lundberg *et al*., 2000; Albert *et al.*, 2011). Nonetheless, dispersal is also recognized to play a role in shaping richness patterns. Specifically, the distribution of fish species across multiple basins may be explained by: (i) river captures in which a river tributary changes its direction and start flowing to the neighbor basin, or (ii) dispersal associated with temporary connections.

As our study (e.g., Figures 2 and 3) and others show (e.g., Chakona *et al*., 2013; Unmack *et al*., 2013, Thomaz *et al*., 2015), dispersal associated with temporary connections that were forged between currently isolated rivers in past drainages (i.e., paleodrainages) when sea-levels

repeatedly decreased may contribute to the spatial structuring and timing of divergence. Nonetheless, it might be argued that this mechanism (i.e., dispersal across drainages via past connections that opened during periods of low sea level) may be relatively species-specific (Waters and Burridge, 2016) unlike river capture and vicariance, which tends to affect communities as a whole (Burridge *et al.*, 2007; Albert *et al.*, 2011). For example, *Hollandichthys* is associated with the presence of riparian forest (Bertaco and Malabarba, 2013), and consequently is distributed in lower land tributaries, which might make downstream dispersal more likely during the cycles of sea-level retreat, given the geographic proximity to the temporary river connections that existed among drainages in the past. However, for fish inhabiting different portions of the rivers (i.e., headwaters, as opposed to lowland tributaries), such temporary connections forged by sea-level retreat might not have been accessible. If such divergence processes act in a species-specific manner, these temporary connections might be helpful to explain differences in species diversity across landscapes (i.e., discord across taxa), and consideration of the species-specific ecologies might explain why the geographic distribution of particular constituents of the ichthyofauna may differ (Waters and Burridge, 2016).

Although the links between the processes structuring genetic variation within species to those structuring species diversity patterns can be tenuous (see Kisel and Barraclough, 2010; Rosenblum *et al.*, 2012; Papadopoulou and Knowles, 2017), there are some noteworthy parallels, but also discordances, between our findings and species diversity studies in freshwater fishes (see Fourtune *et al.*, 2016; Vellend and Geber, 2005). For example, genetic diversity does not only reflects drainage area (Figure IV.4), but species richness-area relationships have been largely observed for current and past drainages over the world and for the study region, the Neotropics (Albert *et al.*, 2011; Dias *et al*., 2014). On the other hand, our focus here was on the recent evolutionary past, and this temporal scale does not correspond to the divergence times estimated for fish species diversification, which often predates the Pleistocene (Lundberg *et al*., 2000). This does not necessarily mean that other mechanisms did not contribute to species diversity patterns in the more distant past. However, because the sea level changes during the LGM were some of the most extreme events and temporally matches with most of the point estimates for divergence times, these recent events would over-ride divergences associated with the more distant past (see Papadopoulou and Knowles, 2015) if the geography of such

divergence patterns were generally coincident with those of the LGM (for an exception, we note the regional split between the northern and southern regions, Figure C.3, which does not coincide with the boundary of recognized paleodrainages; see also Thomaz *et al*., 2015). This argument is also predicated on the presumed importance of divergence associated with geographic isolation, and it does not address whether other evolutionary processes (e.g., selection) might have played more or less of an important role in the past relative to the present.

Of the populations of *Hollandichthys* studied here, individuals from paleodrainage 11 (Figure IV.1) have recently been described as a putative new species, *H. taramandahy* (Bertaco and Malabarba, 2013). The strong structuring of genetic variation in *Hollandichthys* may be indicative of a putative species boundary, and consequently, suggest that paleodrainages may be responsible for long-term isolation that culminates in speciation. However, the degree of genomic differentiation for this putative species is similar to those observed between the populations from other paleodrainages in *Hollandichthys*, as is estimates of its divergence (i.e., approximately 24ky, Figure IV.3). It is unknown if any of a set of geographically isolated regions/populations will become new species (see Sukumaran and Knowles, 2017), but the clear morphological characters (Bertaco and Malabarba, 2013) of the proposed new taxon may suggest that this lineage has preceded beyond what might be expected from geographic isolation at the microevolutionary level (i.e., population isolation; see Rosemblum *et al.*, 2012). Further analyses that tests for morphological differences across individuals in the other paleodrainages are required to determine whether the differentiation observed in the new putative species *H. taramandahy* (Bertaco and Malabarba, 2013) is statistically equivalent to other divergences separating paleodrainage populations that are not recognized as different species (e.g., e.g., Solis-Lemus *et al.*, 2014; Huang and Knowles, 2016).

Conclusions

Our study not only highlights the effect of Pleistocene paleodrainages on patterns of genetic variation in a freshwater fish species along basins of the Brazilian Atlantic coast, but the findings also may help explain why support for paleodrainages as drivers of divergence across taxa and continents have not been consistent. Specifically, the properties that impact population isolation and connectivity in riverine systems may be linked to those of past paleodrainages, not necessarily the current landscape. Given these phenomena occur over short evolutionary time

scales, we also highlight how biases in the test applied and/or interpretation of results can contribute to ambiguities regarding the effects of past river landscapes, as well as how the development of new tools for modeling riverine environments that parallel those from the terrestrial realm will promote more refined hypotheses that could help explain differences in genetic variation across regions and/or taxa.

References

Albert JS, Petry P, Reis RE (2011). Major biogeographic and phylogenetic patterns. In: Albert JS, Reis RE (eds) *Historical biogeography of Neotropical freshwater fishes*, University of California Press: Berkley. pp 21-57.

Bertaco VA, Malabarba LR (2013). A new species of the characid genus *Hollandichthys* Eigenmann from coastal rivers of southern Brazil (Teleostei: Characiformes) with a discussion on the diagnosis of the genus. *Neotrop Ichthyol* **11**: 767-778.

Burridge, C. P., Craw, D., & Waters, J. M. (2007). An empirical test of freshwater vicariance via river capture. *Mol Ecol* **16**: 1883-1895.

Carstens BC, Knowles LL (2007). Shifting distributions and speciation: species divergence during rapid climate change. *Mol Ecol* **16**:619-627.

Catchen J, Hohenlohe P, Bassham S, Amores A, Cresko W (2013). Stacks: an analysis tool set for population genomics. *Mol Ecol* **22**: 3124-3140.

Chakona A, Swartz ER, Gouws G (2013). Evolutionary drivers of diversification and distribution of a southern temperate stream fish assemblage: testing the role of historical isolation and spatial range expansion. *PloS one* **8:** 1-13.

Dias MS, Oberdorff T, Hugueny B, Leprieur F, Jézéquel C, Cornu JF, *et al*. (2014). Global imprint of historical connectivity on freshwater fish biodiversity. *Ecol Lett* **17**: 1130-1140.

Domisch S, Amatulli G, Jetz W (2015). Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. *Scient Data*, ***2***.

Earl DA, vonHoldt BM (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* **4**: 359-361.

Edwards S, Beerli P (2000). Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54**: 1839-1854.

Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611-2620.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013). Robust demographic inference from genomic and SNP data. *PLoS Genet* **9**: 1-17.

Excoffier L, Foll M (2011). Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinform* **27**: 1332-1334.

Excoffier L, Lischer HE (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* **10**: 564-567

Fourtune L, Paz-Vinas I, Loot G, Prunier JG, Blanchet S (2016). Lessons from the fish: a multi-species analysis reveals common processes underlying similar species-genetic diversity correlations. *Freshwater Biol* **61**: 1830-1845.

He Q, Edwards DL, Knowles LL (2013). Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, **67**: 3386-3402.

Huang H, Knowles LL (2014). Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Syst Biol* **65**: 357-365.

Huang JP, Knowles LL (2016) The species versus subspecies conundrum: quantitative delimitation from integrating multiple data types within a single Bayesian approach in Hercules beetles. *Syst. Biol.*65: 685-699.

Hudson RR, Coyne JA (2002). Mathematical consequences of the genealogical species concept. *Evolution* **56**: 1557-1565.

Kisel Y, Barraclough TG (2010) Speciation has a spatial scale that depends on levels of gene flow. *Am Nat* **175**: 316–334

Knowles LL (2009). Statistical phylogeography. *Ann Rev Ecol Evol Syst* **40**: 593-612.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* **15**: 1179-1191.

Lundberg JG, Kottelat M, Smith GR, Stiassny ML, Gill AC (2000). So many fishes, so little time: an overview of recent ichthyological discovery in continental waters. *Ann Mo Bot Gard* **87**: 26-62.

Lynch M (2010). Evolution of the mutation rate. *Trends Genet* **26**: 345-352.

Massatti R, Knowles LL (2014). Microhabitat differences impact phylogeographic concordance of codistributed species: genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. *Evolution* **68**: 2833-2846.

Massatti R, Knowles LL (2016). Contrasting support for alternative models of genomic variation based on microhabitat preference: species-specific effects of climate change in alpine sedges. *Mol Ecol* **25**:3974-3986.

Mazerolle, MJ (2016). AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). R package version 2.0-4. http://CRAN.R-project.org/package=AICcmodavg.

Medrano JF, Aasen E, Sharrow L (1990). DNA extraction from nucleated red blood cells. *Biotechniques*, **8**.

Morrissey MB, de Kerckhove DT (2009). The maintenance of genetic variation due to asymmetric gene flow in dendritic metapopulations. *Am Nat* **174**: 875-889.

Neuenschwander S, Largiadèr CR, Ray N, Currat M, Vonlanthen P, Excoffier L (2008). Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Mol Ecol* **17**: 757-772.

Papadopoulou A, Knowles LL (2015). Genomic tests of the species-pump hypothesis: recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. *Evolution* **69**: 1501-1517.

Papadopoulou A, Knowles LL (2016). Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *PNAS* **113**: 8018-8024.

Papadopoulou A, Knowles LL (2017). Linking micro- and macroevolutionary perspectives to evaluate the role of Quaternary sea-level oscillations in island diversification. *Evolution* in review.

Parchman TL, Gompert Z, Mudge J, Schilkey FD, Benkman CW, Buerkle C (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol Ecol* **21**: 2991-3005.

Paz-Vinas I, Loot G, Stevens VM, Blanchet S (2015). Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Mol Ecol* **24**: 4586-4604.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559-575.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rosenblum EB, Sarver BA, Brown JW, Des Roches S, Hardwick KM, Hether TD *et al*. (2012). Goldilocks meets Santa Rosalia: an ephemeral speciation model explains patterns of diversification across time scales. *Evoly Biol* **39**: 255-261.

Ryan PG, Bloomer P, Moloney CL, Grant TJ, Delport W (2007). Ecological speciation in South Atlantic island finches. *Science* **315**: 1420-1423.

Solis-Lemus C*,* Knowles LL, Ané C (2014). Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution* **69**:492-507.

Sukumaran J, Knowles LL (2017). The multiple coalescent delimits structure not species. *PNAS* **144**: 1607-1612.

Tedesco PA, Leprieur F, Hugueny B, Brosse S, Dürr HH, Beauchard O, *et al*. (2012). Patterns and processes of global riverine fish endemism. *Glob Ecol Biogeogr* **21**: 977-987.

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

Thomaz AT, Malabarba LR, Bonatto SL (2010). The phylogenetic placement of *Hollandichthys* Eigenmann 1909 (Teleostei: Characidae) and related genera. *Mol Phyl Evol* **57**: 1347-1352.

Thomaz AT, Malabarba LR, Bonatto SL, Knowles LL (2015). Testing the effect of palaeodrainages versus habitat stability on genetic divergence in riverine systems: study of a Neotropical fish of the Brazilian coastal Atlantic Forest. *J Biogeogr* **42**: 2389-2401.

Unmack PJ, Hammer MP, Adams M, Johnson JB, Dowling TE (2013). The role of continental shelf width in determining freshwater phylogeographic patterns in south-eastern Australian pygmy perches (Teleostei: Percichthyidae). *Mol Ecol* **22**: 1683-1699.

Vellend M, Geber MA (2005). Connections between species diversity and genetic diversity. *Ecology letters*, **8**: 767-781.

Waters JM, Burridge CP (2016). Fine-scale habitat preferences influence within-river population connectivity: a case-study using two sympatric New Zealand Galaxias fish species. *Freshwater Biol*, **61**: 51-56.

APPENDIX C

Supplementary material from chapter IV

Supplementary Text

Inference of paleodrainages

Palaeodrainages during the Last Glacial Maximum (LGM) were inferred from topographical and bathymetric information extracted from the digital elevation model (DEM) GEBCO_08 at 30 arc-second resolution (c. 1 km; http://www.gebco.net/) in ArcGIS10 using Hydrological tools. Past connections between current riverine basins and palaeodrainages were identified from the inferred flow directions based on topographical relief. Specifically, land area exposed during the LGM (i.e., a maximum shift in sea level of -125 m) was identified using the tool *Contour* followed by *Mask*, and the Fill option was used to cover localized depressions in the surface. The *Flow Direction* tool was then used to identify the steepest descent from each cell, followed by the *Basin* tool to identify ridges between the basins that delineate the basins' borders. For further details see Thomaz *et al*., 2015.

Table C.1. Sampling and genomic sequences per individual pre- and post-processing in STACKS; * marks individuals removed from the analysis because of poor sequence quality, large numbers of missing loci, or a single individual in that population. Note that populations associated to the Paranaguá estuary (i.e., Guaraqueçaba and Paranaguá) were not included in analyses because of ambiguities with paleodrainage assignment.

Population	Voucher	Locality	Longitude	Latitude	Raw read count	Utilized reads	Polymorphic loci
Paraty	UFRGS 11776/TEC909 A	Paraty, RJ	-44.5950	-23.0419	1098168	900046	34705
Paraty	UFRGS 11776/TEC909 B	Paraty, RJ	-44.5950	-23.0419	726529	618667	19976
Paraty	UFRGS 11776/TEC909 C	Paraty, RJ	-44.5950	-23.0419	2868700	2542999	59831
Paraty*	UFRGS 11776/TEC909 D	Paraty, RJ	-44.5950	-23.0419	620526	449556	18025
Paraty	UFRGS 11776/TEC909 E	Paraty, RJ	-44.5950	-23.0419	1310075	1110254	39533
Paraty	UFRGS 11776/TEC909 F	Paraty, RJ	-44.5950	-23.0419	2099253	1827632	54146
Paraty	UFRGS 11776/TEC909 G	Paraty, RJ	-44.5950	-23.0419	1389264	1078034	39152
Paraty	UFRGS 11776/TEC909 H	Paraty, RJ	-44.5950	-23.0419	1464882	1206051	37498
Paraty	UFRGS 11776/TEC909 I	Paraty, RJ	-44.5950	-23.0419	2409686	2221319	28161
Paraty	UFRGS 11776/TEC909 J	Paraty, RJ	-44.5950	-23.0419	2496914	2229227	55359
Paraty 2*	MCP30665	Paraty, RJ	-44.6975	-23.0764	970498	703921	-
Toca do Boi	UFRGS 11775/TEC908 A	Paraty, RJ (Toca do Boi)	-44.6819	-23.3292	1453427	1185664	38351
Toca do Boi	UFRGS 11775/TEC908 B	Paraty, RJ (Toca do Boi)	-44.6819	-23.3292	1574027	1333708	45032
Toca do Boi	UFRGS 11775/TEC908 C	Paraty, RJ (Toca do Boi)	-44.6819	-23.3292	1666329	1443786	40977
Toca do Boi	UFRGS 11775/TEC908 D	Paraty, RJ (Toca do Boi)	-44.6819	-23.3292	1132355	882935	32063
Toca do Boi	UFRGS 11775/TEC908 E	Paraty, RJ (Toca do Boi)	-44.6819	-23.3292	1682263	1366221	46376
Toca do Boi	UFRGS 11775/TEC908 F	Paraty, RJ (Toca do Boi)	-44.6819	-23.3292	1361996	1107613	39835
Toca do Boi	UFRGS 11775/TEC908 G	Paraty, RJ (Toca do Boi)	-44.6819	-23.3292	1434266	1118482	41791
Toca do Boi	UFRGS 11775/TEC908 H	Paraty, RJ (Toca do Boi)	-44.6819	-23.3292	2040740	1742324	54556
Toca do Boi	MCP30664	Paraty, RJ (Toca do Boi)	-44.6819	-23.3292	1484778	1165930	46045
Toca do Boi	MCP30664	Paraty, RJ (Toca do Boi)	-44.6819	-23.3292	1572158	1206032	53987
Ubatuba 1	UFRGS 11789/TEC866 A	Ubatuba, SP	-44.8706	-23.3467	2429300	1984838	49603
Ubatuba 1	UFRGS 11789/TEC866 B	Ubatuba, SP	-44.8706	-23.3467	1074156	793710	27320
Ubatuba 2	MCP30663	Ubatuba, SP	-45.1286	-23.4278	1480671	1083984	51420
Ubatuba 2	UFRGS 11774/TEC900 A	Ubatuba, SP	-45.1144	-23.4125	1321716	935442	33954
Ubatuba 2	UFRGS 11774/TEC900 B	Ubatuba, SP	-45.1144	-23.4125	2151002	1915290	55226
Ilhabela	MCP30661	Ilhabela, RJ	-45.3633	-23.8250	1318964	1074290	39392
Ilhabela	MCP30661	Ilhabela, RJ	-45.3633	-23.8250	3161748	2944413	60933
Ilhabela	MCP30662	Ilhabela, RJ	-45.3589	-23.8200	1880301	1595523	42626
Ilhabela	MCP30662	Ilhabela, RJ	-45.3589	-23.8200	1732334	1455896	51335
Ilhabela*	UFRGS 11773/TEC897 C	Ilhabela, RJ	-45.3536	-23.8214	190840	170846	-
Ilhabela	UFRGS 11773/TEC897 D	Ilhabela, RJ	-45.3536	-23.8214	1095256	778180	22161
Ilhabela	UFRGS 11773/TEC897 E	Ilhabela, RJ	-45.3536	-23.8214	1894348	1618334	52875
Ilhabela	UFRGS 11773/TEC897 F	Ilhabela, RJ	-45.3536	-23.8214	775095	504082	18669
Ilhabela*	UFRGS 11773/TEC897 G	Ilhabela, RJ	-45.3536	-23.8214	604004	52725	-

São Sebastião 1	UFRGS 11795/TEC1229 A	São Sebastião, SP	-45.6064	-23.7733	1998679	1729076	35868
São Sebastião 1	UFRGS 11795/TEC1229 B	São Sebastião, SP	-45.6064	-23.7733	2850510	2373213	49884
São Sebastião 1	UFRGS 11795/TEC1229 C	São Sebastião, SP	-45.6064	-23.7733	3256867	2953768	53940
São Sebastião 1	UFRGS 11795/TEC1229 D	São Sebastião, SP	-45.6064	-23.7733	2501502	2161569	48853
São Sebastião 1	UFRGS 11795/TEC1229 E	São Sebastião, SP	-45.6064	-23.7733	3023378	2681668	55260
São Sebastião 1	UFRGS 11795/TEC1229 F	São Sebastião, SP	-45.6064	-23.7733	3416573	3185316	52682
São Sebastião 1	UFRGS 11795/TEC1229 G	São Sebastião, SP	-45.6064	-23.7733	1479360	1176663	38548
São Sebastião 1	UFRGS 11795/TEC1229 H	São Sebastião, SP	-45.6064	-23.7733	1109731	867293	29378
São Sebastião 1	UFRGS 11795/TEC1229 I	São Sebastião, SP	-45.6064	-23.7733	1909069	1689597	37167
São Sebastião 1	UFRGS 11795/TEC1229 J	São Sebastião, SP	-45.6064	-23.7733	2352458	2146528	29691
São Sebastião 2*	MCP30658	São Sebastião, SP	-45.7144	-23.7617	8339	4351	-
São Sebastião 2	UFRGS 11772/TEC891 A	São Sebastião, SP	-45.7203	-23.7628	2300416	2131299	20896
São Sebastião 2	UFRGS 11772/TEC891 B	São Sebastião, SP	-45.7203	-23.7628	2585949	2107327	48106
São Sebastião 2*	UFRGS 11772/TEC891 C	São Sebastião, SP	-45.7203	-23.7628	224285	193786	-
São Sebastião 2*	UFRGS 11772/TEC891 D	São Sebastião, SP	-45.7203	-23.7628	490259	421046	-
São Sebastião 2*	UFRGS 11772/TEC891 E	São Sebastião, SP	-45.7203	-23.7628	21494	10027	-
São Sebastião 2	UFRGS 11772/TEC891 G	São Sebastião, SP	-45.7203	-23.7628	3404274	3089411	55153
São Sebastião 2	UFRGS 11772/TEC891 H	São Sebastião, SP	-45.7203	-23.7628	2675823	2351809	55137
São Sebastião 2	UFRGS 11772/TEC891 I	São Sebastião, SP	-45.7203	-23.7628	2901160	2660446	28722
São Sebastião 3*	MCP30660	São Sebastião, SP	-45.5519	-23.7894	1219213	864402	-
São Sebastião 4*	UFRGS 11788/TEC859 A	São Sebastião, SP	-45.4522	-23.8239	1165919	603275	-
São Sebastião 4*	UFRGS 11788/TEC859 B	São Sebastião, SP	-45.4522	-23.8239	1201194	599398	-
Bertioga*	UFRGS 11771/TEC888 C	Bertioga, SP	-46.0031	-23.7794	199120	186254	-
Bertioga	UFRGS 11771/TEC888 D	Bertioga, SP	-46.0031	-23.7794	1702483	1340406	44758
Bertioga	UFRGS 11771/TEC888 E	Bertioga, SP	-46.0031	-23.7794	3797857	3466754	48290
Bertioga	UFRGS 11771/TEC888 F	Bertioga, SP	-46.0031	-23.7794	2530946	2115001	55958
Bertioga*	UFRGS 11771/TEC888 G	Bertioga, SP	-46.0031	-23.7794	66882	6311	-
Bertioga	UFRGS 11777/TEC980 B	Bertioga, SP	-46.0031	-23.7794	2935697	2550960	59988
Bertioga	UFRGS 11777/TEC980 C	Bertioga, SP	-46.0031	-23.7794	1770554	1410510	44228
Bertioga	UFRGS 11777/TEC980 D	Bertioga, SP	-46.0031	-23.7794	1856994	1445558	49637
Bertioga	UFRGS 11777/TEC980 E	Bertioga, SP	-46.0031	-23.7794	2460277	1934988	53758
Bertioga*	UFRGS 11777/TEC980 G	Bertioga, SP	-46.0031	-23.7794	23591	10754	-
Upper Tietê	UFRGS 11786/TEC826 A	Paranapiacaba, SP	-46.3347	-23.7664	804503	637740	22305
Upper Tietê	UFRGS 11786/TEC826 B	Paranapiacaba, SP	-46.3347	-23.7664	1239103	996612	35240
Upper Tietê	UFRGS 11786/TEC826 C	Paranapiacaba, SP	-46.3347	-23.7664	1312201	937231	43466
Upper Tietê	UFRGS 11786/TEC826 D	Paranapiacaba, SP	-46.3347	-23.7664	1816278	1383835	58940
Upper Tietê	UFRGS 11786/TEC826 E	Paranapiacaba, SP	-46.3347	-23.7664	1317731	939323	51208
Upper Tietê	UFRGS 11787/TEC827 A	Paranapiacaba, SP	-46.3136	-23.7725	1721145	1281748	63759
Upper Tietê	UFRGS 11787/TEC827 B	Paranapiacaba, SP	-46.3136	-23.7725	2209345	2032282	50874
Upper Tietê	UFRGS 11787/TEC827 C	Paranapiacaba, SP	-46.3136	-23.7725	1340661	943311	56496

Upper Tietê	UFRGS 11787/TEC827 D	Paranapiacaba, SP	-46.3136	-23.7725	2512031	2432574	35786
Upper Tietê	UFRGS 11787/TEC827 E	Paranapiacaba, SP	-46.3136	-23.7725	1651191	1233524	52644
Santos	MCP30560	Cubatão, SP	-46.3281	-23.8425	1652764	1358598	47102
Santos	MCP30560	Cubatão, SP	-46.3281	-23.8425	934064	760014	30578
Santos	MCP30559	Cubatão, SP	-46.3483	-23.8564	1410664	1125864	49925
Santos	MCP30559	Cubatão, SP	-46.3483	-23.8564	1569792	1233696	55180
Santos	UFRGS 11785/TEC820 A	Santos, SP	-46.3489	-23.8569	1428856	1050598	44907
Santos	UFRGS 11785/TEC820 B	Santos, SP	-46.3489	-23.8569	1473939	1090577	48344
Santos	UFRGS 11785/TEC820 C	Santos, SP	-46.3489	-23.8569	2490488	2222585	49345
Santos	UFRGS 11785/TEC820 D	Santos, SP	-46.3489	-23.8569	1379889	1059057	43971
Santos	UFRGS 11785/TEC820 E	Santos, SP	-46.3489	-23.8569	1670347	1282315	56363
Santos	UFRGS 11785/TEC820 F	Santos, SP	-46.3489	-23.8569	1606014	1219205	52808
Peruíbe	MCP30554	Peruíbe, SP	-47.0372	-24.3597	1519051	1159490	33328
Peruíbe	MCP30555	Peruíbe, SP	-47.0547	-24.3681	1291554	940643	47835
Peruíbe	MCP30561	Peruíbe, SP	-47.0361	-24.3586	1912939	1559830	48793
Peruíbe	UFRGS 11783/TEC817 A	Peruíbe, SP	-47.0372	-24.3600	1289234	1028955	32102
Peruíbe	UFRGS 11783/TEC817 C	Peruíbe, SP	-47.0372	-24.3600	1219687	859228	46229
Peruíbe*	UFRGS 11783/TEC817 E	Peruíbe, SP	-47.0372	-24.3600	10481	2863	-
Peruíbe*	UFRGS 11784/TEC722 A	Peruíbe, SP	-47.0550	-24.3683	4449	2478	-
Peruíbe	UFRGS 11784/TEC722 B	Peruíbe, SP	-47.0550	-24.3683	1556261	1114347	60164
Peruíbe	UFRGS 11784/TEC722 D	Peruíbe, SP	-47.0550	-24.3683	1530139	1145238	48197
Peruíbe	UFRGS 11784/TEC722 E	Peruíbe, SP	-47.0550	-24.3683	1598988	1256610	56475
Ribeira de Iguape	UFRGS 11781/TEC806 A	Iguape, SP	-47.4931	-24.6614	2077778	1583210	64526
Ribeira de Iguape	UFRGS 11781/TEC806 B	Iguape, SP	-47.4931	-24.6614	1570785	1305293	40490
Ribeira de Iguape	UFRGS 11781/TEC806 C	Iguape, SP	-47.4931	-24.6614	1514447	1226833	37948
Ribeira de Iguape*	UFRGS 11781/TEC806 D	Iguape, SP	-47.4931	-24.6614	673196	449983	-
Ribeira de Iguape	UFRGS 11782/TEC739 A	Iguape, SP	-47.4836	-24.6500	1644564	1264159	44018
Ribeira de Iguape	UFRGS 11782/TEC739 B	Iguape, SP	-47.4836	-24.6500	1572629	1234041	44278
Guaraqueçaba*	MCP30558	Guaraqueçaba, PR	-48.4336	-25.2092	1570277	1277232	-
Guaraqueçaba*	UFRGS 11779/TEC763 A	Guaraqueçaba, PR	-48.4203	-25.1739	1331361	938170	-
Guaraqueçaba*	UFRGS 11779/TEC763 B	Guaraqueçaba, PR	-48.4203	-25.1739	1343074	995952	-
Guaraqueçaba*	UFRGS 11779/TEC763 C	Guaraqueçaba, PR	-48.4203	-25.1739	1833265	1494428	-
Guaraqueçaba*	UFRGS 11779/TEC763 D	Guaraqueçaba, PR	-48.4203	-25.1739	1841856	1578082	-
Guaraqueçaba*	UFRGS 11778/TEC755 A	Tagaçaba, PR	-48.4500	-25.2200	1365569	1097390	-
Guaraqueçaba*	UFRGS 11778/TEC755 B	Tagaçaba, PR	-48.4500	-25.2200	1541716	1171014	-
Guaraqueçaba*	UFRGS 11778/TEC755 C	Tagaçaba, PR	-48.4500	-25.2200	713284	392307	-
Guaraqueçaba*	UFRGS 11778/TEC755 D	Tagaçaba, PR	-48.4500	-25.2200	1306469	977174	-
Guaraqueçaba*	UFRGS 11778/TEC755 E	Tagaçaba, PR	-48.4500	-25.2200	1301434	985575	-
Paranaguá*	MHNCI ncat	Antonina, PR	-48.7000	-25.4333	1278109	935893	-
Paranaguá*	MHNCI ncat	Antonina, PR	-48.7000	-25.4333	1021488	706960	-

Paranaguá*	MHNCI ncat	Antonina, PR	-48.7000	-25.4333	1235019	906103	-
Paranaguá*	MHNCI ncat	Antonina, PR	-48.7000	-25.4333	1200292	807300	-
Paranaguá*	MHNCI ncat	Antonina, PR	-48.7000	-25.4333	1339796	923487	-
S. Franc. Sul	MCP30553	Ilha S. Franc. Sul, SC	-48.6511	-26.3267	1358622	1041275	41629
S. Franc. Sul*	UFRGS 10579/TEC105	Ilha S. Franc. Sul, SC	-48.6511	-26.3267	455531	247839	-
S. Franc. Sul	UFRGS 9359/TEC345	Ilha S. Franc. Sul, SC	-48.5892	-26.2930	1484971	1257981	41183
S. Franc. Sul	UFRGS 10570/TEC106 A	Ilha S. Franc. Sul, SC	-48.5892	-26.2930	2738799	2506161	45275
S. Franc. Sul	UFRGS 10570/TEC106 B	Ilha S. Franc. Sul, SC	-48.5892	-26.2930	3186947	2892456	58120
S. Franc. Sul	UFRGS 10570/TEC106 C	Ilha S. Franc. Sul, SC	-48.5892	-26.2930	1525797	1326424	26844
S. Franc. Sul	UFRGS 10570/TEC106 D	Ilha S. Franc. Sul, SC	-48.5892	-26.2930	2686698	2495012	50451
S. Franc. Sul	UFRGS 10570/TEC106 E	Ilha S. Franc. Sul, SC	-48.5892	-26.2930	2700678	2429372	43223
S. Franc. Sul	UFRGS 10570/TEC106 F	Ilha S. Franc. Sul, SC	-48.5892	-26.2930	2708515	2440012	40768
S. Franc. Sul	UFRGS 10570/TEC106 G	Ilha S. Franc. Sul, SC	-48.5892	-26.2930	2541149	2313342	47087
Babitonga	MCP30552	Araquari, SC	-48.7258	-26.3828	1209118	946763	38190
Babitonga	MCP30667	Joinville, SC	-48.9289	-26.1919	1272942	994028	41098
Guaratuba	UFRGS 10578/TEC109	Guaratuba, PR	-48.5833	-25.8833	2158302	1857011	53492
Guaratuba	UFRGS 9358/TEC346	Guaratuba, PR	-48.5833	-25.8833	1267302	956980	47120
Florianópolis 1	MCP28747	Florianópolis, SC	-48.4925	-27.5086	1353998	965540	47645
Florianópolis 1	MCP28732	Florianópolis, SC	-48.4864	-27.5111	1839766	1628125	48753
Florianópolis 1	MCP28732	Florianópolis, SC	-48.4864	-27.5111	1438780	1085898	51650
Florianópolis 1	MCP38317	Florianópolis, SC	-48.4372	-27.4831	1355850	962147	45766
Florianópolis 2	MCP37635	Florianópolis, SC	-48.4231	-27.4831	812468	604630	22296
Florianópolis 2	MCP37635	Florianópolis, SC	-48.4231	-27.4831	851942	507815	19579
Florianópolis 2	MCP37635	Florianópolis, SC	-48.4231	-27.4831	1098851	871131	33719
Florianópolis 3	MCP28737	Florianópolis, SC	-48.4775	-27.5881	1571100	1155499	56042
Florianópolis 3	MCP28737	Florianópolis, SC	-48.4775	-27.5881	1500857	1148858	40702
Florianópolis 3	MCP28737	Florianópolis, SC	-48.4775	-27.5881	1128815	909977	32920
Garopaba	UFRGS 16587/TEC2919 A	Garopaba, SC	-48.7005	-28.0715	2783833	2469992	55060
Garopaba	UFRGS 16587/TEC2919 B	Garopaba, SC	-48.7005	-28.0715	1774893	1545472	31964
Garopaba	UFRGS 16587/TEC2919 C	Garopaba, SC	-48.7005	-28.0715	1489435	1230239	38200
Garopaba	UFRGS 16587/TEC2919 D	Garopaba, SC	-48.7005	-28.0715	2483493	2231984	30321
Garopaba	UFRGS 16587/TEC2919 E	Garopaba, SC	-48.7005	-28.0715	2135018	1850607	38721
Garopaba	UFRGS 16587/TEC2919 F	Garopaba, SC	-48.7005	-28.0715	1296886	1108358	24878
Garopaba	UFRGS 16587/TEC2919 G	Garopaba, SC	-48.7005	-28.0715	1062519	836586	36259
Garopaba	UFRGS 16587/TEC2919 H	Garopaba, SC	-48.7005	-28.0715	1101909	929652	22307
Araranguá*	UFRGS 11791/TEC840 A	Araranguá, SC	-49.9289	-28.7908	407508	227797	-
Araranguá	UFRGS 11791/TEC840 B	Araranguá, SC	-49.9289	-28.7908	1170918	964684	30093
Araranguá	UFRGS 11791/TEC840 C	Araranguá, SC	-49.9289	-28.7908	1466012	1256026	40385
Araranguá	UFRGS 11791/TEC840 D	Araranguá, SC	-49.9289	-28.7908	1463739	1159837	35706
Mampituba	MCP23625	Mampituba, SC	-49.9969	-29.1703	1688056	1300026	56730

Mampituba	UFRGS 11790/TEC839	Mampituba, SC	-49.9822	-29.1692	1628963	1381686	44281
Mampituba	UFRGS 11792/TEC841 A	Mampituba, SC	-50.1167	-29.2528	2271189	2019993	49734
Mampituba	UFRGS 11792/TEC841 B	Mampituba, SC	-50.1167	-29.2528	2993304	2662523	65269
Mampituba	UFRGS 11792/TEC841 C	Mampituba, SC	-50.1167	-29.2528	1701233	1480050	44906
Mampituba	UFRGS 11792/TEC841 D	Mampituba, SC	-50.1167	-29.2528	1385623	1156003	39969
Mampituba	UFRGS 11792/TEC841 E	Mampituba, SC	-50.1167	-29.2528	1814810	1516244	48066
Mampituba	UFRGS 11792/TEC841 F	Mampituba, SC	-50.1167	-29.2528	1749054	1464074	45636
Mampituba	UFRGS 11792/TEC841 G	Mampituba, SC	-50.1167	-29.2528	1405917	1095388	38826
Mampituba	UFRGS 11792/TEC841 H	Mampituba, SC	-50.1167	-29.2528	908667	709308	28043
Três Forquilhas	UFRGS 16513/TEC2841 A	Itati, RS	-50.1789	-29.4267	1128460	966580	21091
Três Forquilhas	UFRGS 16513/TEC2841 B	Itati, RS	-50.1789	-29.4267	1234567	931387	39848
Três Forquilhas	UFRGS 16513/TEC2841 C	Itati, RS	-50.1789	-29.4267	1272236	1070280	28453
Três Forquilhas	UFRGS 16513/TEC2841 D	Itati, RS	-50.1789	-29.4267	1019104	815821	26866
Três Forquilhas	UFRGS 16513/TEC2841 E	Itati, RS	-50.1789	-29.4267	974598	707924	34658
Três Forquilhas	UFRGS 16513/TEC2841F	Itati, RS	-50.1789	-29.4267	1121395	894136	34944
Três Forquilhas	UFRGS 16513/TEC2841 G	Itati, RS	-50.1789	-29.4267	1352268	1025133	41640
Três Forquilhas	UFRGS 16513/TEC2841 H	Itati, RS	-50.1789	-29.4267	1286903	1050800	26655
Maquiné	MCP26969	Maquiné, RS	-50.2833	-29.5833	1196972	854115	48981
Maquiné	UFRGS 11793/TEC842 A	Maquiné, RS	-50.2933	-29.5906	1823115	1510955	46622
Maquiné	UFRGS 11793/TEC842 B	Maquiné, RS	-50.2933	-29.5906	1267088	1087900	34763
Maquiné	UFRGS 11793/TEC842 C	Maquiné, RS	-50.2933	-29.5906	2054826	1733752	49519
Maquiné	UFRGS 11793/TEC842 D	Maquiné, RS	-50.2933	-29.5906	1922076	1692089	49124
Maquiné	UFRGS 11793/TEC842 E	Maquiné, RS	-50.2933	-29.5906	2375671	2065855	48407
Maquiné	UFRGS 11793/TEC842 F	Maquiné, RS	-50.2933	-29.5906	2542248	1678713	39745
Maquiné	UFRGS 11793/TEC842 G	Maquiné, RS	-50.2933	-29.5906	1976228	1683519	52671
Maquiné	UFRGS 11793/TEC842 H	Maquiné, RS	-50.2933	-29.5906	2039575	1780893	51597
Maquiné	UFRGS 11793/TEC842 I	Maquiné, RS	-50.2933	-29.5906	2178200	1941389	29609

Table C.2. Population genetic summary statistics for each of the 23 sampled populations, which are listed in order of the northern most populations (PAR) to the southern most population (MAQ)(see Figure 1). Different paleodrainages are demarcated sequentially by white and grey shading. The number of individuals retained per population (N_R) after processing of genomic data and which were used to calculate the reported summary statistics are given, as well as the number of individuals used in STRUCTURE analyses based on a maximum of 25% missing data (N_{25}). Summary statistics are presented only for polymorphic sites, and include average nucleotide diversity, π, average observed heterozygosity per locus, H_{OBS}, and average expected heterozygosity per locus, H_{EXP}. See Table S1 for complete list of samples and summary of genomic data collected for each individual.

Paleo drainage	Pop. Code	Populations	N_R	N_{25}	Private	π	H_{OBS}	H_{EXP}
1	PAR	Paraty	10	7	1069	0.013	0.013	0.012
2	TB	Toca do Boi	10	10	1660	0.021	0.020	0.020
3	UB1	Ubatuba 1	2	2	1001	0.038	0.034	0.028
	UB2	Ubatuba 2	3	3	1152	0.037	0.034	0.031
4	IB	Ilhabela	7	5	958	0.025	0.022	0.023
	SS1	São Sebastião 1	10	7	980	0.039	0.036	0.037
5	SS2	São Sebastião 2	5	3	549	0.032	0.028	0.028
	BER	Bertioga	7	6	2653	0.057	0.052	0.052
6	SAN	Santos	10	9	2295	0.053	0.049	0.050
	UT	Upper Tietê	10	6	638	0.039	0.037	0.037
7	PER	Peruíbe	8	6	1535	0.028	0.025	0.026
8	RI	Ribeira de Iguape	5	5	5272	0.084	0.073	0.075
9	GUA	Guaratuba	2	2	1946	0.067	0.060	0.050
	BAB	Babitonga	2	2	1403	0.070	0.057	0.053
	SFS	S. Francisco do Sul	9	7	3241	0.069	0.059	0.065
10	FL1	Florianópolis 1	4	3	2130	0.057	0.047	0.049
	FL2	Florianópolis 2	3	1	558	0.048	0.045	0.040
	FL3	Florianópolis 3	3	3	822	0.009	0.009	0.008
	GAR	Garopaba	8	2	813	0.015	0.015	0.014
11	ARA	Araranguá	3	3	982	0.016	0.016	0.013
	MAM	Mampituba	10	10	1362	0.019	0.016	0.018
	TF	Três Forquilhas	8	5	439	0.014	0.014	0.013
	MAQ	Maquiné	10	9	551	0.008	0.008	0.007

Table C.3. Pairwise F_{ST}-values between populations with comparisons between the southern and northern regions shown in grey. Bold numbers mark statistically significant values at an α = 0.05 (below the diagonal) and at an α = 0.0002 (above the diagonal) after correcting for multiple comparisons using a Bonferroni correction.

		NORTH												**SOUTH**										
		PAR	TB	UB1	UB2	IB	SS1	SS2	BER	SAN	UT	PER	RI	GUA	BAB	SFS	FL1	FL2	FL3	GAR	ARA	MAM	TF	MAQ
NORTH	PAR	-	**0.67**	0.64	0.58	0.67	**0.74**	0.76	**0.70**	**0.70**	**0.77**	**0.84**	**0.67**	0.88	0.87	**0.77**	0.93	0.85	**0.83**	0.92	**0.93**	**0.92**	**0.92**	0.95
	TB	**0.67**	-	0.53	0.47	**0.58**	**0.71**	**0.67**	**0.68**	**0.69**	**0.74**	**0.81**	0.65	0.84	0.83	**0.76**	0.90	0.83	0.78	**0.89**	0.90	**0.90**	**0.89**	**0.93**
	UB1	**0.64**	**0.53**	-	0.02	0.32	0.61	0.59	0.50	0.54	0.64	0.75	0.42	0.70	0.67	0.64	0.89	0.71	0.68	0.90	0.87	0.89	0.89	0.93
	UB2	**0.58**	**0.47**	0.02	-	0.27	0.57	0.49	0.49	0.53	0.61	0.73	0.43	0.70	0.67	0.63	0.85	0.71	0.64	0.87	0.84	0.88	0.87	0.92
	IB	**0.67**	**0.58**	**0.32**	**0.27**	-	**0.63**	0.59	0.58	**0.60**	**0.66**	**0.76**	0.54	0.79	0.77	**0.68**	0.88	0.77	0.72	**0.88**	0.87	**0.88**	**0.87**	0.92
	SS1	**0.74**	**0.71**	**0.61**	**0.57**	**0.63**	-	**0.17**	**0.12**	**0.38**	**0.46**	**0.61**	0.34	0.71	0.69	**0.62**	0.81	0.71	0.63	**0.82**	0.80	**0.83**	**0.81**	**0.87**
	SS2	**0.76**	**0.67**	0.59	**0.49**	**0.59**	**0.17**	-	-0.08	**0.17**	0.33	**0.54**	0.06	0.63	0.60	0.51	0.85	0.62	0.66	0.85	0.81	0.82	0.83	0.89
	BER	**0.70**	**0.68**	**0.50**	**0.49**	**0.58**	**0.12**	-0.08	-	**0.32**	**0.33**	**0.54**	0.30	0.62	0.59	**0.57**	0.74	0.64	0.51	**0.76**	0.74	**0.81**	**0.76**	**0.84**
	SAN	**0.70**	**0.69**	**0.54**	**0.53**	**0.60**	**0.38**	**0.17**	**0.32**	-	**0.16**	**0.49**	**0.31**	0.65	0.62	**0.59**	0.74	0.66	0.54	**0.75**	0.74	**0.80**	**0.75**	**0.83**
	UT	**0.77**	**0.74**	**0.64**	**0.61**	**0.66**	**0.46**	**0.33**	**0.33**	**0.16**	-	**0.58**	0.31	0.71	0.68	**0.60**	0.81	0.71	0.65	**0.81**	0.80	**0.83**	**0.81**	**0.87**
	PER	**0.84**	**0.81**	**0.75**	**0.73**	**0.76**	**0.61**	**0.54**	**0.54**	**0.49**	**0.58**	-	0.42	0.78	0.75	**0.67**	0.87	0.77	0.72	**0.87**	0.86	**0.88**	**0.86**	**0.91**
	RI	**0.67**	**0.65**	**0.42**	**0.43**	**0.54**	**0.34**	0.06	**0.30**	**0.31**	**0.31**	**0.42**	-	0.51	0.48	0.50	0.66	0.56	0.39	0.71	0.67	0.77	0.71	**0.81**
SOUTH	GUA	**0.88**	**0.84**	0.70	0.70	**0.79**	**0.71**	**0.63**	**0.62**	**0.65**	**0.71**	**0.78**	**0.51**	-	0.09	0.08	0.69	0.37	0.21	0.72	0.69	0.77	0.75	0.85
	BAB	**0.87**	**0.83**	0.67	0.67	**0.77**	**0.69**	0.60	**0.59**	**0.62**	**0.68**	**0.75**	**0.48**	0.09	-	-0.10	0.65	0.30	0.16	0.69	0.65	0.75	0.73	0.83
	SFS	**0.77**	**0.76**	**0.64**	**0.63**	**0.68**	**0.62**	**0.51**	**0.57**	**0.59**	**0.60**	**0.67**	**0.50**	**0.08**	-0.10	-	0.53	0.28	0.06	**0.51**	0.55	**0.64**	**0.57**	**0.70**
	FL1	**0.93**	**0.90**	0.89	0.85	**0.88**	**0.81**	**0.85**	**0.74**	**0.74**	**0.81**	**0.87**	**0.66**	0.69	0.65	**0.53**	-	0.57	0.60	0.82	0.88	0.85	0.88	0.92
	FL2	**0.85**	**0.83**	0.71	**0.71**	**0.77**	**0.71**	**0.62**	**0.64**	**0.66**	**0.71**	**0.77**	**0.56**	0.37	0.30	**0.28**	**0.57**	-	0.12	0.57	0.66	0.74	0.70	0.81
	FL3	**0.83**	**0.78**	0.68	0.64	**0.72**	**0.63**	**0.66**	**0.51**	**0.54**	**0.65**	**0.72**	**0.39**	0.21	0.16	**0.06**	0.60	**0.12**	-	0.59	0.65	0.68	0.68	0.80
	GAR	**0.92**	**0.89**	**0.90**	**0.87**	**0.88**	**0.82**	**0.85**	**0.76**	**0.75**	**0.81**	**0.87**	**0.71**	**0.72**	**0.69**	**0.51**	**0.82**	**0.57**	**0.59**	-	0.86	**0.84**	**0.86**	**0.90**
	ARA	**0.93**	**0.90**	0.87	0.84	**0.87**	**0.80**	**0.81**	**0.74**	**0.74**	**0.80**	**0.86**	**0.67**	0.69	0.65	**0.55**	0.88	**0.66**	0.65	**0.86**	-	0.74	0.76	0.85
	MAM	**0.92**	**0.90**	**0.89**	**0.88**	**0.88**	**0.83**	**0.82**	**0.81**	**0.80**	**0.83**	**0.88**	**0.77**	**0.77**	**0.75**	**0.64**	**0.85**	**0.74**	**0.68**	**0.84**	**0.74**	-	**0.44**	**0.62**
	TF	**0.92**	**0.89**	**0.89**	**0.87**	**0.87**	**0.81**	**0.83**	**0.76**	**0.75**	**0.81**	**0.86**	**0.71**	**0.75**	**0.73**	**0.57**	**0.88**	**0.70**	**0.68**	**0.86**	**0.76**	**0.44**	-	**0.57**
	MAQ	**0.95**	**0.93**	**0.93**	**0.92**	**0.92**	**0.87**	**0.89**	**0.84**	**0.83**	**0.87**	**0.91**	**0.81**	**0.85**	**0.83**	**0.70**	**0.92**	**0.81**	**0.80**	**0.90**	**0.85**	**0.62**	**0.57**	-

Table C.4. Point estimate of demographic parameters estimated across 40 runs of FASTSIMCOAL2 from divergence models for each geographically proximate pair of paleodrainages, with the 95% confidence interval for T_{DIV} and N_2 shown in parentheses. Note that the population size of the first population listed in each paleodrainage pair (N_1) was not estimated under a divergence model, but instead was calculated directly from the empirical data (i.e., it is a fixed parameter in the model) to improve the accuracy of the other parameters estimated from the SFS (following the recommendations for the program; see Excoffier and Foll, 2011). Also shown are number of loci used to calculate the SFSs and the paleodrainage pair in bold highlights the older divergence between northern and southern regions.

Paleodrainage pair	**Loci**	N_1 **(fixed)**	T_{DIV}	N_{ANC}	N_2
1 - 2	4077	1 = 15625	22804 (22741-25107)	76807	2 = 21494 (20014-22657)
2 - 3	5280	2 = 22321	27440 (27333-29874)	81899	3 = 104191 (94039-112579)
3 - 4	9209	3 = 36830	23573 (23645-25861)	209577	4 = 67786 (63811-72052)
4 -5	14038	4 = 34598	12169 (12166-13264)	69108	5 = 36736 (34557-38767)
5 - 6	13512	5 = 43527	25932 (26129-27921)	106179	6 = 57688 (54163-60800)
6 - 7	7893	6 = 47991	24391 (24231-27443)	122119	7 = 16887 (16123-18516)
7 - 8	5988	7 = 29018	44871 (44901-48553)	209764	8 = 286465 (238814-319669)
8 - 9	**11431**	**8 = 78125**	**80560 (51225-88149)**	**102297**	**9 = 84645 (80039-89770)**
9 - 10	16049	9 = 63244	17989 (17897-1915)	55460	10 = 26707 (25442-28488)
10 - 11	14407	10 = 31250	23539 (24104-25920)	73143	11 = 14692 (14334-15552)

Table C.5. Physical properties and population genetic summary statistics for each paleodrainage that were used in linear regression analyses to test for an association between the physical properties of paleodrainages and patterns of genetic variation of constituent populations. Paleodrainages are listed from the most northern to the southern most paleodrainage (see Figure IV.1). Hydrological maps were used to calculate the currently exposed area and the entire area of a paleodrainage (i.e., the exposed and submerged area), as well as the number of rivers of the present area of each paleodrainage (see methods for details). Population genetic summary statistics were calculated as the average over all sampled populations from a single paleodrainage, and include estimates of nucleotide diversity, π, and observed heterozygosity, H_{EXP}.

Paleo-drainage	**Current exposed area (Km^2)**	**Entire area (Km^2)**	**# rivers**	$\boldsymbol{\pi}$	$\boldsymbol{H_{EXP}}$
1	4901	8385	3	0.013	0.012
2	97	2754	1	0.021	0.02
3	898	4837	1	0.037 (±0)	0.03 (±0.002)
4	1084	3728	2	0.032 (±0.01)	0.03 (±0.01)
5	618	4962	1	0.044 (±0.018)	0.04 (±0.017)
6	1381	9072	2	0.046 (±0.01)	0.044 (±0.009)
7	1514	5599	2	0.028	0.026
8	25259	31546	3	0.084	0.075
9	23507	30924	5	0.069 (±0.002)	0.056 (±0.008)
10	5673	12743	4	0.033 (±0.024)	0.028 (±0.02)
11	14372	23837	8	0.014 (0.005)	0.013 (±0.005)

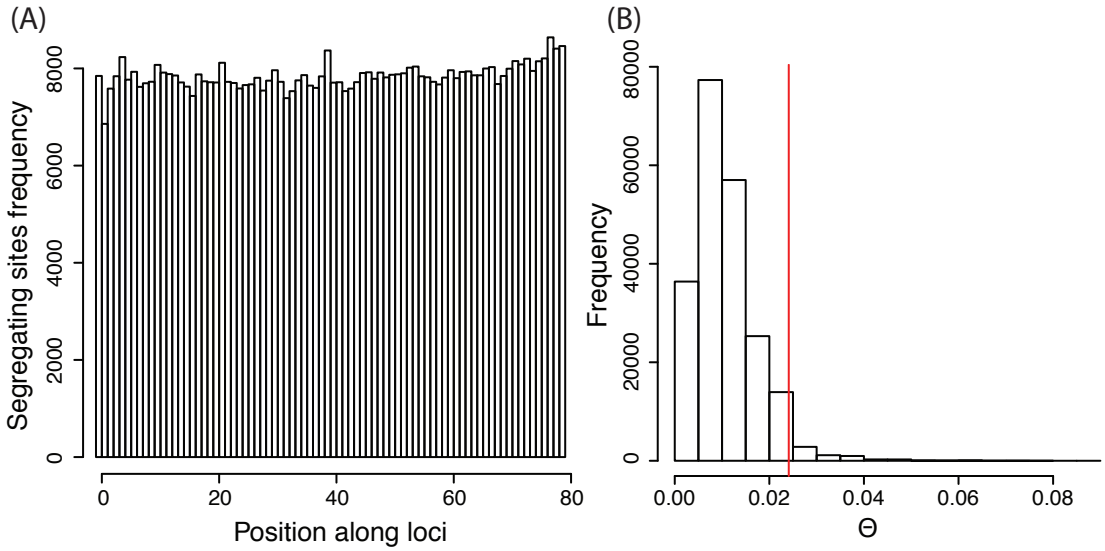


Figure C.1. Summary of the frequency of segregating sites for each base-pair position of a locus (A), and the distribution of theta, θ, per loci (B), with the red line marking the θ-values in the 95 percentile that were excluded from analyses to avoid including variation likely reflective of sequencing and/or assembly errors.

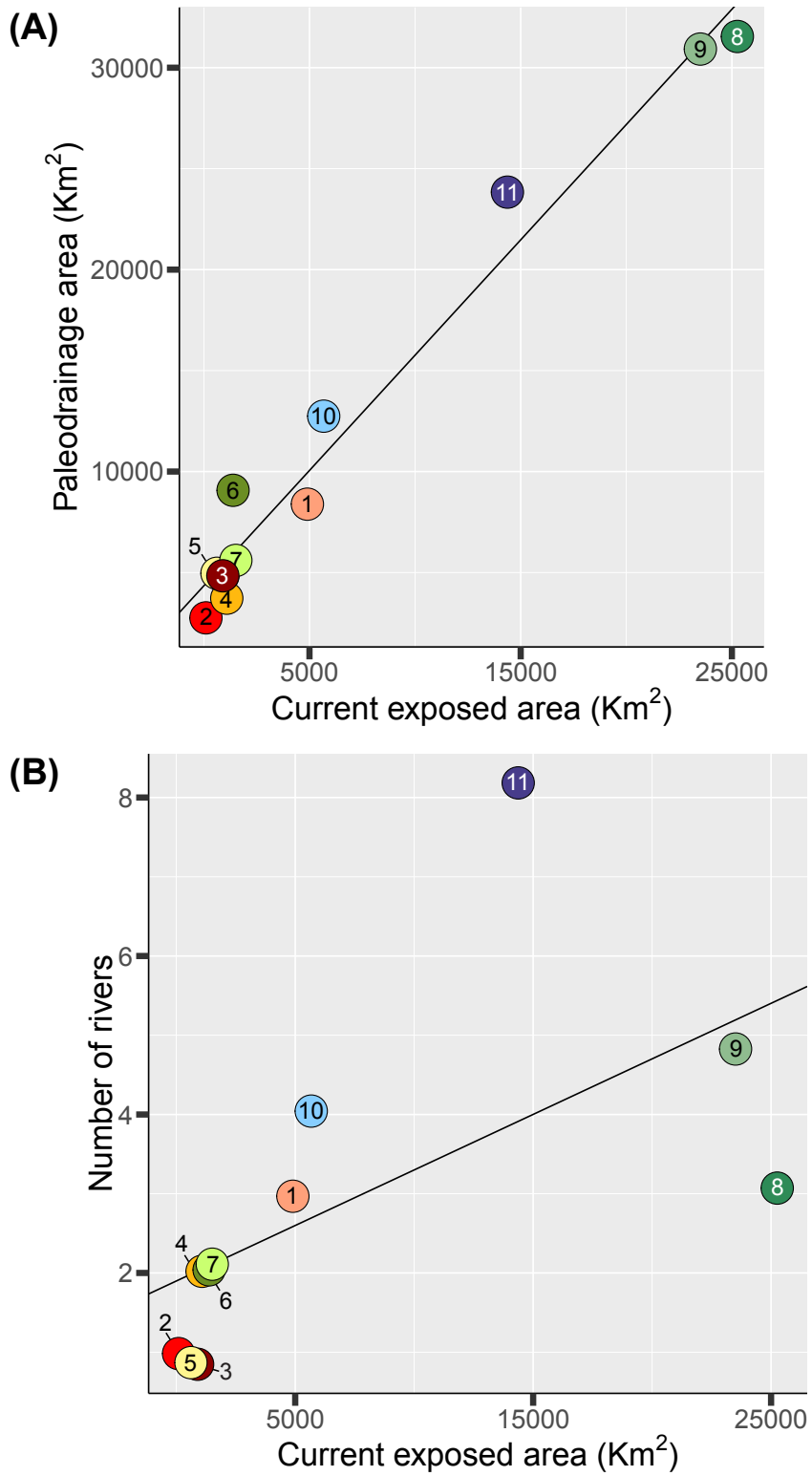


Figure C.2. Correlation between the total area and the currently exposed area of each paleodrainage (A; R^2 = 0.97; p-value <0.001) and between the currently exposed area and the number of rivers in each paleodrainage (B; R^2 = 0.39, p-value = 0.04). The colored dots correspond to the paleodrainages presented in Figure 1.

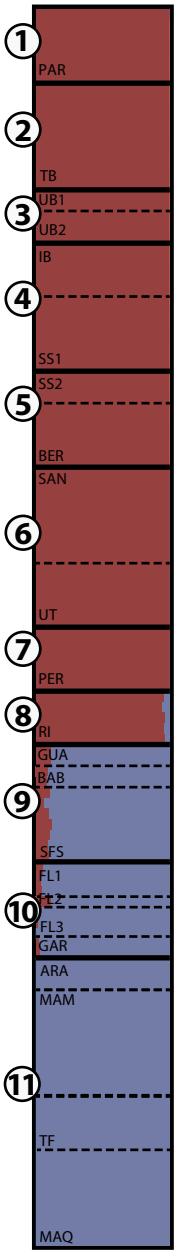


Figure C.3. The two most probable genetic clusters, $K = 2$, based on analyses with the full dataset, which corresponds to a northern group (shown in dark red) and a southern group (shown in purple). Thick black lines and numbers in circles demarcate paleodrainages and dashed lines the populations within a paleodrainage, whose names are listed on the left, arranged from north (PAR) to south (MAQ) (see Figure IV.1). The posterior probabilities of the ancestry of each individual are shown (i.e., the relative proportion of dark red and purple).

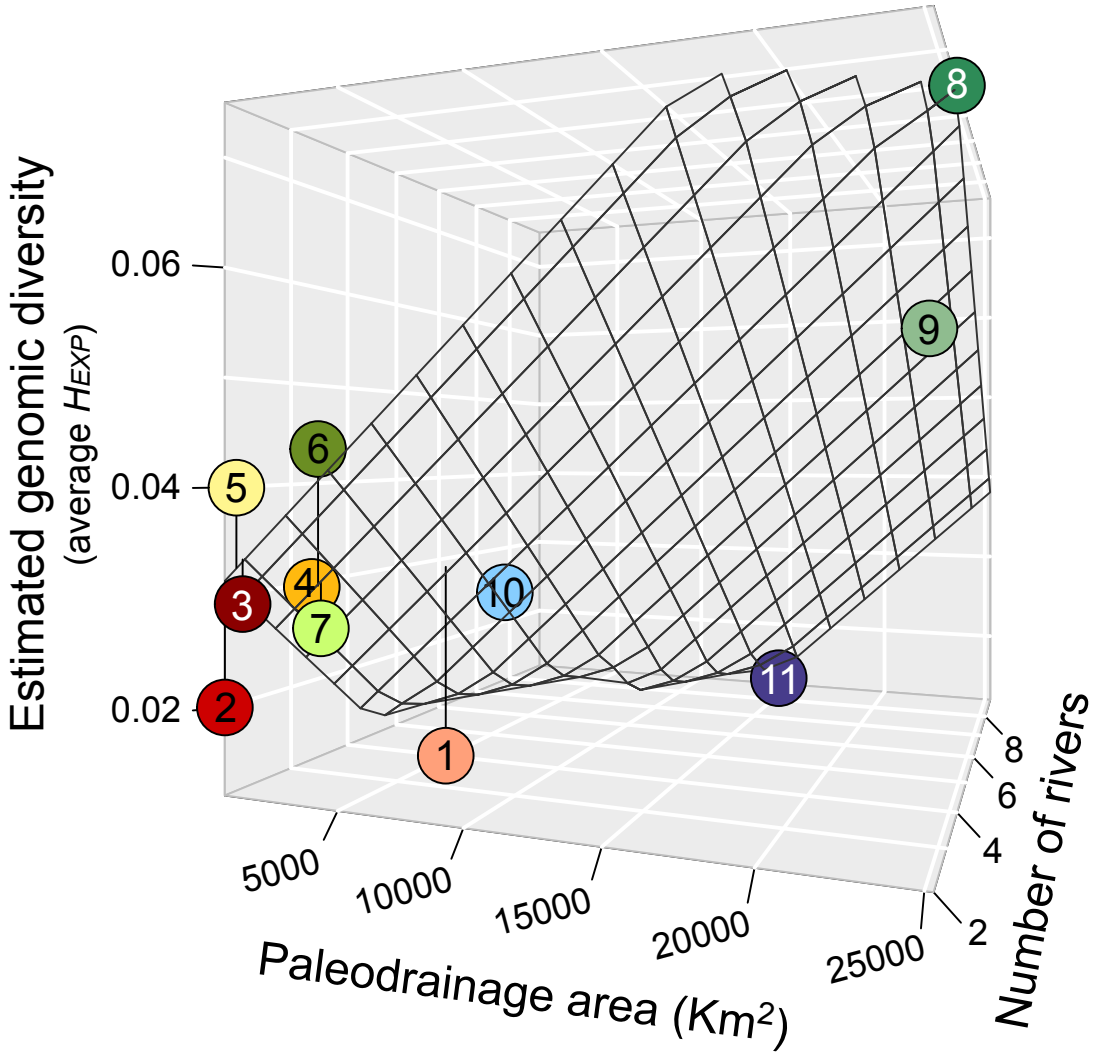


Figure C.4. General linear model fit between paleodrainage properties (i.e., area and number of rivers) and nucleotide diversity (H_{EXP}; R^2 = 0.79, p-value <0.01). The colored dots correspond to the individual paleodrainages on the map in Figure 1.

CHAPTER V

Spatial and temporal congruence of regional genomic structure across a Brazilian coastal fish community

Andréa T. Thomaz and L. Lacey Knowles

Abstract

Biological, geological and climatic processes may act in concert, determining how barriers structure biodiversity. Genetic data can identify possible geographic barriers, as well as regional divergence patterns, to test the extent to which different taxa are affected similarly. Such comparative analyses can highlight the possible interaction between historical events and species-specific responses through the detection of discordant genetic structure in taxa. For dynamic environments, such as coastal areas, identifying regional barriers is challenging as they were dramatically affected by geological and climatic events over time. Here, we use SNPs to test for concordant genetic structure among four co-distributed species of tetras (Teleostei: Characidae) along the Brazilian Atlantic coast. Based on estimated population relationships and hierarchical analyses of genetic structure we identify a geographic barrier that is common to all four taxa, as well as geographic regions for which corresponding genetic differentiation is apparent in only a subset of taxa. In addition to spatial concordance, we test the extent to which the effects of the barriers have a common origin – that is, we test for temporal concordance of divergence. Model-based estimates of divergence times indicate that all divergences date to the Pleistocene; however, simultaneous divergence times across taxa for any given barrier across species is not supported. We discuss how our findings, and in particular the different degrees of spatial and temporal concordance among species, relate to the turnover driven by sea-level fluctuations during the Pleistocene in coastal areas.

Introduction

Spatial congruence in the distribution of species or in the structure of populations between co-occurring taxa has long been recognized as a signal of shared evolutionary history (Bermingham and Avise, 1986; Edwards and Beerli, 2000; Donoghue and Moore, 2003). Such congruence has helped to identify geographical areas that act as barriers to the biological community, especially in cases where a physical barrier is not readily evident, such as ephemeral, climatic and ecological barriers (Carnaval *et al.,* 2009; Edwards *et al.*, 2012; Papadopoulou and Knowles, 2016) or even when genetic discontinuities do not result from a complete barrier but are a function of dispersal and demographic traits (Irwin, 2002). The identification of these barriers also provides insights into diversification and speciation processes (e.g., allopatry; Coyne 1994; Barraclough and Vogler 2000), especially in taxa with constrained dispersal routes, such as freshwater fishes. For example, with the exception of classical cases of speciation in lakes, diversification and speciation between isolated drainages is the main process invoked to explain the majority of freshwater fish species, especially in riverine environments (Seehausen and Wagner, 2014).

The role of barriers as dispersal constraints is not limited to just a single taxon, but also is thought to apply to entire communities of freshwater fishes (Burridge *et al.*, 2006; Albert and Carvalho, 2011; Chakona *et al.,* 2013). However, for coastal regions that are constantly affected by changes in their environment, their impact on the survival, dispersal rates, and population dynamics of constituent taxa raises the question of the extent to which these communities would show congruent divergence patterns (Dias *et al*., 2014). For example, as a consequence of these repeated population reconnection cycles, coastal areas may be subject to a high spatial and/or temporal species turnover. Furthermore, identifying congruent geographic barriers in coastal areas can give insights into the general patterns of spatial connectivity across a geographic region. However, spatial congruence in patterns of genetic divergence might not be accompanied by corresponding temporal congruence in divergence times as a consequence of cyclical processes, species ecology and demographic factors. Considering the relative timing of divergence associated with geographic barriers in a comparative framework can clarify if a congruent spatial barrier between species is the result of a single event or not (i.e., pseudocongruency; Cunningham and Collins, 1994; Donoghue and Moore, 2003).

Here we apply a coalescent-based framework using SNP data to test for spatially and temporally congruence genetic structure in co-distributed fish from the eastern Atlantic coast of Brazil. This region has been directly impacted by numerous geological and climatic events. For example, studies of terrestrial organisms have identified how geographic barriers along the coast that predate the Last Glacial Maximum (LGM) have structured communities of the Brazilian Atlantic Forest, as species were displaced into refuges during the Pleistocene (Carnaval *et al.*, 2009; Thomé *et al.*, 2014). Similarly, for freshwater fishes, two lines of evidence suggest diversification processes may be tied to the geographic barriers associated with geological and climatic changes. This includes a pre-Pleistocene diversification of the ichthyofauna with a colonization of Brazilian coastal basins based on phylogenetic analyses that show relationships between inland and coastal basins lineages indicative of river captures (Ribeiro 2006; Roxo *et al.*, 2012). Following colonization of the coastal basins, sea level fluctuations (which occurred throughout the Pleistocene; Miller *et al.*, 2011) facilitated dispersal along the coast by uniting currently isolated coastal drainages during periods of low sea levels, which then subsequently became isolated as rising sea levels reinstated the barriers among drainages. These connections were ephemeral, but nonetheless they have left a detectable signature in patterns of genetic structure (Weiztman *et al.*, 1988; Thomaz *et al.*, 2015). However, the extent to which fish communities respond similarly to these cyclical changes has not been tested. As noted, the repeated and frequent shifts in sea level may give rise to high species and/or population turnover, such that even if geographic barriers were shared among taxa, there may not be a general temporal concordance in divergence times. Likewise, because fish communities exhibit species-specific ecologies associated with different habitats (Waters and Burridge, 2016) or inherent dispersal tendencies (Radinger and Wolter, 2014), the degree of spatial and temporal congruence may vary among drainages in these dynamic coastal regions, especially if differences in bathymetric data imply different periods of connectivity (see Papadopoulou and Knowles 2015, 2016).

To examine if lineages diversification and genetic divergence is spatially and temporally congruent among freshwater taxa inhabiting the coastal basins, we studied four characid taxa (Ostariophysi: Characiformes), commonly known as tetras. Specifically, we selected four co-distributed lineages that are endemic to the Brazilian coastal basins (rarely distributed in inland basins; Figure V.1): *Mimagoniates microlepis, Hyphessobrycon boulengeri* (with a putative new

species; Carvalho 2006), *Hollandichthys* (encompassing two valid species: *H. multifasciatus* and the recently described *H. taramandahy*; Bertaco and Malabarba, 2013) and the coastal *Bryconamericus* group (and hereafter referred as *Bryconamericus*; encompassing several closely related species of this genus with a coastal distribution, such as: *B. microcephalus, B. ornaticeps, B. tenuis* and *B. lethostigmus*; Pezzi da Silva, 1998; Hirschman *et al.*, 2017). Based on the matches and mismatches observed between spatial and temporal breaks across species we are able to identify the effects of historical events on the freshwater fish community in the study area, and we discuss the relative contribution of different processes shaping genetic diversity in coastal communities in general.

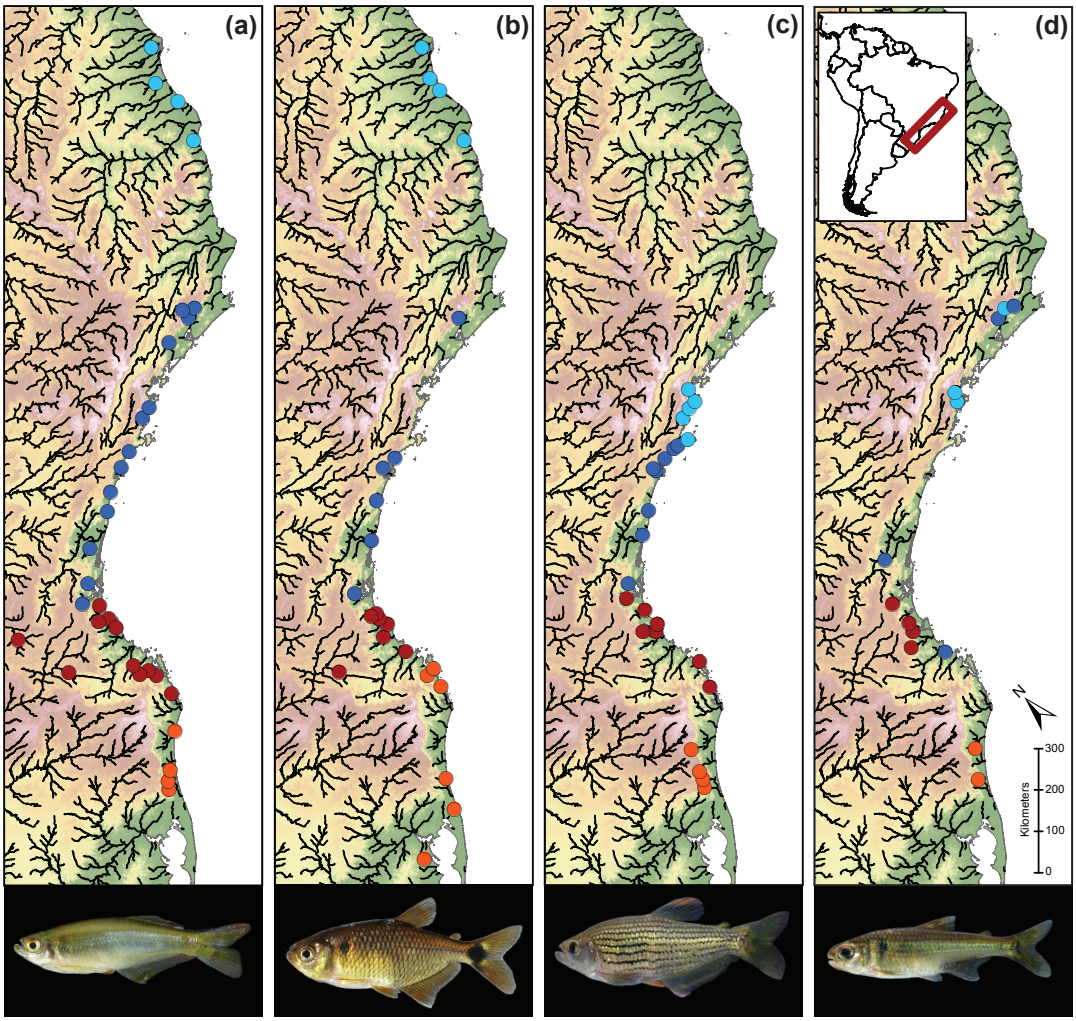


Figure V.1. Distributional map and a specimen photo of (A) *M. microlepis* (38 mm Standard Lengh - SL), (B) *H. boulengeri* (47.8 mm SL), (C) *Hollandichthys* (*H. multifasciatus;* 99.5 mm SL), and (D) *Bryconamericus* (*B. microcephalus;* 57 mm SL) with sampled populations for genomic analyses labeled as colored dots; see small inset of South America for area of study. Different colors depict clusters of genetic differentiation obtained with hierarchical analyses among populations of each species (see Figure V.2 and results for details).

Material and Methods

Sampling and genomic library preparation and processing

Specimens for each of the four species were collected across their entire distributions and tissue samples were extracted from body muscle and preserved in ethanol; vouchers and tissues for this study were catalogued in the ichthyology collection at the Universidade Federal do Rio Grande do Sul (UFRGS), Brazil. Additional tissues (approximately 10% of the samples) were obtained from the Museu the Ciências e Tecnologia, Pontifícia Universidade Católica do Rio Grande do Sul (MCP) and Museu the História Natural Capão da Imbuia (MHNCI) (see complete list in Supporting Table V.1). Genomic DNA was extracted using Qiagen DNeasy kits.

Eight double digest Restriction Associated DNA (ddRAD) libraries were constructed: three libraries for *Mimagoniates microlepis* (118 individuals for this study, out of the 240 total individuals that sequenced across these libraries, which include samples not targeted in this study), two libraries for *Hyphessobrycon boulengeri* (136 individuals), two libraries for *Hollandichthys* (182 individuals), and one library for *Bryconamericus* (87 individuals). We followed the protocol of Peterson *et al.* (2012), except for the two *Hollandichthys* libraries (see Thomaz *et al.,* 2017 for preparation details that followed Parchman *et al.* 2012). Between 300 and 400ng of each DNA sample was double digested with two restriction enzymes (*EcoRI* and *MseI*), followed by a ligation step to add unique barcodes. Samples for each library were pooled and fragments between 350-450bp were selected using a PippinPrep machine. A PCR with 10 cycles was used to add Illumina flowcell adapters. All steps described above were followed by a clean-up step using AMPure beads (1.6x ratio; except after Pippin Prep) to remove small DNA fragments such as primers, and by a high sensitivity Qubit quantification assay. Each library was sequenced in one lane of an Illumina HiSeq2500 to generate single-end 150bp reads (100bp for *Hollandichthys*) at The Centre for Applied Genomics, Toronto, Canada.

Genomic data were demultiplexed and processed separately for each species with the STACKS version 1.41 pipeline (Catchen *et al.*, 2013). For quality filtering, reads with more than one mismatch in the adapter sequence or a barcode distance greater than two (as specified in *process radtags)* were removed, and individuals with less than 300K reads were excluded. To create stacks within each sample, USTACKS was run with a minimum depth of coverage of five and an error bound of $\varepsilon = 0.1$, followed by CSTACKS with a maximum of two mismatches between sequences within a given stack in order to build a catalog of all loci. The stacks of

individual samples were matched against the catalog using SSTACKS with default options. To obtain a vcf output file containing all variable sites from STACKS, we ran the POPULATIONS module with "loose" parameters (i.e., -r 0 -p 2 -m 5 --min_maf 0 --max_obs_het 0.5). We processed this output file in R version 3.3.1 (R Core Team, 2016) to create a whitelist that excluded highly variable positions at the 3' end of all loci and loci with θ-values above the 95th percentile of this distribution, to avoid errors associated with sequencing and assembly (see Figure D.1 and D.2). Then, based on this whitelist, we re-ran the POPULATIONS module. All bioinformatics processing with STACKS was performed in the Advanced Research Computing Technology Services at the University of Michigan.

Because of the various requirements of different analyses used to characterize the geographic structuring of genomic variation, such as sensitivity to missing data (Huang and Knowles, 2016) and computational feasibility, three datasets were generated varying the amount of missing data and the numbers of individuals. One dataset was comprised of unlinked SNPs with maximum of 50% missing data, and hereafter referred to as the SNP dataset (i.e., one random SNP per locus; see Supporting Table D.2 for details), which it was used for estimates of population trees. The other dataset included linked or unlinked SNPs with maximum 25% missing data (except for *M. microlepis*, for which we allowed 35% missing), and hereafter referred to simply as the genomic dataset, and was used in most of the analyses including summary statistics and STRUCTURE analysis. The last dataset, hereafter referred as the reduced dataset, was used to run FASTSIMCOAL2 analysis for each identified geographic break. To generate this dataset, we selected the 20 individuals per species and per identified geographic break (see below) with the smallest amount of missing data from each side of the geographic break (40 individuals in total), using only SNPs with less than 10% missing data (see details below). For all these datasets, individuals with considerably fewer SNPs in comparison to other individuals of the same population were also excluded. All filtering steps were performed using the toolset PLINK v.1.90 (Purcell *et al*., 2007) in association with R scripts (available in GitHub: https://github.com/deathomaz).

Characterizations of genetic diversity and population structure

A population is defined here as all the samples from the same river basin/drainage or island (e.g., FLO and IB populations). Population genetic summary statistics were calculated

from the genomic dataset for each population or between each pair of populations per taxa. This included calculations of π and H_{OBS} using the POPULATION module from STACKS based on a whitelist created after the filtering with PLINK. F_{ST}-values and their statistical significance were calculated in Arlequin 3.5.2.2 (Excoffier and Lisher, 2010) with 10,000 replicates and a Bonferroni correction for multiple comparisons.

To examine how evolutionary relationships correspond to spatial configuration of drainages along the Brazilian coast, we estimated a population tree (Knowles and Cartens, 2007) accounting for the coalescent variation associated with random sorting of gene lineages among loci, and incomplete lineage sorting for any given locus using the program SVDquartets (Chifman and Kubatko, 2014), as implemented in PAUP* 4.0 (Swofford, 2003) under the multispecies coalescent model with all possible quartets evaluated. Branch support was assessed with 1,000 bootstrap replicates and midpoint rooting was used given the absence of outgroups in our datasets.

Hierarchical STRUCTURE analyses (Pritchard *et al*., 2000) were used to evaluate if the probabilistic assignment of individuals in each species to clusters is species specific or if there is a general pattern between species along the Brazilian coast. To access substructure, we performed analyses with the full distribution of each species and then, followed by sequential analyses for each of the data subsets identified as distinct genetic clusters (except for *H. boulengeri* – see results below). The genomic datasets with unlinked SNPs were used, and individuals were not conditioned on any population membership (i.e., no populations were used as priors). Each dataset was analyzed with *K*-values ranging from 1 to 5 or 10 (see Table V.1 for specific information for each species). We performed ten independent runs under the "Admixture" and "Allele Frequencies Correlated" models for 500,000 MCMC iterations following a burn-in period of 200,000 iterations for each analysis. The ΔK of Evanno *et al*. (2005) implemented in STRUCTURE HARVESTER (Earl *et al*., 2012) was used to identify for each species the most probable number of genetic clusters for each species, with the graphical probabilistic assignment of individuals to clusters performed using the CLUMPAK pipeline (Kopelman *et al*., 2015).

Divergence time estimates

Focusing on the major spatial breaks identified among the different species based on the phylogenetic tree and STRUCTURE analyses, we estimated divergence times of these breaks in each species separately using the composite-likelihood method FASTSIMCOAL2 (Excoffier and Foll, 2011; Excoffier *et al*., 2013) based on the folded joint Site Frequency Spectrum (SFS; i.e., for the minor allele since we do not have information from outgroups to obtain the derived state). In *Bryconamericus* the divergence time for one geographic break was also estimated after removing some individuals that showed evidence of gene flow based on the STRUCTURE analyses to check for potential biases in the divergence estimated (see results below). To avoid over-parameterization, and because in all other cases there was no obvious evidence of gene flow, we chose not to use a more complex model with additional migration parameters.

To maximize the number of loci with no missing data per locus, a subset of 15 individuals per locus was sampled from the reduced dataset using a custom script since SFS calculations require no missing data per locus; the script is available on GitHub: https://github.com/deathomaz and is modified from Papadopoulou and Knowles (2015). To improve the performance of parameter estimates from the SFS (following recommendations of the program; see Excoffier and Foll, 2011), we calculated an effective population size of one of the two populations (specifically, N_1) directly from empirical data (i.e., specifically, from the nucleotide diversity (π) of fixed and variable sites). The remaining parameters (i.e., N_2, ancestral population size N_{ANC} and divergence time T_{DIV}) were estimated based on the SFS, with a mutation rate, μ, estimated from the size of a genome (see formula in Lynch, 2010), where a close relative to each species was used (see Table D.6 for details). A generation time of one year was used for all species, which is the common generation time in characids (Azevedo, 2010). FASTSIMCOAL2 runs were performed with 40 replicates for each group pair with 100,000 to 250,000 simulations per likelihood estimation based upon a stopping criterion of 0.001, and 10 to 40 expectation-conditional cycles (ECM). We performed 100 parametric bootstrap iterations by simulating SFS with the same characteristics from the best maximum likelihood estimate and re-estimating the parameters with 40 runs for each one of the 100 simulated SFS, reporting here the 95% confidence interval.

Results

Genetic diversity and population structure

We obtained a total of 165 million to 325 million reads per species. After applying filters for missing data, genotyping rates ranged from 0.67 to 0.72 for the SNP dataset and from 0.85 to 0.92 for the genomic dataset across species (see Table D.1 and D.2 for information per species and individuals). A total of 47 drainages (populations) were sampled across the four species, with an average of 23 drainages sampled per species (Table D.3). Species showed similar levels of genetic diversity (see Table D.4), with *Bryconamericus* showing the highest genetic diversities compared with somewhat lower diversities within *M. microlepis* and *H. boulengeri*. Genetic differentiation as measured by F_{ST} (Table D.5) tended to be high among drainages in all four species, ranging on average from 0.65 in *M. microlepis* to 0.7 in *H. boulengeri*.

There is a strong correspondence between geography and genetic differentiation in all four species. Specifically, a latitudinal pattern of relatedness is evident from the phylogenetic analyses (Figure V.2), except for a few geographically distant drainages in *Bryconamericus* that were closely related genetically.

Analyses of the full dataset in STRUCTURE identified $K = 2$ as the most probable value of K based on ΔK (Evanno *et al.*, 2005) in three species (*M. microlepis*, *Hollandichthys* and *Bryconamericus*), and a $K = 4$ for *H. boulengeri* (Table V.1). These results, as with estimated phylogenetic trees, identified a geographic break in the center of the species distributions at the Paranaguá estuary (hereafter referred to as the central break). This central break is apparent in all four species, separating a northern and southern region in each species. This barrier seems to be permeable with some gene flow occurring between geographically distant populations of *Bryconamericus*, but not in the other species.

Subsequent STRUCTURE analyses performed in the northern and southern subgroups (except for *H. boulengeri*) to account for the hierarchical structure within populations identified $K=2$ as the most probable value of K for all subgroups in each species. In the northern portion of the distribution, two species have smaller distributional ranges (*Hollandichthys* and *Bryconamericus*). These results suggest that the geographic breaks for these species are unique (i.e., not shared among other species). For the two broadly distributed species *M. microlepis* and *H. boulengeri*, a common geographic break is observed above the mouth of the Paraíba do Sul River (hereafter referred to as the north break), suggesting isolation of the very northern

populations that is further supported by the absence of the other two taxa in this area. In the south sub-break (hereafter referred to as the south break), a common structure is identified between three out of the four species. While for *H. boulengeri* there is a major geographic break between the island population of Florianópolis (population 9) and the inland Itajaí (population 12) basin, for all other species there is a sub-break between Araranguá (population 6) and D'Una (population 7) river basins - but for *Bryconamericus* it could be related to a sampling gap (Figure V.1). All these clusters are in agreement with the clades proposed in the phylogenetic trees. For *H. boulengeri*, we did not perform hierarchical STRUCTURE analysis because $K = 4$ was identified as the most probable value of K based on ΔK (Evanno *et al.*, 2005).

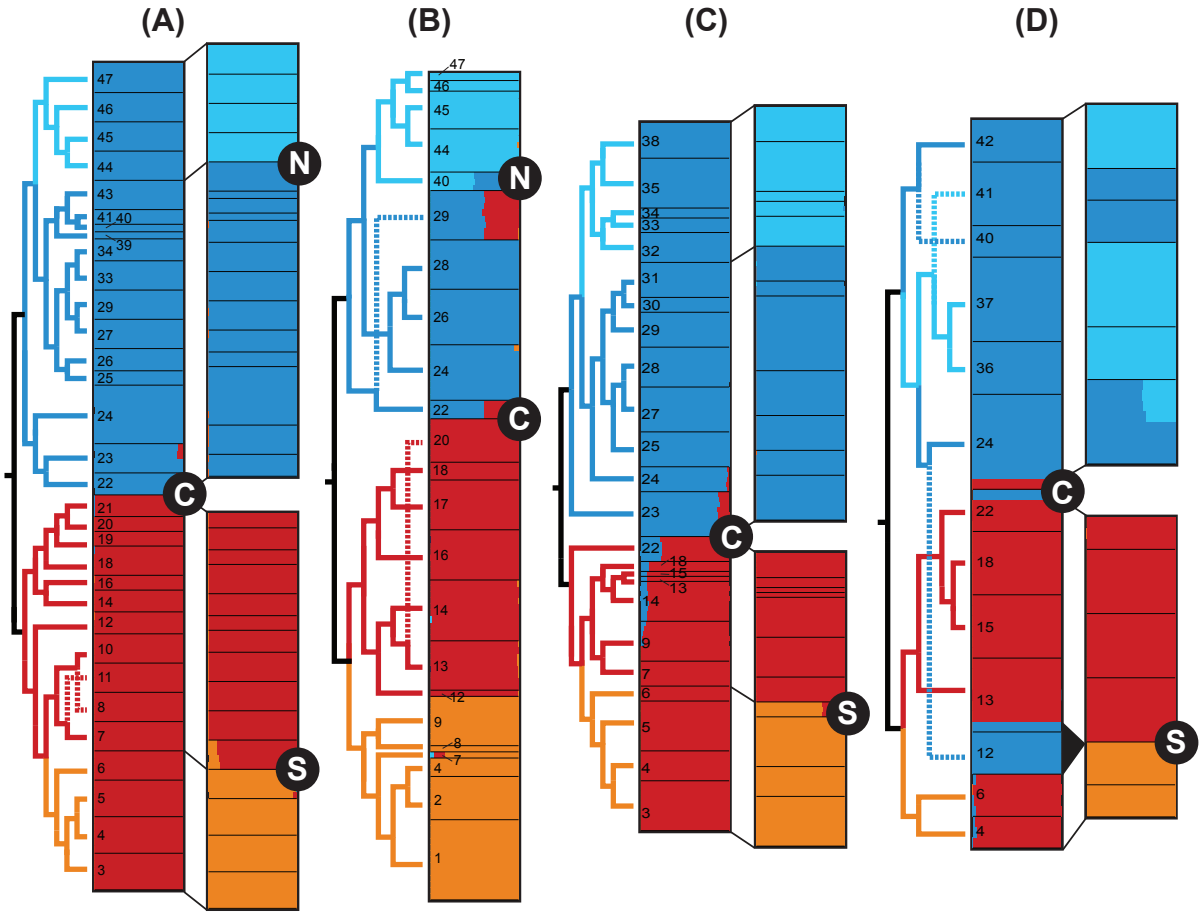


Figure V.2. Estimates of population relationships and genetic clusters in (A) *M. microlepis*, (B) *H. boulengeri*, (C) *Hollandichthys*, and (D) *Bryconamericus*, from SVDquartets and STRUCTURE analyses. Congruent patterns of divergence are emphasized by black circles with the letter corresponding to the geographic break (N = North, C = Central Paranaguá, S = South), which are also highlighted on the distributional maps (see colored dots in Figure V.1). Note the blue group in South *Bryconamericus* cluster was removed from the hierarchical analysis.

Table V.1. Results of hierarchical STRUCTURE analyses, with the full dataset (All) and the population subsets (North and South) for each species. For each analysis (i.e., row), the first and second most probable *K*-values identified using Evanno method are reported along with the correspondent *ΔK*. The total number of loci and individuals analyzed are given, as well as the total individual genotyping rate (Gen. rate).

Taxa	Level	Loci	Inds.	Gen. rate	K tested	1st K	ΔK	2nd K	ΔK
M. microlepis	All	1,800	113	0.79	10	2	9,054.0	4	20.0
	North	1,042	59	0.87	5	2	5,780.7	4	1,155.1
	South	1,441	54	0.88	5	2	2,110.3	3	2,078.7
H. boulengeri	All	6,129	134	0.86	10	4	34.3	2	3.0
Hollandichthys	All	6,902	142	0.87	5	2	19,511.8	3	5.9
	South	6,335	59	0.87	5	2	12,095.1	3	885.9
Bryconamericus	All	4,276	69	0.95	10	2	10,261.1	3	144.3
	South	4,180	28	0.95	5	2	3,331.4	3	998.2

Divergence time estimates

In order to determine if the same process was responsible for the spatial agreement between species for each one of the barriers identified above, divergence time estimates around each break were estimated for each species. Most point estimates among species and breaks were positioned in the Upper Pleistocene (<126 kya; Figure V.3 and Table D.6). However, there is high diversity between species divergence time estimations across each break. In general, it is possible to infer that each spatial break had processes that shaped the genetic structure of species in at least two different temporal events (Figure V.3) based on the distribution of the divergence time point estimates and the confidence intervals. In the central break of Paranaguá, *Hollandichthys* and *Bryconamericus* diverged at older times (>60 kya) while *M. microlepis* and *H. boulengeri* have much recent point estimates around the LGM (20-30 kya). We want to highlight that two divergence times were calculated for *Bryconamericus* given the possibility of gene flow demonstrated in the phylogeny and in the STRUCTURE analysis (e.g., population 12 from the south is genetically related to the northern clade; Figure V.2). We ran one analysis with all populations and another (here called "No admixture") removing individuals with a signal of admixture. Results from these two runs showed a difference in point estimates of 15 kya; however, this difference did not change qualitative interpretation about spatial congruence (see Figure V.3 and Table D.6). That is, despite the difference in divergence times including and excluding individuals that might be indicative of admixture, the estimated time is still in the distant past, indicating that there was no recent admixture event between northern and southern groups in this taxon.

The northern break corresponds to two different divergence times for the two species whose distributions extend across the northern region. Specifically, an older divergence is observed in *H. boulengeri* (~135 kya), which is the oldest estimation among all comparisons, while for *M. microlepis* a more recent divergence was estimated (~25 kya), corresponding to the LGM. The south break presents the most recent divergence time estimates among all comparisons, which are ~7 and 10 kya for *M. microlepis* and *Hollandichthys*, respectively. This geographic break has a deeper divergence in *Bryconamericus* (~75 kya; Figure V.3).

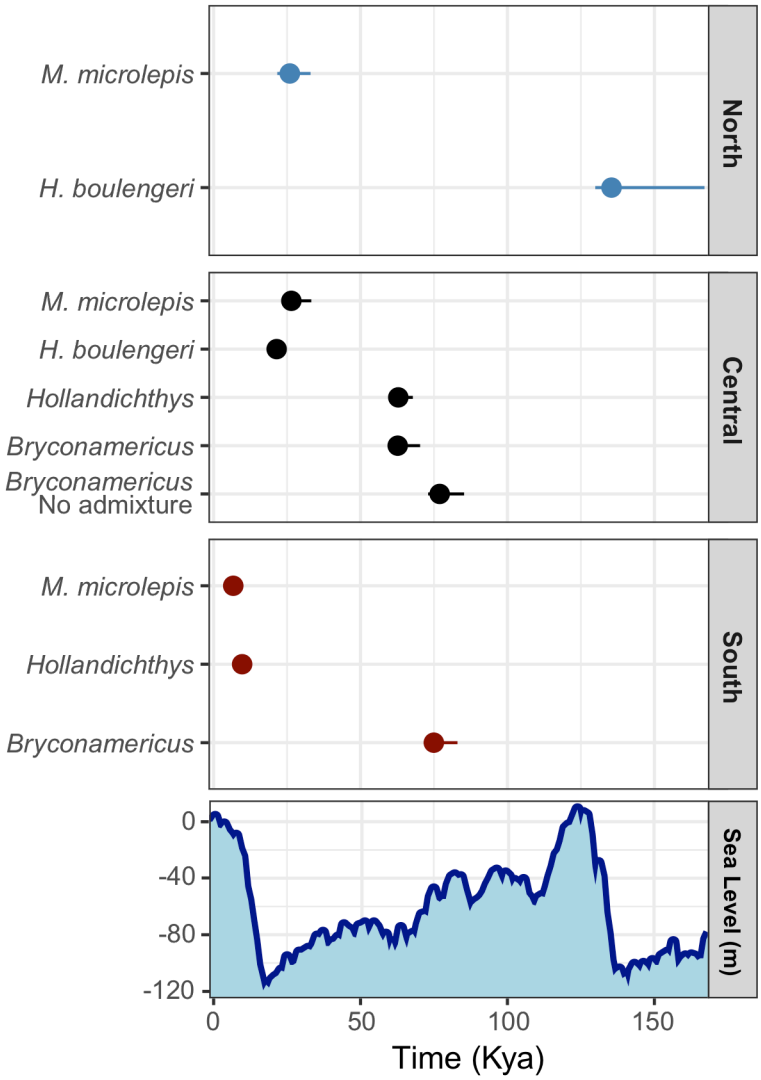


Figure V.3. Divergence times and 95% confidence interval estimated with FASTSIMCOAL2 per species for each geographic break (i.e., North, Central and South) along the Brazilian coast with the estimation of sea level for the same time period (Miller *et al.*, 2011).

Discussion

Congruent genetic divergence patterns demonstrate that the effects of geographic barriers are not taxon-specific, but have a common effect across coastal freshwater fishes of Brazil. This finding indicates that regional patterns caused by Pleistocene climatic changes are not restricted, or unique, to terrestrial communities, even though this has received most of the attention in the area (Carnaval *et al*., 2009; Leite *et al.*, 2016). Nevertheless, despite the spatial congruency

observed among taxa that identify common geographic barriers structuring regional divergence patterns in the freshwater fishes, temporal congruence patterns varied across barriers. Specifically, divergence times dating mostly to the Upper Pleistocene implies there is high population turnover such that the older divergences may have been lost. Here we discuss the impacts of these findings in understanding divergence processes in coastal regions. In particular, we refer to the hypothesized role of paleodrainages themselves, and how our comparative data raises questions about the generalizability of this hypothesis. This includes a more nuanced approach for understand the role of barriers in structuring ichthyofauna communities.

Divergence processes in coastal basins

The ephemerality associated with coastal regions during the Pleistocene has been demonstrated mostly in terrestrial organisms and islands environments (Edwards *et al.,* 2012; Ali and Aitchison, 2014; Papadopoulou and Knowles, 2015; Leite *et al.,* 2016). However, the constraints imposed by riverine systems on their fauna (Thomaz *et al.,* 2016) make freshwater fishes ideal organisms to evaluate the effect that dynamic environments, in this case coastal regions, have on genetic patterns.

The strong spatial congruency at the regional level is observed in the entire community despite differential dispersal capabilities that each lineage present based on differences in their ecology. For example, among the set of taxa sampled here, *M. microlepis* is the most generalist species, being widely distributed in small to large streams and frequently associated with vegetation (Menezes and Weitzman, 2009), *H. boulengeri* lives in lowland areas (e.g., slow flowing streams and lakes; Carvalho, 2006), *Hollandichthys* is associated with small streams, slow flowing waters, and dense riparian vegetation (Bertaco and Malabarba, 2013), and *Bryconamericus* inhabits upper sections of fast flowing water of river and streams running over rocky substrate (Hirschmann *et al.*, 2017). The lack of taxon-specific differences in the spatial pattern observed in genetic data reinforces the commonalities associated with riverine species such as strong genetic structure and relatively small population sizes (Tedesco *et al.,* 2012) as a consequence of the constraint imposed by riverine environments.

The strong spatial congruence in divergence patterns across species suggests that abiotic factors supersede any taxon-specific differences in their ecologies that might make some barriers more or less effective. However, caution is warranted given the scale of the barriers studied here.

Specifically, congruence in this case was observed at the regional level. We cannot rule out that connectivity patterns might vary in a species-specific manner at a more local scale (see Papadopoulou *et al.,* 2015; Massatti and Knowles, 2016; Papadopoulou and Knowles, 2016). Moreover, for riverine fishes, there is contrasting evidence regarding the structuring of genetic divergence at smaller scales, such as between paleodrainages, which might be indicative of taxon-specific effects of barriers (Chakona *et al*., 2013; de Bruyn *et al.,* 2013; Tscha *et al.,* 2017; although see Thomaz et al. 2015; Thomaz *et al*., in review).

Moreover, with at least three different divergences times estimated for the geographic breaks, the hypothesis of paleodrainages driving regional divergence through sea-level changes is not clear (Thomaz *et al*., 2015). The older (~135ka) and younger divergences (<25ka) match mostly with periods of sea level uplift, while the 60-70ka divergences, mostly in the central break, occurred in a period of several fluctuations (Figure V.3). Although sea level changes during Pleistocene may have structured some of these groups, we cannot clearly assign all events to a vicariant model that establishes formation of genetic breaks during transgression periods. Based on this evidence it is possible that paleodrainages structure genetic patterns to a subset of taxa (i.e., *Hollandichthys*; Thomaz *et al*., in review), but not for the entire community. This conclusion is general robust to errors in divergence time estimates (e.g., those associated with biases related to mutation rate) given that the timing of divergence differed among barriers within a species, in addition to across species for any specific barrier (Fig. V.3).

There some partial temporal congruence observed here across taxa and barriers. This agrees with previous reports in a handful of comparative studies with a diverse set of animals in island-like environments that have been suggested to reflect differences in temporal co-diversification as a function of environmental stability (e.g., unstable coastal areas or stable inland; Papadopoulou and Knowles, 2016; Shaw and Gillespie 2017). For example, the species that inhabits the lower portion of rivers, *H. boulengeri*, and the habitat generalist, *M. microlepis*, are in environments that may promote higher dispersal opportunities and are also the species with wider distributions (Figure V.1). Both species have relatively young divergence times (except by *H. boulengeri* at the north break), whereas the species inhabiting the upper river portions and/or are habitat specialists (i.e., *Hollandichthys*) are characterized primarily by older divergence times of 60-70 ky. Whether these ecological differences may be linked causally to the postulated high lineage turnover that would result in recent (as opposed to older) divergence time

estimates is intriguing, but will require further work. Specifically, it may be suggestive of differences in temporal stability of populations. Alternatively, it might reflect general difference in connectivity as a function of dispersal propensities. For example, the generalist species *M. microlepis*, as well as taxa in the lower portion of rivers like *H. boulengeri*, may move readily among temporally accessible connections among drainages, and hence exhibit relatively recent divergence times dating to around the LGM. In contrast, these temporary coastal connections may be less accessible to the taxa inhabiting upper river portions or have more restrictive movement as a function of habitat specialization.

The high levels of endemism observed in the area (up to 94% according to Bizerril, 1994), along with several species with small and disjunct distribution and some areas presenting depauperate ichthyofauna (Ribeiro, 2006) give support for the spatial species/populations turnover in this coastal area, to the extent that diversification processes are tied to the cycles of isolation and connection that characterizes the coastal region. The high levels of genetic differentiation observed within taxa support such a mechanism (Knowles and Alvarado-Serrano, 2010). Moreover, the timing of divergences estimated here suggests that these recent demographic processes could play a role in promoting differentiation in the coastal region. This contrasts with previous studies that have proposed diversification as result of geographic events during Miocene-Pliocene time period in this region (Roxo *et al.,* 2012; Roxo *et al.,* 2014). Hence, this evidence indicates a complex history of the ichthyofauna in the region is likely framed by processes that occurred millions of years ago, but it may also be impacted by the Pleistocene climatic changes.

Relationship to diversity patterns of ichthyofauna along the Brazilian coastal basins

Previous evaluations of scenarios and population genetics studies provide supporting evidence for some of common geographic breaks found for fishes in the Brazilian coastal basins (Pereira *et al.,* 2013; Thomaz et a., 2015; Tschá *et al*, 2017). For example, the north break matches with the transition between haplogroups for a widespread species (*Hoplias malabaricus*; Pereira *et al*., 2013), indicating that this geographic break may be recovered in other species besides the ones studied here. Furthermore, the south and north breaks correspond to the boundary between freshwater ecoregions (Abell *et al.,* 2008) providing evidence that genetic differentiation within widespread species may be associated with faunal turnover. Hence, 23

species are reported to be endemic to the southern portion of this south geographic break (e.g., *Diapoma itaimbe*, *Mimagoniates rheocharis*, *Pareiorhaphis nudulus*, *Jenynsia unitaenia* and *Ituglanis boitata*; Ferrer *et al.,* 2015), including *Hollandichthys taramandahy* and *Bryconamericus lethostigmus* sampled here. This southern break is congruent with the limit between two paleodrainages reconstructed in Thomaz *et al*. (2015). Northern and southern breaks are located at previously suggested geographic barriers promoting divergence associated with geological structures on the Brazilian coast, such as the Serra do Tabuleiro that is a prominent mountain chain that extends eastward (south break), and the Cabo Frio Magmatic Lineament, which is a region characterized by mountainous relief of granite-gneiss crystalline basement (north Break; Villwock et al., 2005; Pereira *et al.,* 2013). In contrast with these two geographic breaks, the central (Paranaguá) break, has received less attention. This could be an artifact of the few population level studies we have available for the area. For example recent study, has also found evidence of this central break based on high F_{ST}-values for three freshwater species of fishes (Tschá *et al*., 2017), indicating recent genetic differentiation in populations around the Paranaguá estuary, despite shared mtDNA haplotypes.

Besides geographic breaks along the coast, river captures between coastal and inland drainages are known to have occurred during the Oligocene and Neogene (i.e., Paraíba do Sul with Upper Tietê river capture; Malabarba, 1998) and are largely used to explain species presence in both areas (Ribeiro, 2006; Ribeiro *et al.,* 2006; Menezes *et al.,* 2008, but see Buckup 2011). Here, three of the four groups (all except *Bryconamericus*) are present in at least one inland basin. The inland basins are Iguaçu (population 20), Paranapanema (population 21) and Upper Tietê (population 28). In all cases, these inland populations are closely related to coastal drainages, indicating extremely recent events of river captures. These recent stream capture events are thought to be result of recent tectonic activity on geological faults and generally hypothesized to have a direction from inland towards coastal basins, in favor of altitudinal gradient (Ribeiro 2006; Ribeiro *et al.,* 2006; Torres and Ribeiro, 2009). Our results suggest an opposite migration direction from coastal to inland basin, against an altitudinal gradient.

Since we have the presence of populations and species within the taxonomic groups studied here, it allows us to comment on the agreement between morphological species assignment and genetic data (Huang and Knowles, 2016). To demonstrate, the older and one of the youngest divergence estimates (i.e., *H. boulengeri* at the north break and *Hollandichthys* in

the south break, respectively; Figure V.3) matches with the limits of putative new species (Carvalho, 2006; Bertaco *et al*., 2013). On the contrary, in the central Paranaguá break all divergences are currently between populations. These results indicate the lack of an agreement between taxonomic status and genetic divergence. Two potential reasons may contribute to this mismatch: (i) some degree of arbitrariness associated with the taxonomic level designation, which is often based on morphology only (Huang and Knowles, 2016), or (ii) that the taxa sampled here are a good example of the diversification mosaic, with different lineages presenting distinguish speciation duration and conversion rates (Dynesius and Jansson, 2013). The latter is especially true given the recency of our divergence estimations, with divergence processes resulting in population structure or speciation (Sukumaran and Knowles, 2017), and demonstrates that Pleistocene divergence is not solely responsible for population diversification but is also responsible for species richness in this area.

References

Abell R, Thieme ML, Revenga C, Bryer M, Kottelat M, Bogutskaya N, *et al.* (2008) Freshwater ecoregions of the world: a new map of biogeographic units for freshwater biodiversity conservation. *BioScience* **58**: 403-414.

Albert JS, Carvalho TP (2011). Neogene assembly of modern faunas. *Historical biogeography of neotropical freshwater fishes. University of California Press, Berkeley*, 119-136.

Ali JR, Aitchison JC (2014). Exploring the combined role of eustasy and oceanic island thermal subsidence in shaping biodiversity on the Galápagos. *J Biogeogr* **41**: 1227-1241.

Barraclough TG, Vogler AP (2000). Detecting the geographical pattern of speciation from species-level phylogenies. *Am Nat* **155**: 419-434.

Bermingham E, Avise JC (1986). Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics* **113**: 939-965.

Bertaco VA, Malabarba LR (2013). A new species of the characid genus *Hollandichthys* Eigenmann from coastal rivers of southern Brazil (Teleostei: Characiformes) with a discussion on the diagnosis of the genus. *Neotrop Ichthyol* **11**: 767-778.

Bizerril CRSF (1994). Análise taxonômica e biogeográfica da ictiofauna de água doce do leste brasileiro. *Acta Biologica Leopoldensia* **16**: 51-80.

de Bruyn M, Rüber L, Nylinder S, Stelbrink B, Lovejoy NR, Lavoué S, Tab HH, Nugroho E, Wowor D, Ng PKL, Azizah MNS, Von Rintelen T, Hall R, Carvalho GR (2013). Palaeo-drainage basin connectivity predicts evolutionary relationships across three southeast Asian biodiversity hotspots. *Syst Biol* **62**: 398-410.

Buckup PA (2011). The eastern Brazilian shield. In: *Historical biogeography of Neotropical Freshwater Fishes*, 203-210.

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

Catchen J, Hohenlohe P, Bassham S, Amores A, Cresko W (2013). Stacks: an analysis tool set for population genomics. *Mol Ecol* **22**: 3124-3140.

Carnaval AC, Hickerson MJ, Haddad CF, Rodrigues MT, Moritz C (2009). Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*. **323**: 785-789.

Carvalho FR (2006). Taxonomia das populações de *Hyphessobrycon boulengeri* (Eigenmann, 1907) e *Hyphessobrycon reticulatus* Ellis, 1911 (Characiformes: Characidae). *Master thesis* UNESP, Brazil

Chakona A, Swartz ER, Gouws G (2013). Evolutionary drivers of diversification and distribution of a southern temperate stream fish assemblage: testing the role of historical isolation and spatial range expansion. *PloS one* **8:** 1-13.

Chifman J, Kubatko L (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics* **30**: 3317-3324.

Coyne JA (1994). Ernst Mayr and the origin of species. *Evolution* **48**: 19-30.

Cunningham CW, Collins TM (1994). Developing model systems for molecular biogeography: vicariance and interchange in marine invertebrates. In *Molecular ecology and evolution: approaches and applications* (pp. 405-433). Birkhäuser Basel.

Dias MS, Oberdorff T, Hugueny B, Leprieur F, Jézéquel C, Cornu JF, *et al*. (2014). Global imprint of historical connectivity on freshwater fish biodiversity. *Ecol Lett* **17**: 1130-1140.

Donoghue MJ, Moore BR (2003). Toward an Integrative Historical Biogeography. *Integr Comp Biol* **43**: 261-270.

Dynesius M, Jansson R (2014). Persistence of within-species lineages: a neglected control of speciation rates. *Evolution* **68**: 923-934.

Earl DA (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* **4**: 359-361.

Edwards S, Beerli P (2000). Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54**: 1839-1854.

Edwards DL, Keogh JS, Knowles LL (2012). Effects of vicariant barriers, habitat stability, population isolation and environmental features on species divergence in the south-western Australian coastal reptile community. *Mol Ecol* **21**: 3809-3822.

Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611-2620.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013). Robust demographic inference from genomic and SNP data. *PLoS Genet* **9**: 1-17.

Excoffier L, Foll M (2011). Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinform* **27**: 1332-1334.

Excoffier L, Lischer HE (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* **10**: 564-567.

Ferrer J, Donin LM, Malabarba LR (2015). A new species of *Ituglanis* Costa & Bockmann, 1993 (Siluriformes: Trichomycteridae) endemic to the Tramandaí-Mampituba ecoregion, southern Brazil. *Zootaxa* **4020**: 375-389.

Hirschmann A, Fagundes NJ, Malabarba LR (2017). Ontogenetic changes in mouth morphology triggers conflicting hypotheses of relationships in characid fishes (Ostariophysi: Characiformes). *Neotr Ichthyol* **15**:

Huang H, Knowles LL (2016). Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Syst Biol* **65**: 357-365.

Huang JP, Knowles LL (2016) The species versus subspecies conundrum: quantitative delimitation from integrating multiple data types within a single Bayesian approach in Hercules beetles. *Syst. Biol.*65: 685-699.

Irwin DE (2002). Phylogeographic breaks without geographic barriers to gene flow. *Evolution* **56**: 2383-2394.

Knowles LL, Alvarado-Serrano DF (2010). Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled

ecological, demographic and genetic models in montane grasshoppers. *Mol Ecol* **19**: 3727-3745.

Knowles LL, Carstens BC (2007) Estimating a geographically explicit model of population divergence. *Evolution* 61:477-493

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* **15**: 1179-1191.

Leite YL, Costa LP, Loss AC, Rocha RG, Batalha-Filho H, Bastos AC, *et al.* (2016). Neotropical forest expansion during the last glacial period challenges refuge hypothesis. *PNAS* **113**: 1008-1013.

Lynch M (2010). Evolution of the mutation rate. *Trends Genet* **26**: 345-352.

Malabarba MC (1998). Phylogeny of fossil Characiformes and paleobiogeography of the Tremembé formation, São Paulo, Brazil. *Phylogeny and classification of Neotropical fishes* 69-84

Massatti R, Knowles LL (2016). Contrasting support for alternative models of genomic variation based on microhabitat preference: species-specific effects of climate change in alpine sedges. *Mol Ecol* **25**: 3974-3986.

Miller KG, Mountain GS, Wright JD, Browning JV (2011). A 180-million record of sea level and ice volume variations from continental margin and deep-sea isotopic records. *Oceanography* **24**: 40-53.

Menezes NA, Ribeiro AC, Weitzman S, Torres RA (2008). Biogeography of Glandulocaudinae (Teleostei: Characiformes: Characidae) revisited: phylogenetic patterns, historical geology and genetic connectivity. *Zootaxa* **1726**: 33-48.

Menezes NA, Weitzman SH (2009). Systematics of the neotropical fish subfamily Glandulocaudinae (Teleostei: Characiformes: Characidae). *Neotr Ichthyol* **7**: 295-370.

Papadopoulou A, Knowles LL (2015). Genomic tests of the species-pump hypothesis: recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. *Evolution* **69**: 1501-1517.

Papadopoulou A, Knowles LL (2016). Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *PNAS* **113**: 8018-8024.

Parchman TL, Gompert Z, Mudge J, Schilkey FD, Benkman CW, Buerkle C (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol Ecol* **21**: 2991-3005.

Pereira TL, Santos U, Schaefer CE, Souza GO, Paiva SR, Malabarba LR, Schmidt EE, Dergam JA (2013). Dispersal and vicariance of *Hoplias malabaricus* (Bloch, 1794) (Teleostei, Erythrinidae) populations of the Brazilian continental margin. *J Biogeogr* **40**: 905-914.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one* **7**

Pezzi da Silva JF (1998). *Revisão taxonômica das espécies de Bryconamericus (Eigenmann, 1907) do Sudeste do Brasil (Ostariophysi, Characidae)* (Doctoral dissertation, Pontificia Universidade Católica do Rio Grande do Sul.).

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559-575.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Radinger J, Wolter C (2014). Patterns and predictors of fish dispersal in rivers. *Fish and Fisheries* **15**: 456-473.

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. *Neotr Ichthyol* **4**: 225–246.

Ribeiro AC, Lima FC, Riccomini C, Menezes NA (2006). Fishes of the Atlantic Rainforest of Boracéia: testimonies of the Quaternary fault reactivation within a Neoproterozoic tectonic province in Southeastern Brazil. *Ichthyol Expl Freshwaters* **17**: 157-164.

Roxo FF, Albert JS, Silva GS, Zawadzki CH, Foresti F, Oliveira C (2014). Molecular phylogeny and biogeographic history of the armored Neotropical catfish subfamilies Hypoptopomatinae, Neoplecostominae and Otothyrinae (Siluriformes: Loricariidae). *PLoS One* **9**.

Roxo FF, Zawadzki CH, Alexandrou MA, Costa Silva GJ, Chiachio MC, Foresti F, Oliveira C (2012). Evolutionary and biogeographic history of the subfamily Neoplecostominae (Siluriformes: Loricariidae). *Ecol Evol* **2**: 2438-2449.

Seehausen O, Wagner CE (2014). Speciation in freshwater fishes. *Annual Rev Ecol Evol Syst* **45**: 621-651.

Shaw KL, Gillespie RG (2016). Comparative phylogeography of oceanic archipelagos: Hotspots for inferences of evolutionary process. *PNAS* **113**: 7986-7993.

Sukumaran J, Knowles LL (2017). The multiple coalescent delimits structure not species. *PNAS* **144**: 1607-1612.

Swofford DL (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tedesco PA, Leprieur F, Hugueny B, Brosse S, Dürr HH, Beauchard *et al*. (2012). Patterns and processes of global riverine fish endemism. *Global Ecol Biogeogr* **21**: 977-987.

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

Thomaz AT, Malabarba LR, Bonatto SL, Knowles LL (2015). Testing the effect of palaeodrainages versus habitat stability on genetic divergence in riverine systems: study of a Neotropical fish of the Brazilian coastal Atlantic Forest. *J Biogeogr* **42**: 2389-2401.

Thomaz AT, Malabarba LR, Bonatto SL, Knowles LL (*in review*). Genomic signatures of paleodrainages in a freshwater fish along the southeastern coast of Brazil: genetic structure reflects past riverine properties. *Heredity*.

Thomé MTC, Zamudio KR, Haddad CF, Alexandrino J (2014). Barriers, rather than refugia, underlie the origin of diversity in toads endemic to the Brazilian Atlantic Forest. *Mol Ecol* **23**: 6152-6164.

Tschá MK, Baggio RA, Marteleto FM, Abilhoa V, Bachmann L, Boeger WA (2017). Sea-level variations have influenced the demographic history of estuarine and freshwater fishes of the coastal plain of Paraná, Brazil. *J Fish Biol* **90**: 968-979.

Torres RA, Ribeiro J (2009). The remarkable species complex *Mimagoniates microlepis* (Characiformes: Glandulocaudinae) from the Southern Atlantic Rain forest (Brazil) as revealed by molecular systematic and population genetic analyses. *Hydrobiologia*, **617**, 157-170.

Villwock JA, Lessa GC, Suguio K, Angulo RJ, Dillenburg SR (2005). Geologia e geomorfologia de regiões costeiras. *Quaternário do Brasil*, 94-113.

Waters JM, Burridge CP (2016). Fine-scale habitat preferences influence within-river population connectivity: a case-study using two sympatric New Zealand Galaxias fish species. *Freshwater Biol*, **61**: 51-56.

Weitzman SH, Menezes NA, Weitzman MJ (1988) Phylogenetic biogeography of the Glandulocaudini (Teleostei: Characiformes, Characidae) with comments on the distributions of other freshwater fishes in eastern and southeastern Brazil. *Proceedings of a workshop on neotropical distribution patterns. Academia Brasileira de Ciências.* pp. 379–427. Academia Brasileira de Ciências, Rio de Janeiro

APPENDIX D

Supplementary material from chapter V

Table D.1. Sampling and genomic sequences pre- and post-processing in STACKS per species for each individual: (a) *M. microlepis*, (B) *H. boulengeri*, (c) *Hollandichthys* and (d) *Bryconamericus*. Populations are organized from the southern most to the northern most population. * marks individuals removed from the analysis because of poor sequence quality or large numbers of missing loci.

(a) *M. microlepis*

Population	Voucher	Longitude	Latitude	Raw read count	Retained reads	Mean coverage (SD)
Maquiné	UFRGS 20507/TEC 5765F	-50.20956	-29.65217	2859537	2825580	76.5 (±218.6)
Maquiné	UFRGS 20508/TEC 5766C	-50.28592	-29.56717	2995427	2958723	63.1 (±249)
Maquiné	UFRGS 20508/TEC 5766F	-50.28592	-29.56717	2910662	2880899	68.5 (±247)
Maquiné	UFRGS 20509/TEC 5767C	-50.24617	-29.53839	3219550	3179406	59.1 (±188)
Maquiné	UFRGS 20509/TEC 5767D	-50.24617	-29.53839	2842865	2816537	66.1 (±166.2)
Três Forquilhas	UFRGS 16544/TEC 2877C	-50.12278	-29.45500	2697645	2656104	59 (±185.6)
Três Forquilhas	UFRGS 20511/TEC 5769H	-50.11389	-29.53925	2670490	2641787	54.9 (±172.5)
Três Forquilhas	UFRGS 20514/TEC 5772C	-50.09222	-29.51208	2357245	2317969	47.7 (±142.9)
Três Forquilhas	UFRGS 20514/TEC 5772E	-50.09222	-29.51208	2549843	2519964	56 (±243.7)
Três Forquilhas	UFRGS 20514/TEC 5772F	-50.09222	-29.51208	2470543	2444984	53.9 (±167.2)
Mampituba	UFRGS 20518/TEC 5776D	-49.93492	-29.31067	2173708	2146723	58.2 (±283.8)
Mampituba	UFRGS 20518/TEC 5776E	-49.93492	-29.31067	2152717	2130222	55.3 (±163.4)
Mampituba	UFRGS 20520/TEC 5778E	-50.01664	-29.23361	1976335	1957299	54.7 (±161.6)
Mampituba	UFRGS 20520/TEC 5778G	-50.01664	-29.23361	3283563	3254724	75.8 (±457.8)
Mampituba	UFRGS 20522/TEC 5780D	-49.81936	-29.07864	1898043	1873146	48.5 (±170.2)
Ararangua	UFRGS 16613/TEC 2941A	-49.26544	-28.84219	2032225	1973088	42.8 (±150)
Ararangua	UFRGS 16613/TEC 2941B	-49.26544	-28.84219	2730519	2698593	57.8 (±198.5)
Ararangua	UFRGS 20521/TEC 5779A	-49.77783	-28.99994	3274240	3240545	64.6 (±215.7)
Ararangua	UFRGS 20521/TEC 5779B	-49.77783	-28.99994	2564370	2539369	60.7 (±247.2)
D'Una	MCP 28743/A	-48.78861	-28.19972	1433083	1411758	44.6 (±90.1)
D'Una	MCP 28743/B	-48.78861	-28.19972	1408790	1391503	45 (±91.8)
D'Una	UFRGS 20500/TEC 5758A	-48.76217	-28.20914	700228	691552	23.7 (±89.4)
D'Una	UFRGS 20500/TEC 5758B	-48.76217	-28.20914	372905	366915	14.8 (±64)
Cubatão Sul	UFRGS 20444/TEC 5703A	-48.72703	-27.69833	985484	969321	30 (±86.5)
Cubatão Sul	UFRGS 20444/TEC 5703B	-48.72703	-27.69833	1264009	1253165	38.1 (±124.9)
Cubatão Sul	UFRGS 20444/TEC 5703C	-48.72703	-27.69833	1266391	1251242	38.4 (±127.7)
Cubatão Sul	UFRGS 20444/TEC 5703D	-48.72703	-27.69833	1077564	1063978	35 (±122.1)
Biguaçu	UFRGS 20451/TEC 5709A	-48.78250	-27.49544	726571	714905	25.1 (±65.2)
Biguaçu	UFRGS 20451/TEC 5709B	-48.78250	-27.49544	1535032	1515730	41.4 (±107.4)
Biguaçu	UFRGS 20453/TEC 5711A	-48.78639	-27.49250	2002923	1980246	52.7 (±180)
Biguaçu	UFRGS 20453/TEC 5711B	-48.78639	-27.49250	1748684	1726628	46.6 (±154.1)

Tijucas	UFRGS 20456/TEC 5714	-48.72239	-27.28267	1883904	1859687	57.9 (±163.5)
Tijucas	UFRGS 20457/TEC 5715	-48.98083	-27.43361	450039	442053	16.9 (±44.4)
Tijucas	UFRGS 20458/TEC 5716	-48.97417	-27.29444	1488172	1470200	42 (±137.2)
Tijucas	UFRGS 18498/TEC 3879	-48.82083	-27.26667	1155380	1138042	34.2 (±188.8)
Itajaí	UFRGS 18598/TEC 3891	-48.94769	-27.19461	2059760	2035570	46.9 (±149.9)
Itajaí*	UFRGS 20464/TEC 5722	-48.71486	-26.99508	2342550	1230770	33.9 (±68.1)
Itajaí	UFRGS 18601/TEC 3894A	-48.71444	-27.00694	1315012	1277632	33.5 (±176.4)
Itajaí	UFRGS 18601/TEC 3894B	-48.71444	-27.00694	2708215	2675608	55.3 (±155.2)
São Francisco do Sul	MCP 31800/A	-48.65111	-26.32667	1518743	1500492	42.5 (±65.8)
São Francisco do Sul	MCP 31800/B	-48.65111	-26.32667	1341107	1323749	40 (±64)
São Francisco do Sul	UFRGS 10697/TEC 134	-48.58917	-26.29306	1948766	1904294	46.3 (±104.7)
Saí-Mirim	UFRGS 20473/TEC 5731A	-48.63267	-26.08328	1320046	1304121	28.6 (±62)
Saí-Mirim	UFRGS 20473/TEC 5731B	-48.63267	-26.08328	2375129	2341127	50.8 (±124.8)
Guaratuba	UFRGS 14580/TEC 476A	-48.70000	-25.95944	2452671	2425735	68.6 (±123.1)
Guaratuba	UFRGS 14580/TEC 476B	-48.70000	-25.95944	1606961	1585520	51.7 (±83.8)
Guaratuba	UFRGS 18638/TEC 3944	-48.83589	-25.97833	2391122	2363516	47.5 (±115.3)
Guaratuba	UFRGS 18650/TEC 3956	-48.83472	-25.97472	2422777	2389807	56.8 (±326.9)
Guaraguaçu	UFRGS 18592/TEC 3874	-48.59361	-25.72639	1849381	1827440	59.8 (±152.0)
Guaraguaçu	UFRGS 18506/TEC 3968	-48.59028	-25.73917	3230690	3197946	35.4 (±61.8)
Iguaçu*	UFRGS 12859/TEC 1569	-49.79194	-25.69944	120845	99782	7.1 (±13.2)
Iguaçu	UFRGS 12872/TEC 1582	-49.77139	-25.72361	1504811	1484685	45.5 (±72.5)
Iguaçu*	UFRGS 12874/TEC 1584	-50.07333	-26.33722	4895	3371	7.3 (±5)
Iguaçu	UFRGS 12878/TEC 1588	-49.88639	-25.54389	1167724	1151084	32 (±98.6)
Paranapanema	UFRGS 12862/TEC 1572A	-50.38639	-25.07333	880308	872109	26.4 (±45.4)
Paranapanema	UFRGS 12862/TEC 1572B	-50.38639	-25.07333	1308867	1282801	32.7 (±76.2)
Paranapanema	UFRGS 12862/TEC 1572C	-50.38639	-25.07333	1434937	1414982	36 (±72.9)
Paranaguá	UFRGS 18663/TEC 3979	-48.83472	-25.45806	1541078	1523201	35.4 (±61.8)
Paranaguá*	MCP 31816	-48.77472	-25.34139	1377439	1361490	38.7 (±63.1)
Paranaguá	MCP 31813	-48.87472	-25.42222	1925707	1895774	42.3 (±94.6)
Paranaguá	UFRGS 14708/TEC 780	-48.74503	-25.30958	1564502	1539504	36.4 (±56)
Guaraqueçaba	UFRGS 14698/TEC 757A	-48.45008	-25.22019	1746477	1730928	50.4 (±113.7)
Guaraqueçaba	UFRGS 14698/TEC 757B	-48.45008	-25.22019	1909453	1888327	50.6 (±131.1)
Guaraqueçaba	UFRGS 14702/TEC 767A	-48.40778	-25.18111	2077321	2051600	44.6 (±68.5)
Guaraqueçaba	UFRGS 14702/TEC 767B	-48.40778	-25.18111	2179690	1387017	32.8 (±49.6)
Ribeira de Iguape	MCP 31804	-48.61083	-24.60306	1363858	813391	22.5 (±46)
Ribeira de Iguape	UFRGS 18680/TEC 4320	-48.29333	-24.64306	1939279	1917803	43.9 (±108.9)
Ribeira de Iguape	MCP 31836	-48.09639	-24.54861	1239882	1225970	34.1 (±116.2)
Ribeira de Iguape	UFRGS 20481/TEC 5739	-47.89050	-24.69406	2187236	2153954	54.1 (±212.5)
Ribeira de Iguape	UFRGS 12422/TEC 807	-47.42639	-24.66139	3036692	3005245	62.6 (±291.2)
Ribeira de Iguape	UFRGS 18681/TEC 4324	-47.85361	-24.31611	1011939	997781	25.4 (±46.9)
Ribeira de Iguape	UFRGS 20483/TEC 5741	-47.62014	-24.25333	1762463	1738957	39.7 (±72.1)
Ribeira de Iguape	UFRGS 20485/TEC 5743	-47.61583	-24.33361	1993479	1969625	49.3 (±162.4)
Guarau	UFRGS 18716/TEC 4366A	-47.06722	-24.37472	2263364	2237431	51.3 (±120.7)
Guarau	UFRGS 18716/TEC 4366B	-47.06722	-24.37472	1799823	1780539	45.9 (±104.2)
Itanhaém	UFRGS 18711/TEC 4361	-46.99111	-24.24167	1045456	1031341	27.4 (±57.9)
Itanhaém	UFRGS 18703/TEC 4353	-47.00500	-24.18333	1272908	1252251	31.5 (±59.1)
Itanhaém	UFRGS 20490/TEC 5748	-46.72389	-24.11036	2657668	2623908	57.8 (±173)
Santos	MCP 31806/A	-46.33528	-23.85139	1276996	1253657	30 (±71.6)
Santos	MCP 31806/B	-46.33528	-23.85139	953269	935027	30.1 (±60)
Santos	UFRGS 20492/TEC 5750A	-46.19472	-23.88000	1487001	1465111	36.7 (±65.1)
Santos	UFRGS 20492/TEC 5750B	-46.19472	-23.88000	1101112	1073421	28.4 (±45.6)
Bertioga	UFRGS 20494/TEC 5752A	-45.81733	-23.74300	1518547	1497150	36.2 (±125.4)
Bertioga	UFRGS 20494/TEC 5752B	-45.81733	-23.74300	1743256	1724231	39.5 (±64.5)

Bertioga	UFRGS 20494/TEC 5752C	-45.81733	-23.74300	2199766	2172383	46.1 (±72.2)
Bertioga	UFRGS 20494/TEC 5752D	-45.81733	-23.74300	1277157	1264328	30.7 (±46.9)
Ubatuba	UFRGS 18770/TEC 4412A	-45.11333	-23.41333	1664401	1645904	39.1 (±84.3)
Ubatuba	UFRGS 18770/TEC 4412B	-45.11333	-23.41333	559203	550934	51.9 (±140)
Ubatuba	UFRGS 14770/TEC 862A	-45.05917	-23.40917	1379008	1343594	38.7 (±74.4)
Ubatuba	UFRGS 14770/TEC 862B	-45.05917	-23.40917	2524168	2489353	17.2 (±28.7)
Picinguaba	UFRGS 18776/TEC 4418	-44.88056	-23.31250	909183	899293	25.6 (±45.2)
Picinguaba*	UFRGS 12895/TEC 902	-44.87000	-23.35000	1681322	1665167	53.4 (±231.3)
Picinguaba	UFRGS 14775/TEC 869	-44.85333	-23.35222	2276504	2249902	50.6 (±111.2)
Picinguaba	UFRGS 18779/TEC 4421	-44.85122	-23.35014	844079	823133	24 (±34.4)
Jaceruba	UFRGS 18808/TEC 4444	-43.56889	-22.60889	980957	961999	25.4 (±44.8)
Guapimirim	UFRGS 18850/TEC 4480	-42.89861	-22.51778	1288668	1276028	31.7 (±48.9)
Macacu	UFRGS 18833/TEC 4464	-42.64850	-22.45106	639440	632018	18.4 (±28.7)
Macacu	UFRGS 18846/TEC 4476	-42.65722	-22.47889	1709758	1687195	39.3 (±65.2)
Paraíba do Sul	UFRGS 18818/TEC 4453A	-42.86997	-22.30334	1121087	1107832	30.9 (±56.7)
Paraíba do Sul	UFRGS 18818/TEC 4453B	-42.86997	-22.30334	946412	929782	26.9 (±54.3)
Paraíba do Sul	UFRGS 18818/TEC 4453C	-42.86997	-22.30334	1649489	1621937	34.6 (±71.1)
Paraíba do Sul	UFRGS 18818/TEC 4453D	-42.86997	-22.30334	559019	548959	18 (±32.3)
Barra do Riacho	UFRGS 18968/TEC 4575A	-40.16550	-19.78003	1752270	1730773	47 (±76)
Barra do Riacho	UFRGS 18968/TEC 4575B	-40.16550	-19.78003	1254498	1228029	32 (±50.7)
Barra do Riacho	UFRGS 18968/TEC 4575C	-40.16550	-19.78003	1926531	1897578	47 (±80.4)
Barra do Riacho	UFRGS 18968/TEC 4575D	-40.16550	-19.78003	1805245	1786162	46.5 (±79.4)
Barra Seca	UFRGS 14813/TEC 926A	-39.84083	-18.93694	1626063	1600532	45.9 (±86.8)
Barra Seca	UFRGS 14813/TEC 926B	-39.84083	-18.93694	1624248	1606357	46.2 (±89)
Barra Seca	UFRGS 11102/TEC 931A	-39.89806	-18.89750	1982433	1959121	46.5 (±88.4)
Barra Seca	UFRGS 11102/TEC 931B	-39.89806	-18.89750	1319481	1294005	35.2 (±62.7)
Itaunas	UFRGS 19022/TEC 4622A	-39.79444	-18.29639	1089820	1075159	31.3 (±53.8)
Itaunas	UFRGS 19022/TEC 4622B	-39.79444	-18.29639	1817530	1796595	44.2 (±90.3)
Itaunas	UFRGS 19028/TEC 4626A	-39.91917	-18.30778	1163679	1144700	34.7 (±60.8)
Itaunas	UFRGS 19028/TEC 4626B	-39.91917	-18.30778	1846484	1823780	47.6 (±80.8)
Caravelas	UFRGS 14823/TEC 941A	-39.44833	-17.67889	2226359	2200474	54.9 (±110)
Caravelas	UFRGS 14823/TEC 941B	-39.44833	-17.67889	1661928	1639185	39.7 (±70.7)
Caravelas	UFRGS 14823/TEC 941C	-39.44833	-17.67889	1333590	1319162	36.5 (±63.3)
Caravelas	UFRGS 14823/TEC 941D	-39.44833	-17.67889	2545746	2499483	54.3 (±107.5)

(b) *H. boulengeri*

Population	Voucher	Longitude	Latitude	Raw read count	Retained reads	Mean coverage (SD)
Patos	UFRGS 12518/TEC 1336A	-51.295556	-31.603333	1847673	1788023	42 (±41.3)
Patos	UFRGS 12518/TEC 1336B	-51.295556	-31.603333	1277945	1238633	34.2 (±33.1)
Patos	UFRGS 12518/TEC 1336C	-51.295556	-31.603333	3683342	3527546	64.2 (±86.7)
Patos	UFRGS 12527/TEC 1345A	-52.133056	-31.382500	1165996	1131511	31 (±29.4)
Patos	UFRGS 12527/TEC 1345B	-52.133056	-31.382500	3261960	3193781	57.8 (±70.6)
Patos	UFRGS 12527/TEC 1345C	-52.133056	-31.382500	2673491	2625853	57.4 (±67)
Patos	UFRGS 12528/TEC 1346A	-51.513889	-30.556389	1584103	1538907	36.8 (±39.6)
Patos	UFRGS 12528/TEC 1346B	-51.513889	-30.556389	2987324	2928618	42.7 (±60.3)
Patos	UFRGS 15546/TEC 1950A	-50.849889	-30.095694	1082592	1044864	28.9 (±25)
Patos	UFRGS 15546/TEC 1950B	-50.849889	-30.095694	1068606	1036985	30.5 (±29.1)
Patos	UFRGS 15546/TEC 1950C	-50.849889	-30.095694	1389194	1358850	35.1 (±35)
Patos	UFRGS 15546/TEC 1950D	-50.849889	-30.095694	1848274	1810718	47.8 (±52.2)
Patos	UFRGS 12454/TEC 698	-52.536333	-30.211944	1918460	1852380	42.7 (±46.3)
Tramandaí	UFRGS 19203/TEC 4898	-50.442494	-30.535433	886449	857102	25.3 (±21.4)

Tramandaí	UFRGS 19201/TEC 4896A	-50.294433	-30.213375	1525371	1475527	35.1 (±37.9)
Tramandaí	UFRGS 19201/TEC 4896B	-50.294433	-30.213375	1364819	1321166	34.4 (±36.3)
Tramandaí	UFRGS 19201/TEC 4896C	-50.294433	-30.213375	1895721	1833593	40.7 (±49.8)
Tramandaí	UFRGS 19201/TEC 4896D	-50.294433	-30.213375	558082	547335	19.8 (±18)
Tramandaí	UFRGS 19201/TEC 4896E	-50.294433	-30.213375	3426471	3339351	65.4 (±88.4)
Tramandaí	UFRGS 19201/TEC 4896F	-50.294433	-30.213375	1470285	1437608	39.7 (±44.2)
Três Forquilhas	UFRGS 19131/TEC 4850A	-49.968041	-29.613938	1833999	1772018	39.6 (±45.9)
Três Forquilhas	UFRGS 19131/TEC 4850B	-49.968041	-29.613938	1287638	1261596	33.9 (±37.9)
Três Forquilhas	UFRGS 19131/TEC 4850C	-49.968041	-29.613938	2382929	2333741	53.4 (±71.9)
D'Una	MCP 28748/-	-48.679167	-28.080833	1195766	1158358	32 (±29.6)
Cubatão Sul	UFRGS 20443/TEC 5702	-48.727028	-27.698333	1907120	1879187	18.5 (±25.9)
Florianópolis	UFRGS 20498/TEC 5756A	-48.518750	-27.685250	1240256	1197094	33.4 (±32.6)
Florianópolis	UFRGS 20498/TEC 5756B	-48.518750	-27.685250	542035	522700	18.2 (±13.8)
Florianópolis	UFRGS 20498/TEC 5756C	-48.518750	-27.685250	567089	548360	19.4 (±15.8)
Florianópolis	UFRGS 20498/TEC 5756D	-48.518750	-27.685250	644879	624246	21.2 (±17.5)
Florianópolis	UFRGS 20498/TEC 5756E	-48.518750	-27.685250	1970387	1935240	54.2 (±57.3)
Florianópolis	UFRGS 20498/TEC 5756F	-48.518750	-27.685250	1842212	1803438	48.5 (±51.8)
Florianópolis	UFRGS 20498/TEC 5756G	-48.518750	-27.685250	1210673	1184078	38.7 (±37.9)
Florianópolis	UFRGS 20498/TEC 5756H	-48.518750	-27.685250	2147120	2104987	55.8 (±63)
Itajaí	UFRGS 18609/TEC 3902	-48.714444	-27.006944	690013	667119	20.3 (±15.1)
Itapocu	UFRGS 18616/TEC 3909A	-48.848611	-26.443056	880315	853196	23.9 (±20.9)
Itapocu	UFRGS 18616/TEC 3909B	-48.848611	-26.443056	957989	929415	25.1 (±22)
Itapocu	UFRGS 18616/TEC 3909C	-48.848611	-26.443056	1116077	1073124	28 (±23.9)
Itapocu	UFRGS 18616/TEC 3909D	-48.848611	-26.443056	747371	723963	20.8 (±16.7)
Itapocu	UFRGS 18616/TEC 3909E	-48.848611	-26.443056	1065569	1042031	27.4 (±24.6)
Itapocu	UFRGS 18616/TEC 3909F	-48.848611	-26.443056	802157	783185	23.6 (±18.3)
Itapocu	UFRGS 18616/TEC 3909G	-48.848611	-26.443056	1380793	1356729	38 (±33.6)
Itapocu	UFRGS 18616/TEC 3909H	-48.848611	-26.443056	830935	812535	24.4 (±19.9)
Itapocu	UFRGS 18613/TEC 3906A	-48.726500	-26.383139	907116	882078	24 (±20.5)
Itapocu	UFRGS 18613/TEC 3906B	-48.726500	-26.383139	423667	415579	15.3 (±10.7)
Sao Francisco do Sul	UFRGS 10571/TEC 0126A	-48.589167	-26.293056	890002	864059	25.3 (±19.9)
Sao Francisco do Sul	UFRGS 10571/TEC 0126B	-48.589167	-26.293056	1349021	1305196	32.2 (±28.5)
Sao Francisco do Sul	UFRGS 10571/TEC 0126C	-48.589167	-26.293056	989664	973667	28.6 (±24.6)
Sao Francisco do Sul	UFRGS 10571/TEC 0126D	-48.589167	-26.293056	3678677	3625315	76.4 (±95)
Sao Francisco do Sul	UFRGS 10563/TEC 0120A	-48.588611	-26.286667	1232286	1195594	31 (±28.5)
Sao Francisco do Sul	UFRGS 10563/TEC 0120B	-48.588611	-26.286667	1156274	1120625	30.3 (±25.8)
Sao Francisco do Sul	UFRGS 10563/TEC 0120C	-48.588611	-26.286667	2060354	2017742	49.7 (±50.5)
Sao Francisco do Sul	UFRGS 10563/TEC 0120D	-48.588611	-26.286667	1931096	1894277	49.9 (±50.1)
Saí-Mirim	UFRGS 20472/TEC 5730A	-48.632667	-26.083278	928179	901751	26.4 (±22.9)
Saí-Mirim	UFRGS 20472/TEC 5730B	-48.632667	-26.083278	1057938	1016252	27.8 (±24.6)
Saí-Mirim	UFRGS 20472/TEC 5730C	-48.632667	-26.083278	550306	533183	18.3 (±13.6)
Saí-Mirim	UFRGS 20472/TEC 5730D	-48.632667	-26.083278	1491290	1442995	36.8 (±35.7)
Saí-Mirim	UFRGS 20472/TEC 5730E	-48.632667	-26.083278	2205001	2175225	59.5 (±60.9)
Saí-Mirim	UFRGS 20472/TEC 5730F	-48.632667	-26.083278	1638328	1618616	48.2 (±47.1)
Saí-Mirim	UFRGS 20472/TEC 5730G	-48.632667	-26.083278	1363790	1341654	41.1 (±38.9)
Saí-Mirim	UFRGS 20472/TEC 5730H	-48.632667	-26.083278	1092912	1071841	34.4 (±28.2)
Sai-Guaçu	UFRGS 20529/TEC 5787A	-48.604333	-25.963167	1416489	1363476	35.8 (±34.9)
Sai-Guaçu	UFRGS 20529/TEC 5787C	-48.604333	-25.963167	1631849	1580518	39.2 (±37.6)
Sai-Guaçu	UFRGS 20529/TEC 5787D	-48.604333	-25.963167	1808793	1750390	42.1 (±43.3)
Sai-Guaçu	UFRGS 20529/TEC 5787E	-48.604333	-25.963167	2222668	2177970	52.7 (±59.3)
Sai-Guaçu	UFRGS 20529/TEC 5787F	-48.604333	-25.963167	2043184	2000026	53.3 (±56.9)
Sai-Guaçu	UFRGS 20529/TEC 5787G	-48.604333	-25.963167	2349755	2299787	56.7 (±63.7)
Sai-Guaçu	UFRGS 20529/TEC 5787H	-48.604333	-25.963167	2229059	2181937	53.6 (±59.6)

Sai-Guaçu	UFRGS 20529/TEC 5787J	-48.604333	-25.963167	2242437	2166865	45.7 (±51.4)
Guaratuba	UFRGS 18643/TEC 3949	-48.834722	-25.974722	1721277	1669595	40.8 (±40.3)
Guaratuba	UFRGS 13825/TEC 108	-48.728056	-25.923889	590609	576956	20.8 (±16.1)
Guaratuba	UFRGS 10562/TEC 0110	-48.586111	-25.906667	1855602	1798195	40.4 (±40.6)
Iguaçu	UFRGS 12875/TEC 1585A	-50.073333	-26.337222	1239601	1203290	30.3 (±25.7)
Iguaçu	UFRGS 12875/TEC 1585B	-50.073333	-26.337222	689033	670158	20.2 (±15.4)
Iguaçu	UFRGS 12875/TEC 1585C	-50.073333	-26.337222	1043123	1013865	27.4 (±23.7)
Iguaçu	UFRGS 12875/TEC 1585D	-50.073333	-26.337222	590078	573422	20 (±14.4)
Iguaçu	UFRGS 12875/TEC 1585E	-50.073333	-26.337222	1139137	1112616	32.9 (±26.6)
Iguaçu	UFRGS 12875/TEC 1585F	-50.073333	-26.337222	1411411	1384906	37 (±32.8)
Iguaçu	UFRGS 12875/TEC 1585G	-50.073333	-26.337222	2057434	2017790	49.9 (±49.9)
Paranaguá	UFRGS 12444/TEC 719A	-48.659778	-25.314806	1479810	1434631	36.5 (±32.1)
Paranaguá	UFRGS 12444/TEC 719B	-48.659778	-25.314806	901025	880570	28.6 (±21.6)
Paranaguá	UFRGS 12444/TEC 719C	-48.659778	-25.314806	1859826	1823965	49.7 (±44.9)
Ribeira de Iguape	UFRGS 14713/TEC 796A	-47.780750	-24.789500	1311412	1268394	31.4 (±28.6)
Ribeira de Iguape	UFRGS 14713/TEC 796B	-47.780750	-24.789500	2772614	2722505	60.8 (±72.5)
Ribeira de Iguape	UFRGS 12378/TEC 798	-47.598611	-24.730000	599369	578393	17.7 (±14.4)
Ribeira de Iguape*	UFRGS 12420/TEC 792	-47.890278	-24.694167	216428	210442	9.9 (±6.9)
Ribeira de Iguape	UFRGS 20480/TEC 5738	-47.890500	-24.694056	1536032	1480825	33.2 (±34.7)
Ribeira de Iguape	UFRGS 14724/TEC 814	-47.426389	-24.661389	1874361	1827227	44.8 (±48.4)
Ribeira de Iguape	UFRGS 12397/TEC 735	-47.532833	-24.078028	1422585	1378170	34.1 (±32.9)
Ribeira de Iguape	UFRGS 18690/TEC 4334A	-47.744167	-24.283611	1406567	1360579	34.1 (±32.6)
Ribeira de Iguape	UFRGS 18690/TEC 4334B	-47.744167	-24.283611	2104860	2047396	48 (±53.8)
Ribeira de Iguape	UFRGS 18690/TEC 4334C	-47.744167	-24.283611	1308497	1277420	35.2 (±33.9)
Itanhaém	UFRGS 18712/TEC 4362A	-46.991111	-24.241667	899229	870778	24.3 (±18.6)
Itanhaém	UFRGS 18712/TEC 4362B	-46.991111	-24.241667	1233645	1194107	30.4 (±26.4)
Itanhaém	UFRGS 18712/TEC 4362C	-46.991111	-24.241667	892012	871549	27.6 (±23.2)
Itanhaém	UFRGS 20487/TEC 5745A	-46.923611	-24.177472	1469953	1423861	34.7 (±31.1)
Itanhaém	UFRGS 20487/TEC 5745B	-46.923611	-24.177472	3345766	3274386	62.2 (±81.9)
Itanhaém	UFRGS 20487/TEC 5745C	-46.923611	-24.177472	3223135	3153918	60.9 (±78.3)
Itanhaém*	MCP 31741/-	-46.732778	-24.116944	1093267	626572	12.7 (±48.8)
Itanhaém	UFRGS 20489/TEC 5747A	-46.723889	-24.110361	1307613	1264536	32.5 (±28.2)
Itanhaém	UFRGS 20489/TEC 5747B	-46.723889	-24.110361	1393500	1347426	34.2 (±30.6)
Itanhaém	UFRGS 20489/TEC 5747C	-46.723889	-24.110361	4076021	3989460	69.1 (±103.5)
Upper Tietê	UFRGS 12432/TEC 829	-46.313750	-23.772528	1647062	1595317	40.1 (±39.1)
Upper Tietê	UFRGS 12427/TEC 825A	-46.334722	-23.766389	1331282	1289472	36.3 (±32.6)
Upper Tietê	UFRGS 12427/TEC 825B	-46.334722	-23.766389	469645	453753	15.2 (±11.4)
Upper Tietê	UFRGS 12427/TEC 825C	-46.334722	-23.766389	612485	595898	19 (±14.7)
Upper Tietê	UFRGS 12427/TEC 825D	-46.334722	-23.766389	2755277	2694254	56.1 (±68.2)
Upper Tietê	UFRGS 12427/TEC 825E	-46.334722	-23.766389	2146580	2105405	35.4 (±45.1)
Upper Tietê	UFRGS 12427/TEC 825F	-46.334722	-23.766389	1396475	1363107	39.4 (±36.4)
Upper Tietê	UFRGS 12427/TEC 825G	-46.334722	-23.766389	1763142	1726593	47.8 (±47.4)
Bertioga	UFRGS 12408/TEC 752A	-46.003153	-23.779469	1046977	1008794	25.9 (±20.7)
Bertioga	UFRGS 12408/TEC 752B	-46.003153	-23.779469	1093755	1060987	27.9 (±22.7)
Bertioga	UFRGS 12408/TEC 752C	-46.003153	-23.779469	2348598	2279306	46.4 (±47.9)
Bertioga	UFRGS 12408/TEC 752D	-46.003153	-23.779469	1504611	1449840	31.2 (±30.3)
Bertioga	UFRGS 12408/TEC 752E	-46.003153	-23.779469	1563612	1522542	40 (±36.2)
Bertioga	UFRGS 12408/TEC 752F	-46.003153	-23.779469	2659199	2600009	54.3 (±59.2)
Bertioga	UFRGS 12408/TEC 752G	-46.003153	-23.779469	1721657	1691182	46.4 (±40.7)
Bertioga	UFRGS 12408/TEC 752H	-46.003153	-23.779469	3538443	3467198	64.8 (±77.5)
Guapimirim	UFRGS 18866/TEC 4493A	-42.898611	-22.517778	1339187	1297980	31.7 (±28.2)
Guapimirim	UFRGS 18866/TEC 4493B	-42.898611	-22.517778	1105006	1085422	32 (±25)
Guapimirim	UFRGS 18866/TEC 4493C	-42.898611	-22.517778	1365621	1341490	38.1 (±31.8)

Barra do Riacho	UFRGS 18971/TEC 4578A	-40.165500	-19.780028	1177801	1139150	32.4 (±25.8)
Barra do Riacho	UFRGS 18971/TEC 4578B	-40.165500	-19.780028	489321	473156	16.4 (±10.8)
Barra do Riacho	UFRGS 18971/TEC 4578C	-40.165500	-19.780028	666778	645884	21.1 (±15.2)
Barra do Riacho	UFRGS 18971/TEC 4578D	-40.165500	-19.780028	671894	655108	23.5 (±19)
Barra do Riacho	UFRGS 18971/TEC 4578E	-40.165500	-19.780028	510501	499673	18.2 (±13.3)
Barra do Riacho	UFRGS 18971/TEC 4578F	-40.165500	-19.780028	597704	585393	22.5 (±17.4)
Barra do Riacho	UFRGS 18971/TEC 4578G	-40.165500	-19.780028	2271118	2227180	53.7 (±58.4)
Barra Seca	UFRGS 11106/TEC 935A	-39.81083333	-18.63083333	1064153	1030541	29.9 (±23.4)
Barra Seca	UFRGS 11106/TEC 935B	-39.81083333	-18.63083333	1282813	1245387	35.8 (±29.1)
Barra Seca	UFRGS 11106/TEC 935C	-39.81083333	-18.63083333	1846537	1816357	53.8 (±53.4)
Barra Seca	UFRGS 11106/TEC 935D	-39.81083333	-18.63083333	2344462	2289974	55.3 (±61.8)
Barra Seca	UFRGS 11106/TEC 935E	-39.81083333	-18.63083333	2130634	2088276	57 (±58.4)
Barra Seca	UFRGS 19005/TEC 4609	-39.793889	-18.543333	1764047	1706487	41.4 (±38.5)
Itaunas	UFRGS 19017/TEC 4617A	-39.794444	-18.296389	1201344	1163117	31.7 (±25.9)
Itaunas	UFRGS 19017/TEC 4617B	-39.794444	-18.296389	2657271	2592633	57.5 (±67.6)
Caravelas	UFRGS 14820/TEC 938	-39.44833333	-17.67888889	1949737	1902617	52 (±56.2)

(c) *Hollandichthys*

Population	Voucher	Longitude	Latitude	Raw read count	Retained reads	Mean coverage (SD)
Maquiné	UFRGS 11793/TEC842 A	-50.2933	-29.5906	1823115	1510955	15.1 (±45.8)
Maquiné	UFRGS 11793/TEC842 B	-50.2933	-29.5906	1267088	1087900	13 (±32.4)
Maquiné	UFRGS 11793/TEC842 C	-50.2933	-29.5906	2054826	1733752	16.2 (±45.5)
Maquiné	UFRGS 11793/TEC842 D	-50.2933	-29.5906	1922076	1692089	15.8 (±46.7)
Maquiné	UFRGS 11793/TEC842 E	-50.2933	-29.5906	2375671	2065855	18.3 (±63.1)
Maquiné	UFRGS 11793/TEC842 F	-50.2933	-29.5906	2542248	1678713	17.8 (±50.9)
Maquiné	UFRGS 11793/TEC842 G	-50.2933	-29.5906	1976228	1683519	14.9 (±41.2)
Maquiné	UFRGS 11793/TEC842 H	-50.2933	-29.5906	2039575	1780893	15.9 (±45.4)
Maquiné	UFRGS 11793/TEC842 I	-50.2933	-29.5906	2178200	1941389	23.3 (±60.9)
Maquiné	MCP26969	-50.2833	-29.5833	1196972	854115	7.7 (±69.4)
Três Forquilhas*	UFRGS 16513/TEC2841 A	-50.1789	-29.4267	1128460	966580	17.9 (±57)
Três Forquilhas	UFRGS 16513/TEC2841 B	-50.1789	-29.4267	1234567	931387	9.1 (±18)
Três Forquilhas	UFRGS 16513/TEC2841 C	-50.1789	-29.4267	1272236	1070280	14.9 (±40.6)
Três Forquilhas	UFRGS 16513/TEC2841 D	-50.1789	-29.4267	1019104	815821	10.6 (±21.3)
Três Forquilhas	UFRGS 16513/TEC2841 E	-50.1789	-29.4267	974598	707924	7.4 (±17.7)
Três Forquilhas	UFRGS 16513/TEC2841F	-50.1789	-29.4267	1121395	894136	9.8 (±18.8)
Três Forquilhas	UFRGS 16513/TEC2841 G	-50.1789	-29.4267	1352268	1025133	9.9 (±27.5)
Três Forquilhas*	UFRGS 16513/TEC2841 H	-50.1789	-29.4267	1286903	1050800	13 (±29.3)
Mampituba	UFRGS 11792/TEC841 A	-50.1167	-29.2528	2271189	2019993	18 (±56.5)
Mampituba	UFRGS 11792/TEC841 B	-50.1167	-29.2528	2993304	2662523	19.1 (±57.8)
Mampituba	UFRGS 11792/TEC841 C	-50.1167	-29.2528	1701233	1480050	14.9 (±42.2)
Mampituba	UFRGS 11792/TEC841 D	-50.1167	-29.2528	1385623	1156003	12 (±30.7)
Mampituba	UFRGS 11792/TEC841 E	-50.1167	-29.2528	1814810	1516244	13.9 (±36.8)
Mampituba	UFRGS 11792/TEC841 F	-50.1167	-29.2528	1749054	1464074	13.8 (±36.8)
Mampituba	UFRGS 11792/TEC841 G	-50.1167	-29.2528	1405917	1095388	11.9 (±30.7)
Mampituba	UFRGS 11792/TEC841 H	-50.1167	-29.2528	908667	709308	9 (±19.9)
Mampituba	MCP23625	-49.9969	-29.1703	1688056	1300026	11.6 (±562.5)
Mampituba	UFRGS 11790/TEC839	-49.9822	-29.1692	1628963	1381686	14 (±52.1)
Araranguá*	UFRGS 11791/TEC840 A	-49.9289	-28.7908	407508	227797	-
Araranguá	UFRGS 11791/TEC840 B	-49.9289	-28.7908	1170918	964684	12 (±34.6)
Araranguá	UFRGS 11791/TEC840 C	-49.9289	-28.7908	1466012	1256026	12.2 (±44.5)
Araranguá	UFRGS 11791/TEC840 D	-49.9289	-28.7908	1463739	1159837	13 (±39.2)

D'Una	UFRGS 16587/TEC2919 A	-48.7005	-28.0715	2783833	2469992	20.1 (±91.6)
D'Una	UFRGS 16587/TEC2919 B	-48.7005	-28.0715	1774893	1545472	18.2 (±55.2)
D'Una	UFRGS 16587/TEC2919 C	-48.7005	-28.0715	1489435	1230239	12.8 (±41.1)
D'Una*	UFRGS 16587/TEC2919 D	-48.7005	-28.0715	2483493	2231984	26.6 (±81.8)
D'Una	UFRGS 16587/TEC2919 E	-48.7005	-28.0715	2135018	1850607	18.3 (±52.3)
D'Una*	UFRGS 16587/TEC2919 F	-48.7005	-28.0715	1296886	1108358	14.4 (±36.4)
D'Una	UFRGS 16587/TEC2919 G	-48.7005	-28.0715	1062519	836586	7.6 (±57)
D'Una*	UFRGS 16587/TEC2919 H	-48.7005	-28.0715	1101909	929652	13.6 (±31.4)
Florianópolis	MCP28737	-48.4775	-27.5881	1571100	1155499	10 (±408.6)
Florianópolis	MCP28737	-48.4775	-27.5881	1500857	1148858	13.1 (±955.2)
Florianópolis	MCP28737	-48.4775	-27.5881	1128815	909977	10.6 (±203.6)
Florianópolis	MCP28732	-48.4864	-27.5111	1839766	1628125	8.8 (±105.6)
Florianópolis	MCP28732	-48.4864	-27.5111	1438780	1085898	12.2 (±53.7)
Florianópolis	MCP28747	-48.4925	-27.5086	1353998	965540	9.8 (±188.5)
Florianópolis	MCP38317	-48.4372	-27.4831	1355850	962147	9 (±74)
Florianópolis*	MCP37635/100	-48.4231	-27.4831	812468	604630	9 (±99.5)
Florianópolis*	MCP37635/101	-48.4231	-27.4831	851942	507815	7.3 (±61.8)
Florianópolis	MCP37635/102	-48.4231	-27.4831	1098851	871131	9.5 (±75.1)
Itapocu	MCP30552	-48.7258	-26.3828	1209118	946763	10 (±396.1)
São Francisco do Sul	MCP30553	-48.6511	-26.3267	1358622	1041275	10.4 (±314.8)
São Francisco do Sul*	UFRGS 10579/TEC105	-48.6511	-26.3267	455531	247839	-
São Francisco do Sul	UFRGS 9359/TEC345	-48.58916667	-26.29305556	1484971	1257981	12.8 (±57.2)
São Francisco do Sul	UFRGS 10570/TEC106 A	-48.58916667	-26.29305556	2738799	2506161	21.2 (±59.7)
São Francisco do Sul	UFRGS 10570/TEC106 B	-48.58916667	-26.29305556	3186947	2892456	21.8 (±62)
São Francisco do Sul*	UFRGS 10570/TEC106 C	-48.58916667	-26.29305556	1525797	1326424	16.9 (±37)
São Francisco do Sul	UFRGS 10570/TEC106 D	-48.58916667	-26.29305556	2686698	2495012	17.5 (±48.6)
São Francisco do Sul	UFRGS 10570/TEC106 E	-48.58916667	-26.29305556	2700678	2429372	21 (±57.6)
São Francisco do Sul	UFRGS 10570/TEC106 F	-48.58916667	-26.29305556	2708515	2440012	23 (±62.1)
São Francisco do Sul	UFRGS 10570/TEC106 G	-48.58916667	-26.29305556	2541149	2313342	18.7 (±53.7)
Cubatão Norte	MCP30667	-48.9289	-26.1919	1272942	994028	9.9 (±60.9)
Guaratuba	UFRGS 10578/TEC109	-48.5833	-25.8833	2158302	1857011	16.4 (±128.7)
Guaratuba	UFRGS 9358/TEC346	-48.5833	-25.8833	1267302	956980	8.8 (±74.1)
Paranaguá	MHNCI ncat	-48.7000	-25.4333	1278109	935893	9.2 (±33)
Paranaguá	MHNCI ncat	-48.7000	-25.4333	1021488	706960	8.1 (±21)
Paranaguá	MHNCI ncat	-48.7000	-25.4333	1235019	906103	9.4 (±46.4)
Paranaguá	MHNCI ncat	-48.7000	-25.4333	1200292	807300	8.2 (±79.1)
Paranaguá	MHNCI ncat	-48.7000	-25.4333	1339796	923487	7.8 (±134.4)
Guaraqueçaba	UFRGS 11778/TEC755 A	-48.4500	-25.2200	1365569	1097390	13.1 (±29.2)
Guaraqueçaba	UFRGS 11778/TEC755 B	-48.4500	-25.2200	1541716	1171014	9.1 (±19.6)
Guaraqueçaba*	UFRGS 11778/TEC755 C	-48.4500	-25.2200	713284	392307	-
Guaraqueçaba	UFRGS 11778/TEC755 D	-48.4500	-25.2200	1306469	977174	9.3 (±19.7)
Guaraqueçaba	UFRGS 11778/TEC755 E	-48.4500	-25.2200	1301434	985575	9.3 (±24.4)
Guaraqueçaba	MCP30558	-48.4336	-25.2092	1570277	1277232	12.9 (±34.4)
Guaraqueçaba	UFRGS 11779/TEC763 A	-48.4203	-25.1739	1331361	938170	8.9 (±19.6)
Guaraqueçaba	UFRGS 11779/TEC763 B	-48.4203	-25.1739	1343074	995952	8.5 (±17.3)
Guaraqueçaba	UFRGS 11779/TEC763 C	-48.4203	-25.1739	1833265	1494428	13.5 (±31.9)
Guaraqueçaba	UFRGS 11779/TEC763 D	-48.4203	-25.1739	1841856	1578082	14 (±36.1)
Ribeira de Iguape	UFRGS 11781/TEC806 A	-47.4931	-24.6614	2077778	1583210	11.9 (±26.5)
Ribeira de Iguape	UFRGS 11781/TEC806 B	-47.4931	-24.6614	1570785	1305293	12.8 (±30.2)
Ribeira de Iguape	UFRGS 11781/TEC806 C	-47.4931	-24.6614	1514447	1226833	12.6 (±30.3)
Ribeira de Iguape*	UFRGS 11781/TEC806 D	-47.4931	-24.6614	673196	449983	8.4 (±16.2)
Ribeira de Iguape	UFRGS 11782/TEC739 A	-47.4836	-24.6500	1644564	1264159	11.9 (±28)
Ribeira de Iguape	UFRGS 11782/TEC739 B	-47.4836	-24.6500	1572629	1234041	12 (±29)

Guarau*	UFRGS 11784/TEC722 A	-47.0550	-24.3683	4449	2478	-
Guarau	UFRGS 11784/TEC722 B	-47.0550	-24.3683	1556261	1114347	9 (±15.6)
Guarau	UFRGS 11784/TEC722 D	-47.0550	-24.3683	1530139	1145238	10.3 (±20.8)
Guarau	UFRGS 11784/TEC722 E	-47.0550	-24.3683	1598988	1256610	8.1 (±15.7)
Guarau	MCP30555	-47.0547	-24.3681	1291554	940643	8 (±12.7)
Guarau*	UFRGS 11783/TEC817 A	-47.0372	-24.3600	1289234	1028955	10.3 (±19.5)
Guarau	UFRGS 11783/TEC817 C	-47.0372	-24.3600	1219687	859228	7.7 (±14)
Guarau*	UFRGS 11783/TEC817 E	-47.0372	-24.3600	10481	2863	-
Guarau	MCP30554	-47.0372	-24.3597	1519051	1159490	9.8 (±91.4)
Guarau	MCP30561	-47.0361	-24.3586	1912939	1559830	14.9 (±34.9)
Santos	UFRGS 11785/TEC820 A	-46.3489	-23.8569	1428856	1050598	10 (±21.5)
Santos	UFRGS 11785/TEC820 B	-46.3489	-23.8569	1473939	1090577	10 (±20.6)
Santos*	UFRGS 11785/TEC820 C	-46.3489	-23.8569	2490488	2222585	20.2 (±79.2)
Santos	UFRGS 11785/TEC820 D	-46.3489	-23.8569	1379889	1059057	10.3 (±21.9)
Santos	UFRGS 11785/TEC820 E	-46.3489	-23.8569	1670347	1282315	11 (±22.5)
Santos	UFRGS 11785/TEC820 F	-46.3489	-23.8569	1606014	1219205	10.7 (±22.7)
Santos	MCP30559	-46.3483	-23.8564	1410664	1125864	10 (±19.4)
Santos	MCP30559	-46.3483	-23.8564	1569792	1233696	10.1 (±20.4)
Santos	MCP30560	-46.3281	-23.8425	1652764	1358598	12.4 (±28.7)
Santos	MCP30560	-46.3281	-23.8425	934064	760014	8.9 (±17)
Upper Tietê	UFRGS 11787/TEC827 A	-46.3136	-23.7725	1721145	1281748	8.9 (±16.8)
Upper Tietê*	UFRGS 11787/TEC827 B	-46.3136	-23.7725	2209345	2032282	16.4 (±52.7)
Upper Tietê	UFRGS 11787/TEC827 C	-46.3136	-23.7725	1340661	943311	7.7 (±15.6)
Upper Tietê*	UFRGS 11787/TEC827 D	-46.3136	-23.7725	2512031	2432574	32.1 (±84.9)
Upper Tietê	UFRGS 11787/TEC827 E	-46.3136	-23.7725	1651191	1233524	10.3 (±20.6)
Upper Tietê	UFRGS 11786/TEC826 A	-46.3347	-23.7664	804503	637740	8.4 (±15)
Upper Tietê	UFRGS 11786/TEC826 B	-46.3347	-23.7664	1239103	996612	9.4 (±21.9)
Upper Tietê	UFRGS 11786/TEC826 C	-46.3347	-23.7664	1312201	937231	8.9 (±17.3)
Upper Tietê	UFRGS 11786/TEC826 D	-46.3347	-23.7664	1816278	1383835	11.6 (±23)
Upper Tietê	UFRGS 11786/TEC826 E	-46.3347	-23.7664	1317731	939323	7.8 (±13.8)
Bertioga*	UFRGS 11771/TEC888 C	-46.0031	-23.7794	199120	186254	-
Bertioga	UFRGS 11771/TEC888 D	-46.0031	-23.7794	1702483	1340406	14.2 (±35.2)
Bertioga	UFRGS 11771/TEC888 E	-46.0031	-23.7794	3797857	3466754	32 (±109.1)
Bertioga	UFRGS 11771/TEC888 F	-46.0031	-23.7794	2530946	2115001	19.3 (±77.3)
Bertioga*	UFRGS 11771/TEC888 G	-46.0031	-23.7794	66882	6311	-
Bertioga	UFRGS 11777/TEC980 B	-46.0031	-23.7794	2935697	2550960	20.7 (±58.8)
Bertioga	UFRGS 11777/TEC980 C	-46.0031	-23.7794	1770554	1410510	15.4 (±40.3)
Bertioga	UFRGS 11777/TEC980 D	-46.0031	-23.7794	1856994	1445558	13.9 (±34.6)
Bertioga	UFRGS 11777/TEC980 E	-46.0031	-23.7794	2460277	1934988	16.8 (±44.1)
Bertioga*	UFRGS 11777/TEC980 G	-46.0031	-23.7794	23591	10754	-
Juquehy*	UFRGS 11772/TEC891 A	-45.7203	-23.7628	2300416	2131299	45.5 (±92.5)
Juquehy	UFRGS 11772/TEC891 B	-45.7203	-23.7628	2585949	2107327	19 (±53.2)
Juquehy*	UFRGS 11772/TEC891 C	-45.7203	-23.7628	224285	193786	-
Juquehy*	UFRGS 11772/TEC891 D	-45.7203	-23.7628	490259	421046	-
Juquehy*	UFRGS 11772/TEC891 E	-45.7203	-23.7628	21494	10027	-
Juquehy	UFRGS 11772/TEC891 G	-45.7203	-23.7628	3404274	3089411	25.3 (±83)
Juquehy	UFRGS 11772/TEC891 H	-45.7203	-23.7628	2675823	2351809	19.4 (±50.3)
Juquehy*	UFRGS 11772/TEC891 I	-45.7203	-23.7628	2901160	2660446	31 (±72.1)
Juquehy*	MCP30658	-45.7144	-23.7617	8339	4351	-
Boiçucanga	UFRGS 11795/TEC1229 A	-45.6064	-23.7733	1998679	1729076	18.7 (±48.6)
Boiçucanga	UFRGS 11795/TEC1229 B	-45.6064	-23.7733	2850510	2373213	20.3 (±63.3)
Boiçucanga	UFRGS 11795/TEC1229 C	-45.6064	-23.7733	3256867	2953768	23.7 (±78.2)
Boiçucanga	UFRGS 11795/TEC1229 D	-45.6064	-23.7733	2501502	2161569	19.3 (±66)

Boiçucanga	UFRGS 11795/TEC1229 E	-45.6064	-23.7733	3023378	2681668	22.1 (±67.5)
Boiçucanga*	UFRGS 11795/TEC1229 F	-45.6064	-23.7733	3416573	3185316	26.1 (±181)
Boiçucanga	UFRGS 11795/TEC1229 G	-45.6064	-23.7733	1479360	1176663	12.6 (±26.1)
Boiçucanga	UFRGS 11795/TEC1229 H	-45.6064	-23.7733	1109731	867293	10.5 (±20.7)
Boiçucanga*	UFRGS 11795/TEC1229 I	-45.6064	-23.7733	1909069	1689597	16.7 (±31.6)
Boiçucanga*	UFRGS 11795/TEC1229 J	-45.6064	-23.7733	2352458	2146528	26.4 (±61.7)
Sao Sebastiao*	UFRGS 11788/TEC859 A	-45.4522	-23.8239	1165919	603275	8.3 (±14)
Sao Sebastiao*	UFRGS 11788/TEC859 B	-45.4522	-23.8239	1201194	599398	6.6 (±9.6)
Sao Sebastiao*	MCP30660	-45.5519	-23.7894	1219213	864402	7.9 (±16.4)
Ilhabela	MCP30661	-45.3633	-23.8250	1318964	1074290	11.3 (±23.8)
Ilhabela	MCP30661	-45.3633	-23.8250	3161748	2944413	15.7 (±48.8)
Ilhabela*	UFRGS 11773/TEC897 C	-45.3536	-23.8214	190840	170846	-
Ilhabela	UFRGS 11773/TEC897 D	-45.3536	-23.8214	1095256	778180	11.4 (±22.6)
Ilhabela	UFRGS 11773/TEC897 E	-45.3536	-23.8214	1894348	1618334	12.4 (±27.4)
Ilhabela*	UFRGS 11773/TEC897 F	-45.3536	-23.8214	775095	504082	7.6 (±12.5)
Ilhabela*	UFRGS 11773/TEC897 G	-45.3536	-23.8214	604004	52725	-
Ilhabela	MCP30662	-45.3589	-23.8200	1880301	1595523	14.1 (±37.5)
Ilhabela	MCP30662	-45.3589	-23.8200	1732334	1455896	13.1 (±31.2)
Ubatuba	MCP30663	-45.1286	-23.4278	1480671	1083984	9.6 (±18.9)
Ubatuba	UFRGS 11774/TEC900 A	-45.1144	-23.4125	1321716	935442	10.3 (±20.2)
Ubatuba	UFRGS 11774/TEC900 B	-45.1144	-23.4125	2151002	1915290	15.5 (±38.9)
Picinguaba	UFRGS 11789/TEC866 A	-44.8706	-23.3467	2429300	1984838	15.6 (±44.2)
Picinguaba	UFRGS 11789/TEC866 B	-44.8706	-23.3467	1074156	793710	10.2 (±19.3)
Toca do Boi	UFRGS 11775/TEC908 A	-44.6819	-23.3292	1453427	1185664	13.4 (±36.7)
Toca do Boi	UFRGS 11775/TEC908 B	-44.6819	-23.3292	1574027	1333708	13.4 (±31)
Toca do Boi	UFRGS 11775/TEC908 C	-44.6819	-23.3292	1666329	1443786	15.7 (±42.8)
Toca do Boi	UFRGS 11775/TEC908 D	-44.6819	-23.3292	1132355	882935	11 (±26)
Toca do Boi	UFRGS 11775/TEC908 E	-44.6819	-23.3292	1682263	1366221	13.4 (±33)
Toca do Boi	UFRGS 11775/TEC908 F	-44.6819	-23.3292	1361996	1107613	12.1 (±28.3)
Toca do Boi	UFRGS 11775/TEC908 G	-44.6819	-23.3292	1434266	1118482	11.6 (±24.2)
Toca do Boi	UFRGS 11775/TEC908 H	-44.6819	-23.3292	2040740	1742324	13.7 (±44.1)
Toca do Boi	MCP30664	-44.6819	-23.3292	1484778	1165930	12 (±29.6)
Toca do Boi	MCP30664	-44.6819	-23.3292	1572158	1206032	11 (±28.9)
Tarituba	UFRGS 11776/TEC909 A	-44.5950	-23.0419	1098168	900046	10.5 (±23.9)
Tarituba*	UFRGS 11776/TEC909 B	-44.5950	-23.0419	726529	618667	10.1 (±19.7)
Tarituba	UFRGS 11776/TEC909 C	-44.5950	-23.0419	2868700	2542999	20.6 (±60.5)
Tarituba*	UFRGS 11776/TEC909 D	-44.5950	-23.0419	620526	449556	7.3 (±13.7)
Tarituba	UFRGS 11776/TEC909 E	-44.5950	-23.0419	1310075	1110254	12.2 (±28.6)
Tarituba	UFRGS 11776/TEC909 F	-44.5950	-23.0419	2099253	1827632	16.2 (±39.9)
Tarituba	UFRGS 11776/TEC909 G	-44.5950	-23.0419	1389264	1078034	11.6 (±27.6)
Tarituba	UFRGS 11776/TEC909 H	-44.5950	-23.0419	1464882	1206051	13.4 (±36.9)
Tarituba*	UFRGS 11776/TEC909 I	-44.5950	-23.0419	2409686	2221319	27.8 (±69.3)
Tarituba	UFRGS 11776/TEC909 J	-44.5950	-23.0419	2496914	2229227	19.3 (±51.2)
Paraty*	MCP30665	-44.6975	-23.0764	970498	703921	8.7 (±17.1)

(d) *Bryconamericus*

Population	Voucher	Longitude	Latitude	Raw read count	Retained reads	Mean coverage (SD)
Três Forquilhas	UFRGS 16209/TEC 2357	-50.09161111	-29.509	1279122	1239762	23.3 (±52.9)
Três Forquilhas	UFRGS 16209/TEC 2358	-50.09161111	-29.509	2119729	2054087	33.4 (±78.3)
Três Forquilhas	UFRGS 16209/TEC 2359	-50.09161111	-29.509	2928005	2840748	41.6 (±113.4)
Três Forquilhas*	UFRGS 16209/TEC 2360	-50.09161111	-29.509	40524	38406	11.5 (±19.2)
Araranguá	UFRGS 16211/TEC 2376	-49.67361111	-28.97875	2918015	2846728	41.8 (±109.7)

Araranguá	UFRGS 16211/TEC 2377	-49.67361111	-28.97875	2280681	2213314	34 (±85.9)
Araranguá	UFRGS 16211/TEC 2378	-49.67361111	-28.97875	3248149	3171295	45.3 (±113.5)
Araranguá	UFRGS 16211/TEC 2379	-49.67361111	-28.97875	2706494	2633775	39.7 (±105)
Itajaí	UFRGS 18476/TEC 3884A	-49.161194	-27.224417	392158	379261	11 (±21)
Itajaí	UFRGS 18476/TEC 3884B	-49.161194	-27.224417	466689	452890	10.2 (±19.1)
Itajaí	UFRGS 18476/TEC 3884C	-49.161194	-27.224417	802105	770570	18 (±35.2)
Itajai	UFRGS 20461/TEC 5719	-48.7225	-27.015	1292849	1261528	26.8 (±52.2)
Itajaí*	UFRGS 18600/TEC 3893A	-48.714444	-27.006944	56007	52150	10 (±16.7)
Itajaí*	UFRGS 18600/TEC 3893B	-48.714444	-27.006944	34129	31681	14.2 (±20.5)
Itajaí	UFRGS 18600/TEC 3893C	-48.714444	-27.006944	1095984	1064556	21.4 (±44.4)
Itajai	UFRGS 20463/TEC 5721	-48.712556	-26.994972	1988499	1934159	33.1 (±72.3)
Itapocu	UFRGS 18485/TEC 3923	-49.178333	-26.419444	2433769	2369823	39.8 (±118.7)
Itapocu	UFRGS 18485/TEC 4670	-49.178333	-26.419444	2158778	2105481	28.9 (±82.9)
Itapocu	UFRGS 18485/TEC 4671	-49.178333	-26.419444	1599110	1558982	32.4 (±85)
Itapocu	UFRGS 18485/TEC 4672	-49.178333	-26.419444	712644	691911	17.6 (±37.9)
Itapocu*	UFRGS 18485/TEC 4673	-49.178333	-26.419444	1157	109	-
Itapocu	UFRGS 18485	-49.178333	-26.419444	2718259	2652589	45.1 (±129.5)
Itapocu	UFRGS 18485	-49.178333	-26.419444	2246814	2191791	40.8 (±115.3)
Itapocu	UFRGS 18485	-49.178333	-26.419444	2991530	2917900	45.3 (±143.8)
Cubatão Norte*	UFRGS 20467/TEC 5725A	-48.922222	-26.197111	79835	53398	8.9 (±9.8)
Cubatão Norte	UFRGS 20467/TEC 5725B	-48.922222	-26.197111	754078	732134	18.6 (±36.2)
Cubatão Norte	UFRGS 20467/TEC 5725C	-48.922222	-26.197111	1476060	1430276	28.5 (±74.7)
Cubatão Norte	UFRGS 18631/TEC 3937A	-48.954083	-26.176611	1447902	1398341	28.2 (±61)
Cubatão Norte	UFRGS 18631/TEC 3937B	-48.954083	-26.176611	2379986	2309263	38.8 (±88.6)
Cubatão Norte	UFRGS 18511/TEC 3931	-48.999764	-26.175833	1826251	1767902	32.4 (±76.4)
Cubatão Norte	UFRGS 18511/TEC 4675	-48.999764	-26.175833	920988	888339	20 (±41.3)
Cubatão Norte*	UFRGS 18511/TEC 4676	-48.999764	-26.175833	117985	110651	8.7 (±15.4)
Guaratuba	UFRGS 18651/TEC 3957A	-48.856111	-26.007778	1856046	1794705	34.6 (±130.1)
Guaratuba	UFRGS 18651/TEC 3957B	-48.856111	-26.007778	1569911	1516407	30.7 (±80.8)
Guaratuba	UFRGS 18651/TEC 3957C	-48.856111	-26.007778	1444455	1387134	27.9 (±78.7)
Guaratuba*	UFRGS 18651/TEC 3957D	-48.856111	-26.007778	3952	1701	-
Guaratuba	UFRGS 18651/TEC 3957E	-48.856111	-26.007778	1294115	1255443	27.2 (±63.3)
Guaratuba	UFRGS 18651/TEC 3957F	-48.856111	-26.007778	2859614	2766823	43.1 (±142.1)
Guaratuba	UFRGS 18651/TEC 3957G	-48.856111	-26.007778	879030	843185	20.3 (±48.2)
Guaratuba*	UFRGS 18651/TEC 3957H	-48.856111	-26.007778	3584	281	-
Paranaguá	UFRGS 18510/TEC 3977A	-48.834722	-25.458056	1243022	1205489	26.2 (±64.6)
Paranaguá	UFRGS 18510/TEC 3977B	-48.834722	-25.458056	2401229	2330032	39.4 (±78.8)
Paranaguá	UFRGS 18510/TEC 3977E	-48.834722	-25.458056	662219	640463	16.6 (±36.9)
Paranaguá*	UFRGS 18510/TEC 3977F	-48.834722	-25.458056	5225	113	-
Paranaguá*	UFRGS 18510/TEC 3977G	-48.834722	-25.458056	4029	84	-
Paranaguá	UFRGS 18510/TEC 3977H	-48.834722	-25.458056	403240	388617	12.3 (±27.2)
Ribeira de Iguape	UFRGS 12417/TEC 778	-48.551389	-25.072222	1160027	1092998	18.6 (±89.1)
Ribeira de Iguape	UFRGS 18536/TEC 4319A	-48.293333	-24.643056	1594720	1558011	29.6 (±71.8)
Ribeira de Iguape	UFRGS 18536/TEC 4319B	-48.293333	-24.643056	1362004	1321566	26.3 (±56.1)
Ribeira de Iguape	UFRGS 18517/TEC 4304A	-48.591389	-24.580000	1489355	1453023	28.6 (±56.7)
Ribeira de Iguape	UFRGS 18517/TEC 4309B	-48.591389	-24.580000	1352136	1323030	26.1 (±54.6)
Ribeira de Iguape	UFRGS 18515/TEC 4683	-48.673056	-24.558333	2904870	2844880	44.4 (±96.2)
Ribeira de Iguape	UFRGS 18515/TEC 4684	-48.673056	-24.558333	2866453	2800335	43.5 (±89.5)
Ribeira de Iguape*	UFRGS 18524/TEC 4346A	-47.653056	-24.104167	322268	311254	11.7 (±25.8)
Ribeira de Iguape*	UFRGS 18524/TEC 4346B	-47.653056	-24.104167	304445	295462	9.8 (±17.3)
Ribeira de Iguape	UFRGS 18524/TEC 4346C	-47.653056	-24.104167	2724659	2655741	41.6 (±83)
Ribeira de Iguape	UFRGS 18524/TEC 4346D	-47.653056	-24.104167	2057572	2011391	33.7 (±78.3)
Perequê-açu	UFRGS 18789/TEC 4429A	-44.788611	-23.215278	1790684	1737918	32.7 (±89.2)

Perequê-açu	UFRGS 18789/TEC 4429B	-44.788611	-23.215278	720890	699474	17.4 (±32.6)
Perequê-açu*	UFRGS 18789/TEC 4429C	-44.788611	-23.215278	257228	246347	8.4 (±16.3)
Perequê-açu	UFRGS 18789/TEC 4429D	-44.788611	-23.215278	1592620	1546938	30 (±64.4)
Perequê-açu	UFRGS 18789/TEC 4429E	-44.788611	-23.215278	2387374	2325328	40.2 (±88.7)
Perequê-açu	UFRGS 18789/TEC 4429F	-44.788611	-23.215278	1343369	1307901	28.4 (±73)
Perequê-açu	UFRGS 18789/TEC 4429G	-44.788611	-23.215278	216700	208781	8.5 (±14.3)
Perequê-açu*	UFRGS 18789/TEC 4429H	-44.788611	-23.215278	8562	269	-
Taquari	UFRGS 18791/TEC 4431A	-44.692500	-23.041389	1386365	1356730	29.6 (±66.9)
Taquari	UFRGS 18791/TEC 4431B	-44.692500	-23.041389	1849566	1805769	32.8 (±64.1)
Taquari	UFRGS 18791/TEC 4431C	-44.692500	-23.041389	2272186	2180744	36.3 (±72.5)
Taquari	UFRGS 18791/TEC 4431D	-44.692500	-23.041389	2693680	2622928	42.8 (±103.2)
Taquari	UFRGS 18791/TEC 4431E	-44.692500	-23.041389	2861613	2783414	44.6 (±89.7)
Taquari	UFRGS 18791/TEC 4431F	-44.692500	-23.041389	1809723	1765553	19.5 (±37.7)
Taquari	UFRGS 18791/TEC 4431G	-44.692500	-23.041389	1493358	1449232	27.4 (±49.9)
Taquari	UFRGS 18791/TEC 4431H	-44.692500	-23.041389	1904893	1832935	30.5 (±63.2)
Guapimirim	UFRGS 18848/TEC 4478A	-42.898611	-22.517778	1463128	1412646	32 (±131)
Guapimirim	UFRGS 18848/TEC 4478B	-42.898611	-22.517778	1749975	1686892	33.8 (±152.4)
Guapimirim	UFRGS 18848/TEC 4478C	-42.898611	-22.517778	2067482	1997891	32.3 (±89.1)
Macacu	UFRGS 18843/TEC 4473	-42.657222	-22.478889	1880337	1802560	32.4 (±89.5)
Macacu*	UFRGS 18834/TEC 4465	-42.648500	-22.451056	136370	125997	32.5 (±89.9)
Macacu	UFRGS 18834/TEC 4747	-42.648500	-22.451056	2405931	2350649	32.5 (±90.3)
Macacu	UFRGS 18834/TEC 4749	-42.648500	-22.451056	2913533	2846707	32.6 (±90.6)
Macacu	UFRGS 18828/TEC 4459	-42.621667	-22.418889	1101578	1060097	32.7 (±91)
Macacu	UFRGS 18828/TEC 4744	-42.621667	-22.418889	1649697	1608699	32.8 (±91.4)
Macacu	UFRGS 18828/TEC 4746	-42.621667	-22.418889	527327	514733	32.9 (±91.8)
São João	UFRGS 18873/TEC 4500	-42.574722	-22.582778	1860924	1808838	34.9 (±93.9)
São João	UFRGS 18892/TEC 4516	-42.465278	-22.545833	3132383	3062687	50.9 (±189.1)
São João	UFRGS 18892/TEC 4754	-42.465278	-22.545833	2124948	2070053	39.7 (±119)
São João	UFRGS 18892/TEC 4756	-42.465278	-22.545833	3458041	3353169	47.4 (±174.3)

Table D.2. Per species libraries and STACKS processing information, with number of individuals and loci in each step of data processing.

	Mimagoniates microlepis	*Hyphessobrycon boulengeri*	*Hollandichthys*	*Bryconamericus*
Total of libraries	3	2	2	1
Total individuals	118*	136	182	87
Total reads (millions)	203625424	241441086	325578791	164516577
Retained reads after *proccess radtags*	198540198	203388694	232013863	136913642
Mean coverage	44.2 (±13.5)	37.1 (±13.6)	13.8 (±5.8)	28.4 (±11.7)
Number base pairs eliminated at 3' end	15	20	5	15
Final sequence length	120	115	80	120
Θ upper 95% quantile	0.02	0.0145	0.024	0.018
Number of individuals after STACKS	116	136	167	83
Number of loci after STACKS	182,799	90,487	204,924	126,378
Number of SNPs after STACKS	508,333	209,853	568,347	379,422
Maximum number of SNP's in one locus	13	9	10	11
Number of individuals used for analyses	113**	134	142	74
Number of populations	32	23	24	13
Number of loci/SNPs with 50% missing data	3,515/30,032	13,881/57,183	26,077/122,762	11,436/75,552
Genotyping rate 50% missing	0.67	0.72	0.66	0.68
Number of loci/SNPs with 25% missing data	1,802/15,896	6,129/27,103	6,902/33,780	4,293/30,686
Genotyping rate 25% missing	0.79	0.87	0.85	0.92

*118 are a subset of 240 individuals that were sequenced in three libraries for this species as part of a larger project.

**35% missing for *M. microlepis*

Table D.3. Species present at each sampled population. Populations are ordered from the southern most (1 - Patos) to the northern most population (47 - Caravelas).

	Population	# sp.	*M. microlepis*	*H. boulengeri*	*Hollandichthys*	*Bryconamericus*
1	Patos	1		X		
2	Tramandaí	1		X		
3	Maquiné	2	X		X	
4	Três Forquilhas	4	X	X	X	X
5	Mampituba	2	X		X	
6	Araranguá	3	X		X	X
7	D'Una	3	X	X	X	
8	Cubatão Sul	2	X	X		
9	Florianópolis	2		X	X	
10	Biguaçu	1	X			
11	Tijucas	1	X			
12	Itajaí	3	X	X		X
13	Itapocu	3		X	X	X
14	São Francisco do Sul	3	X	X	X	
15	Cubatão Norte	2			X	X
16	Saí-Mirim	2	X	X		
17	Saí-Guaçu	1		X		
18	Guaratuba	4	X	X	X	X
19	Guaraguaçu	1	X			
20	Iguaçu	2	X	X		
21	Paranapanema	1	X			
22	Paranaguá	4	X	X	X	X
23	Guaraqueçaba	2	X		X	
24	Ribeira de Iguape	4	X	X	X	X
25	Guarau	2	X		X	
26	Itanhaém	2	X	X		
27	Santos	2	X		X	
28	Upper Tietê	2		X	X	
29	Bertioga	3	X	X	X	
30	Juquehy	1			X	
31	Boiçucanga	1			X	
32	Ilhabela	1			X	
33	Ubatuba	1	X		X	
34	Picinguaba	1	X		X	
35	Toca do Boi	1			X	
36	Pereque-Açu	1				X
37	Taquari	1				X
38	Tarituba	1			X	
39	Jaceruba	1	X			
40	Guapimirim	3	X	X		X
41	Macacu	2	X			X
42	São João	1				X
43	Paraíba do Sul	1	X			
44	Barra do Riacho	2	X	X		
45	Barra Seca	2	X	X		
46	Itaunas	2	X	X		
47	Caravelas	2	X	X		

Table D.4. Population genetic summary statistics for each species per population: (a) *M. microlepis*, (b) *H. boulengeri*, (c) *Hollandichthys* and (d) *Bryconamericus*. Populations are listed in order of the southern most populations (1 - Patos) to the northern most population (47 - Caravelas)(see Figure 1). The number of individuals retained per population after processing of genomic data for the 25% missing data (N_{25}), which were used to calculate the reported summary statistics are given. Summary statistics are presented only for polymorphic sites, and include average nucleotide diversity, π, average observed heterozygosity per locus, H_{OBS}, and average expected heterozygosity per locus, H_{EXP}. See Table S1 for complete list of samples and summary of genomic data collected for each individual.

(a) *M. microlepis*

Pop. Number	Population	N_{25}	Poly Sites	% Poly Loci	Private	π	H_{OBS}	H_{EXP}	*Fis*
3	Maquiné	5	463	0.1925	85	0.013	0.011	0.011	0.003
4	Três Forquilhas	5	496	0.2034	104	0.011	0.010	0.010	0.002
5	Mampituba	5	430	0.1794	159	0.010	0.009	0.009	0.002
6	Araranguá	4	544	0.2222	256	0.016	0.010	0.014	0.012
7	D'Una	4	415	0.1727	269	0.013	0.011	0.010	0.003
8	Cubatão Sul	4	394	0.1665	219	0.011	0.011	0.010	0.001
10	Biguaçu	4	476	0.198	240	0.013	0.012	0.011	0.003
11	Tijucas	4	516	0.2142	163	0.015	0.012	0.013	0.005
12	Itajaí	3	1286	0.5172	404	0.041	0.021	0.033	0.035
14	São Francisco do Sul	3	828	0.3427	288	0.030	0.024	0.022	0.010
16	Saí-Mirim	2	858	0.353	277	0.033	0.030	0.024	0.006
18	Guaratuba	4	1011	0.4119	255	0.030	0.021	0.025	0.018
17	Guaraguaçu	2	919	0.3753	217	0.035	0.030	0.025	0.008
20	Iguaçu	2	467	0.1983	40	0.022	0.020	0.014	0.003
21	Paranapanema	3	843	0.3503	108	0.027	0.022	0.022	0.008
22	Paranaguá	3	1265	0.529	465	0.040	0.031	0.032	0.017
23	Guaraqueçaba	4	1232	0.5152	464	0.038	0.028	0.031	0.018
24	Ribeira de Iguape	8	2339	0.9657	1319	0.047	0.028	0.042	0.046
25	Guarau	2	433	0.1934	93	0.018	0.017	0.013	0.001
26	Itanhaém	3	1100	0.4584	240	0.037	0.027	0.029	0.018
27	Santos	4	1107	0.4587	182	0.036	0.015	0.030	0.037
29	Bertioga	4	1431	0.6023	364	0.041	0.029	0.034	0.022
33	Ubatuba	4	948	0.3995	276	0.030	0.022	0.025	0.015
32	Picinguaba	3	986	0.417	272	0.035	0.022	0.027	0.021
39	Jaceruba*	1	417	0.2024	180	0.032	0.032	0.016	0.000
40	Guapimirim*	1	408	0.1948	156	0.031	0.031	0.016	0.000
41	Macacu	2	924	0.4191	368	0.040	0.034	0.028	0.009
43	Paraíba do Sul	4	828	0.358	380	0.026	0.021	0.022	0.009
44	Barra do Riacho	4	577	0.271	238	0.018	0.015	0.015	0.006
45	Barra Seca	4	554	0.25	208	0.018	0.015	0.015	0.005
46	Itaunas	4	647	0.2901	259	0.019	0.016	0.016	0.006
47	Caravelas	4	795	0.3652	469	0.024	0.021	0.021	0.006

(b) *H. boulengeri*

Pop. Number	Population	N_{25}	Poly Sites	% Poly Loci	Private	π	H_{OBS}	H_{EXP}	*Fis*
1	Patos	13	3292	0.3852	872	0.032	0.024	0.030	0.022
2	Tramandaí	7	2272	0.2663	267	0.030	0.025	0.027	0.010
4	Três Forquilhas	3	1382	0.1635	101	0.024	0.023	0.020	0.003
7	D'Una*	1	239	0.0332	289	0.011	0.011	0.005	0.000
8	Cubatão Sul*	1	223	0.0344	80	0.011	0.011	0.006	0.000
9	Florianópolis	8	411	0.0497	156	0.006	0.006	0.005	0.000
12	Itajaí*	1	867	0.1163	115	0.037	0.037	0.019	0.000
13	Itapocu	10	5829	0.6802	1311	0.063	0.053	0.060	0.027
14	São Francisco do Sul	8	2900	0.3393	263	0.040	0.035	0.037	0.011
16	Saí-Mirim	8	2097	0.2455	45	0.024	0.022	0.022	0.005
17	Saí-Guaçu	8	1199	0.1411	25	0.017	0.015	0.015	0.003
18	Guaratuba	3	1236	0.1457	17	0.022	0.018	0.018	0.008
20	Iguaçu	7	3389	0.3955	330	0.047	0.042	0.043	0.011
22	Paranaguá	3	1362	0.1653	65	0.026	0.023	0.021	0.004
24	Ribeira de Iguape	9	4103	0.4901	2923	0.044	0.037	0.041	0.017
26	Itanhaém	9	2837	0.3364	615	0.038	0.033	0.035	0.012
28	Upper Tietê	8	1413	0.1705	72	0.020	0.018	0.018	0.002
29	Bertioga	8	2817	0.3342	644	0.038	0.034	0.035	0.009
40	Guapimirim	3	2474	0.3475	3095	0.052	0.047	0.042	0.009
44	Barra do Riacho	7	1608	0.2743	600	0.024	0.022	0.022	0.004
45	Barra Seca	6	1626	0.2799	619	0.026	0.023	0.024	0.007
46	Itaunas	2	587	0.1028	184	0.018	0.017	0.013	0.001
47	Caravelas*	1	222	0.0406	135	0.013	0.013	0.006	0.000

(c) *Hollandichthys*

Pop. Number	Population	N_{25}	Poly Sites	% Poly Loci	Private	π	H_{OBS}	H_{EXP}	*Fis*
3	Maquiné	10	641	0.1103	170	0.006	0.006	0.006	0.001
4	Três Forquilhas	6	964	0.1703	204	0.013	0.012	0.011	0.000
5	Mampituba	10	1566	0.2681	531	0.015	0.013	0.014	0.007
6	Araranguá	3	854	0.1505	496	0.013	0.013	0.010	0.001
7	D'Una	5	958	0.1681	513	0.013	0.012	0.011	0.002
9	Florianópolis	8	4990	0.8544	1488	0.055	0.024	0.051	0.071
13	Itapocu*	1	1196	0.2361	214	0.043	0.043	0.021	0.000
14	São Francisco do Sul	8	5700	0.9747	1632	0.058	0.047	0.052	0.024
15	Cubatão Norte*	1	1194	0.2371	211	0.043	0.043	0.021	0.000
18	Guaratuba	2	2610	0.4538	580	0.048	0.045	0.034	0.004
22	Paranaguá	5	3844	0.6649	1165	0.049	0.038	0.042	0.022
23	Guaraqueçaba	9	4811	0.8232	1886	0.050	0.040	0.047	0.026
24	Ribeira de Iguape	5	5460	0.9353	2056	0.063	0.056	0.055	0.014
25	Guarau	7	1806	0.3114	666	0.021	0.018	0.019	0.008
27	Santos	9	4007	0.6848	847	0.043	0.039	0.040	0.009
28	Upper Tietê	8	2827	0.4838	413	0.032	0.030	0.029	0.004
29	Bertioga	7	4263	0.7289	1147	0.045	0.041	0.041	0.009
30	Juquehy	3	1605	0.2793	309	0.025	0.024	0.020	0.002
31	Boiçucanga	7	2748	0.4708	579	0.032	0.028	0.029	0.007
32	Ilhabela	6	1812	0.3108	638	0.021	0.019	0.019	0.006
33	Ubatuba	3	1992	0.3422	580	0.030	0.028	0.024	0.003
34	Picinguaba	2	1394	0.2464	400	0.028	0.026	0.019	0.003
35	Toca do Boi	10	1587	0.2724	676	0.017	0.016	0.016	0.002
38	Tarituba	7	976	0.1682	555	0.010	0.010	0.009	0.001

(d) *Bryconamericus*

Pop. Number	Population	N_{25}	Poly Sites	% Poly Loci	Private	π	H_{OBS}	H_{EXP}	*Fis*
4	Três Forquilhas	3	2559	0.4472	1233	0.044	0.036	0.036	0.014
6	Araranguá	4	2671	0.4706	1316	0.040	0.034	0.035	0.013
12	Itajaí	6	1692	0.3024	1050	0.026	0.024	0.023	0.004
13	Itapocu	7	4987	0.8688	1324	0.056	0.020	0.052	0.128
15	Cubatão Norte	6	2458	0.4286	1328	0.029	0.026	0.026	0.007
18	Guaratuba	6	1594	0.2784	674	0.021	0.019	0.019	0.003
22	Paranaguá	4	4957	0.8649	1785	0.086	0.023	0.073	0.133
24	Ribeira de Iguape	11	7790	1.3568	3655	0.082	0.027	0.077	0.180
36	Pereque-Açu	6	697	0.1229	312	0.011	0.011	0.010	0.001
37	Taquari	8	1529	0.2668	632	0.022	0.020	0.020	0.004
40	Guapimirim	3	570	0.1095	1113	0.012	0.012	0.010	0.000
41	Macacu	6	821	0.1442	483	0.013	0.012	0.012	0.002
42	São João	4	1224	0.2317	1738	0.020	0.019	0.017	0.001

Table D.5. Pairwise F_{ST}-values between populations per species: (a) *M. microlepis*, (B) *H. boulengeri*, (c) *Hollandichthys* and (d) *Bryconamericus*. Bold numbers mark statistically significant values at an $\alpha = 0.05$ (below the diagonal) and at an α after correcting for multiple comparisons using a Bonferroni correction (above the diagonal; α = 0.0001, 0.0003, 0.0002 and 0.0006, respectively).

(a) *M. microlepis*

	Maquiné	Três Forquilhas	Mampituba	Araranguá	D'Una	Cubatão Sul	Biguaçu	Tijucas	Itajaí	Sat-Mirim	São Francisco do Sul	Guaratuba	Iguaçu	Paranapanema	Guaraguaçu	Paranaguá	Guaraqueçaba	Ribeira de Iguape	Guarau	Itanhaém	Santos	Bertioga	Ubatuba	Picinguaba	Macacu	Paraíba do Sul	Barra do Riacho	Barra Seca	Itaunas	Caravelas
Maquiné	-	0.3	0.35	0.37	0.71	0.73	0.72	0.72	0.53	0.66	0.65	0.66	0.72	0.68	0.64	0.75	0.76	0.7	0.86	0.78	0.76	0.75	0.8	0.79	0.8	0.82	0.86	0.87	0.86	0.84
Três Forquilhas	**0.3**	-	0.1	0.52	0.72	0.76	0.75	0.74	0.59	0.7	0.67	0.69	0.74	0.71	0.69	0.78	0.78	0.72	0.87	0.8	0.78	0.78	0.82	0.8	0.82	0.84	0.88	0.88	0.88	0.86
Mampituba	**0.35**	**0.1**	-	0.54	0.73	0.77	0.75	0.74	0.59	0.7	0.67	0.69	0.74	0.72	0.69	0.78	0.78	0.72	0.87	0.8	0.78	0.78	0.82	0.8	0.82	0.84	0.88	0.88	0.88	0.86
Araranguá	**0.37**	**0.52**	**0.54**	-	0.67	0.7	0.68	0.68	0.48	0.61	0.59	0.62	0.67	0.64	0.6	0.72	0.73	0.67	0.83	0.75	0.73	0.72	0.78	0.75	0.76	0.8	0.85	0.85	0.84	0.82
D'Una	**0.71**	**0.72**	**0.73**	**0.67**	-	0.69	0.67	0.68	0.47	0.62	0.61	0.64	0.7	0.65	0.6	0.72	0.75	0.67	0.85	0.76	0.73	0.72	0.79	0.76	0.78	0.81	0.86	0.87	0.86	0.83
Cubatão Sul	**0.73**	**0.76**	**0.77**	**0.7**	**0.69**	-	0.6	0.67	0.52	0.65	0.63	0.66	0.72	0.68	0.64	0.76	0.77	0.71	0.86	0.78	0.76	0.76	0.8	0.78	0.8	0.82	0.87	0.87	0.87	0.85
Biguaçu	**0.72**	**0.75**	**0.75**	**0.68**	**0.67**	**0.6**	-	0.66	0.51	0.64	0.61	0.65	0.7	0.67	0.63	0.75	0.76	0.71	0.85	0.78	0.75	0.76	0.8	0.78	0.8	0.82	0.87	0.87	0.86	0.85
Tijucas	**0.72**	**0.74**	**0.74**	**0.68**	**0.68**	**0.67**	**0.66**	-	0.34	0.59	0.56	0.61	0.66	0.64	0.59	0.73	0.74	0.68	0.84	0.76	0.73	0.73	0.78	0.76	0.78	0.81	0.86	0.86	0.85	0.83
Itajaí	**0.53**	**0.59**	**0.59**	**0.48**	**0.47**	**0.52**	**0.51**	0.34	-	0.23	0.19	0.34	0.32	0.37	0.28	0.59	0.6	0.59	0.69	0.62	0.61	0.61	0.66	0.62	0.62	0.7	0.76	0.74	0.76	0.73
Saí-Mirim	**0.66**	0.7	0.7	0.61	0.62	0.65	0.64	0.59	0.23	-	0.12	0.37	0.37	0.39	0.27	0.62	0.64	0.61	0.74	0.65	0.64	0.64	0.7	0.66	0.65	0.73	0.8	0.8	0.79	0.77
São Francisco do Sul	**0.65**	**0.67**	**0.67**	**0.59**	**0.61**	**0.63**	**0.61**	**0.56**	0.19	0.12	-	0.39	0.39	0.36	0.26	0.6	0.63	0.58	0.74	0.64	0.62	0.61	0.69	0.65	0.65	0.72	0.78	0.79	0.78	0.75
Guaratuba	**0.66**	**0.69**	**0.69**	**0.62**	**0.64**	**0.66**	**0.65**	**0.61**	**0.34**	0.37	**0.39**	-	0.33	0.31	0.2	0.63	0.65	0.61	0.76	0.68	0.65	0.65	0.7	0.67	0.68	0.73	0.79	0.79	0.79	0.77
Iguaçu	**0.72**	**0.74**	**0.74**	0.67	0.7	0.72	0.7	0.66	0.32	0.37	0.39	0.33	-	0.01	0.15	0.61	0.66	0.59	0.78	0.67	0.64	0.63	0.72	0.67	0.67	0.74	0.81	0.83	0.81	0.78
Paranapanema	**0.68**	**0.71**	**0.72**	**0.64**	**0.65**	**0.68**	**0.67**	**0.64**	0.37	0.39	0.36	**0.31**	0.01	-	0.2	0.65	0.66	0.63	0.77	0.69	0.67	0.67	0.72	0.69	0.7	0.75	0.81	0.81	0.81	0.78
Guaraguaçu	0.64	**0.69**	**0.69**	0.6	0.6	0.64	0.63	0.59	0.28	0.27	0.26	0.2	0.15	0.2	-	0.59	0.61	0.59	0.72	0.63	0.62	0.63	0.68	0.64	0.63	0.72	0.78	0.78	0.78	0.76
Paranaguá	**0.75**	**0.78**	**0.78**	**0.72**	**0.72**	**0.76**	**0.75**	**0.73**	0.59	0.62	0.6	**0.63**	0.61	0.65	0.59	-	0.11	0.23	0.56	0.45	0.46	0.48	0.54	0.51	0.51	0.6	0.7	0.67	0.69	0.68
Guaraqueçaba	**0.76**	**0.78**	**0.78**	**0.73**	**0.75**	**0.77**	**0.76**	**0.74**	**0.6**	0.64	**0.63**	**0.65**	0.66	**0.66**	0.61	**0.11**	-	0.26	0.6	0.49	0.5	0.5	0.58	0.54	0.55	0.62	0.71	0.7	0.7	0.68
Ribeira de Iguape	**0.7**	**0.72**	**0.72**	**0.67**	**0.67**	**0.71**	**0.71**	**0.68**	**0.59**	**0.61**	**0.58**	**0.61**	**0.59**	**0.63**	**0.59**	**0.23**	**0.26**	-	0.49	0.41	0.44	0.46	0.5	0.48	0.49	0.56	0.63	0.6	0.63	0.62
Guarau	**0.86**	**0.87**	**0.87**	0.83	0.85	0.86	0.85	0.84	0.69	0.74	0.74	0.76	0.78	0.77	0.72	0.56	0.6	**0.49**	-	0.28	0.37	0.39	0.54	0.47	0.59	0.67	0.78	0.79	0.77	0.75
Itanhaém	**0.78**	**0.8**	**0.8**	**0.75**	**0.76**	**0.78**	**0.78**	**0.76**	0.62	0.65	0.64	**0.68**	0.67	0.69	0.63	0.45	**0.49**	**0.41**	0.28	-	0.16	0.25	0.37	0.33	0.46	0.56	0.69	0.68	0.68	0.66
Santos	**0.76**	**0.78**	**0.78**	**0.73**	**0.73**	**0.76**	**0.75**	**0.73**	**0.61**	0.64	**0.62**	**0.65**	0.64	**0.67**	0.62	**0.46**	**0.5**	**0.44**	0.37	0.16	-	0.17	0.32	0.27	0.43	0.53	0.65	0.64	0.65	0.63
Bertioga	**0.75**	**0.78**	**0.78**	0.72	0.72	**0.76**	**0.76**	**0.73**	**0.61**	0.64	**0.61**	**0.65**	0.63	**0.67**	0.63	**0.48**	**0.5**	**0.46**	0.39	**0.25**	0.17	-	0.25	0.22	0.4	0.5	0.66	0.63	0.66	0.64
Ubatuba	**0.8**	**0.82**	**0.82**	**0.78**	**0.79**	**0.8**	**0.8**	**0.78**	**0.66**	0.7	**0.69**	**0.7**	0.72	**0.72**	0.68	**0.54**	**0.58**	**0.5**	0.54	**0.37**	**0.32**	**0.25**	-	0.3	0.48	0.56	0.71	0.71	0.71	0.68
Picinguaba	**0.79**	**0.8**	**0.8**	**0.75**	**0.76**	**0.78**	**0.78**	**0.76**	0.62	0.66	0.65	**0.67**	0.67	0.69	0.64	0.51	**0.54**	**0.48**	0.47	0.33	**0.27**	**0.22**	**0.3**	-	0.41	0.51	0.69	0.67	0.67	0.65
Macacu	**0.8**	**0.82**	**0.82**	0.76	0.78	0.8	0.8	0.78	0.62	0.65	0.65	0.68	0.67	0.7	0.63	0.51	0.55	**0.49**	0.59	0.46	0.43	0.4	0.48	0.41	-	0.39	0.71	0.69	0.7	0.67
Paraíba do Sul	**0.82**	**0.84**	**0.84**	**0.8**	**0.81**	**0.82**	**0.82**	**0.81**	**0.7**	0.73	**0.72**	**0.73**	0.74	**0.75**	0.72	**0.6**	**0.62**	**0.56**	0.67	**0.56**	**0.53**	**0.5**	**0.56**	**0.51**	0.39	-	0.74	0.73	0.74	0.71
Barra do Riacho	**0.86**	**0.88**	**0.88**	**0.85**	**0.86**	**0.87**	**0.87**	**0.86**	**0.76**	0.8	**0.78**	**0.79**	0.81	**0.81**	0.78	**0.7**	**0.71**	**0.63**	0.78	**0.69**	**0.65**	**0.66**	**0.71**	**0.69**	0.71	**0.74**	-	0.17	0.34	0.51
Barra Seca	**0.87**	**0.88**	**0.88**	**0.85**	**0.87**	**0.87**	**0.87**	**0.86**	**0.74**	0.8	**0.79**	**0.79**	0.83	**0.81**	0.78	**0.67**	**0.7**	**0.6**	0.79	**0.68**	**0.64**	**0.63**	**0.71**	**0.67**	0.69	**0.73**	**0.17**	-	0.29	0.46
Itaunas	**0.86**	**0.88**	**0.88**	**0.84**	**0.86**	**0.87**	**0.86**	**0.85**	**0.76**	0.79	**0.78**	**0.79**	0.81	**0.81**	0.78	**0.69**	**0.7**	**0.63**	0.77	**0.68**	**0.65**	**0.66**	**0.71**	**0.67**	0.7	**0.74**	**0.34**	**0.29**	-	0.42
Caravelas	**0.84**	**0.86**	**0.86**	**0.82**	**0.83**	**0.85**	**0.85**	**0.83**	**0.73**	0.77	**0.75**	**0.77**	0.78	**0.78**	0.76	**0.68**	**0.68**	**0.62**	0.75	**0.66**	**0.63**	**0.64**	**0.68**	**0.65**	0.67	**0.71**	**0.51**	**0.46**	**0.42**	-

(b) *H. boulengeri*

	Patos	*Três Forquilhas*	*Tramandaí*	*Florianópolis*	*Itapocu*	*Iguaçu*	*São Francisco do Sul*	*Sai-Mirim*	*Sai-Guaçu*	*Guaratuba*	*Paranaguá*	*Ribeira de Iguape*	*Itanhaém*	*Uper Tietê*	*Bertioga*	*Guapimirim*	*Barra do Riacho*	*Barra Seca*	*Itaunas*
Patos	-	0.19	**0.13**	**0.71**	**0.58**	**0.65**	**0.68**	**0.74**	**0.77**	0.74	0.74	**0.72**	**0.74**	**0.79**	**0.70**	0.77	**0.82**	**0.82**	0.82
Três Forquilhas	**0.19**	-	0.09	0.85	0.54	0.64	0.68	0.78	0.83	0.79	0.78	0.72	0.74	0.84	0.71	0.77	0.87	0.86	0.87
Tramandaí	**0.13**	**0.09**	-	**0.76**	**0.56**	0.65	**0.68**	**0.76**	**0.80**	0.75	0.76	**0.72**	**0.74**	**0.81**	**0.71**	0.77	**0.84**	0.84	0.83
Florianópolis	**0.71**	**0.85**	**0.76**	-	**0.66**	**0.76**	**0.78**	**0.86**	0.89	0.90	0.90	**0.81**	**0.83**	0.90	**0.81**	0.89	**0.93**	**0.92**	0.95
Itapocu	**0.58**	**0.54**	**0.56**	**0.66**	-	**0.13**	**0.20**	**0.37**	**0.43**	0.34	0.52	**0.59**	**0.60**	**0.66**	**0.51**	0.61	**0.70**	**0.70**	0.67
Iguaçu	**0.65**	**0.64**	**0.65**	**0.76**	**0.13**	-	**0.33**	**0.49**	**0.56**	0.47	0.63	**0.66**	**0.66**	**0.74**	0.60	0.68	**0.78**	0.77	0.75
São Francisco do Sul	**0.68**	**0.68**	**0.68**	**0.78**	**0.20**	**0.33**	-	**0.55**	0.61	0.53	0.67	**0.68**	**0.69**	**0.76**	**0.63**	0.72	**0.80**	0.79	0.77
Sai-Mirim	**0.74**	**0.78**	**0.76**	**0.86**	**0.37**	**0.49**	**0.55**	-	**0.08**	0.09	0.77	**0.75**	**0.75**	**0.83**	**0.71**	0.80	**0.86**	**0.85**	0.85
Sai-Guaçu	**0.77**	**0.83**	**0.80**	**0.89**	**0.43**	**0.56**	**0.61**	**0.08**	-	0.09	0.82	**0.77**	**0.78**	0.86	0.75	0.84	**0.89**	0.88	0.89
Guaratuba	**0.74**	0.79	**0.75**	**0.90**	**0.34**	**0.47**	**0.53**	**0.09**	**0.09**	-	0.78	0.73	0.74	0.84	0.70	0.78	0.88	0.87	0.87
Paranaguá	**0.74**	0.78	**0.76**	**0.90**	**0.52**	**0.63**	**0.67**	**0.77**	**0.82**	0.78	-	0.70	0.71	0.82	0.15	0.76	0.86	0.85	0.86
Ribeira de Iguape	**0.72**	**0.72**	**0.72**	**0.81**	**0.59**	**0.66**	**0.68**	**0.75**	**0.77**	**0.73**	**0.70**	-	**0.65**	**0.72**	**0.68**	0.70	**0.78**	**0.78**	0.76
Itanhaém	**0.74**	**0.74**	**0.74**	**0.83**	**0.60**	**0.66**	**0.69**	**0.75**	**0.78**	**0.74**	**0.71**	**0.65**	-	**0.21**	**0.68**	0.71	**0.79**	0.79	0.77
Uper Tietê	**0.79**	**0.84**	**0.81**	**0.90**	**0.66**	**0.74**	**0.76**	**0.83**	**0.86**	**0.84**	**0.82**	**0.72**	**0.21**	-	**0.76**	0.80	**0.86**	0.86	0.87
Bertioga	**0.70**	**0.71**	**0.71**	**0.81**	**0.51**	**0.60**	**0.63**	**0.71**	**0.75**	**0.70**	**0.15**	**0.68**	**0.68**	**0.76**	-	0.72	**0.80**	**0.80**	0.78
Guapimirim	**0.77**	0.77	**0.77**	**0.89**	**0.61**	**0.68**	**0.72**	**0.80**	**0.84**	0.78	0.76	**0.70**	**0.71**	**0.80**	**0.72**	-	0.80	0.79	0.76
Barra do Riacho	**0.82**	**0.87**	**0.84**	**0.93**	**0.70**	**0.78**	**0.80**	**0.86**	**0.89**	**0.88**	**0.86**	**0.78**	**0.79**	**0.86**	**0.80**	**0.80**	-	**0.07**	0.31
Barra Seca	**0.82**	**0.86**	**0.84**	**0.92**	**0.70**	**0.77**	**0.79**	**0.85**	**0.88**	**0.87**	**0.85**	**0.78**	**0.79**	**0.86**	**0.80**	**0.79**	**0.07**	-	0.20
Itaunas	**0.82**	0.87	**0.83**	**0.95**	**0.67**	**0.75**	**0.77**	**0.85**	**0.89**	0.87	0.86	**0.76**	**0.77**	**0.87**	**0.78**	0.76	**0.31**	**0.20**	-

(c) *Hollandichthys*

	Maquiné	*Três Forquilhas*	*Mampituba*	*Araranguá*	*D'Una*	*Florianópolis*	*São Francisco do Sul*	*Guaratuba*	*Paranaguá*	*Guaraqueçaba*	*Ribeira de Iguape*	*Guarau*	*Santos*	*Upper Tietê*	*Bertioga*	*Juquehy*	*Bouçucanga*	*Ilhabela*	*Ubatuba*	*Picinguaba*	*Toca do Boi*	*Tarituba*
Maquiné	-	**0.59**	**0.6**	0.85	0.91	**0.71**	**0.73**	0.86	**0.81**	**0.81**	0.82	**0.92**	**0.84**	**0.89**	**0.85**	0.93	**0.89**	**0.92**	0.93	0.94	**0.93**	**0.95**
Três Forquilhas	**0.59**	-	**0.44**	0.77	0.87	0.6	0.63	0.77	0.72	**0.74**	0.73	0.88	**0.78**	0.84	0.79	0.89	**0.84**	0.89	0.88	0.9	**0.9**	0.93
Mampituba	**0.6**	**0.44**	-	0.74	0.84	**0.67**	**0.68**	0.78	**0.75**	**0.78**	0.78	**0.88**	**0.81**	**0.85**	**0.82**	0.88	**0.86**	**0.89**	0.88	0.88	**0.9**	**0.92**
Araranguá	**0.85**	**0.77**	**0.74**	-	0.87	0.57	0.6	0.71	0.68	0.72	0.69	0.87	0.76	0.82	0.76	0.87	0.83	0.88	0.86	0.87	0.9	0.93
D'Una	**0.91**	**0.87**	**0.84**	**0.87**	-	0.41	0.57	0.72	0.67	0.71	0.7	0.87	0.76	0.82	0.77	0.88	0.83	0.88	0.87	0.89	0.89	0.93
Florianópolis	**0.71**	**0.6**	**0.67**	**0.57**	**0.41**	-	**0.32**	0.38	0.44	**0.59**	0.56	**0.71**	**0.65**	**0.68**	0.64	0.67	**0.68**	0.72	0.68	0.66	**0.78**	**0.78**
São Francisco do Sul	**0.73**	**0.63**	**0.68**	**0.6**	**0.57**	**0.32**	-	0.2	0.36	**0.57**	0.55	**0.71**	**0.64**	**0.67**	**0.62**	0.68	**0.68**	0.72	0.68	0.66	**0.77**	0.78
Guaratuba	**0.86**	**0.77**	**0.78**	0.71	0.72	**0.38**	**0.2**	-	0.39	0.61	0.56	0.79	0.67	0.74	0.66	0.74	0.74	0.8	0.73	0.71	0.85	0.87
Paranaguá	**0.81**	**0.72**	**0.75**	**0.68**	**0.67**	**0.44**	**0.36**	**0.39**	-	0.56	0.53	0.73	0.63	0.69	0.63	0.69	0.69	0.75	0.69	0.67	0.8	0.81
Guaraqueçaba	**0.81**	**0.74**	**0.78**	**0.72**	**0.71**	**0.59**	**0.57**	**0.61**	**0.56**	-	0.39	**0.61**	**0.51**	**0.56**	**0.5**	0.56	**0.56**	**0.65**	0.6	0.57	**0.71**	**0.72**
Ribeira de Iguape	**0.82**	**0.73**	**0.78**	**0.69**	**0.7**	**0.56**	**0.55**	0.56	**0.53**	**0.39**	-	0.46	0.35	0.41	0.36	0.43	0.45	0.58	0.49	0.45	0.67	0.67
Guarau	**0.92**	**0.88**	**0.88**	**0.87**	**0.87**	**0.71**	**0.71**	**0.79**	**0.73**	**0.61**	**0.46**	-	**0.53**	**0.63**	**0.58**	0.72	**0.67**	0.78	0.75	0.76	**0.82**	0.85
Santos	**0.84**	**0.78**	**0.81**	**0.76**	**0.76**	**0.65**	**0.64**	**0.67**	**0.63**	**0.51**	**0.35**	**0.53**	-	**0.28**	**0.39**	0.49	**0.48**	0.63	0.58	0.56	**0.7**	**0.71**
Upper Tietê	**0.89**	**0.84**	**0.85**	**0.82**	**0.82**	**0.68**	**0.67**	**0.74**	**0.69**	**0.56**	**0.41**	**0.63**	**0.28**	-	**0.46**	0.6	**0.57**	**0.71**	0.68	0.67	**0.76**	**0.79**
Bertioga	**0.85**	**0.79**	**0.82**	**0.76**	**0.77**	**0.64**	**0.62**	**0.66**	**0.63**	**0.5**	**0.36**	**0.58**	**0.39**	**0.46**	-	0.33	0.3	0.62	0.56	0.54	**0.7**	0.71
Juquehy	**0.93**	**0.89**	**0.88**	0.87	**0.88**	**0.67**	**0.68**	0.74	**0.69**	**0.56**	**0.43**	**0.72**	**0.49**	**0.6**	**0.33**	-	0.44	0.74	0.68	0.69	0.8	0.84
Bouçucanga	**0.89**	**0.84**	**0.86**	**0.83**	**0.83**	**0.68**	**0.68**	**0.74**	**0.69**	**0.56**	**0.45**	**0.67**	**0.48**	**0.57**	**0.3**	**0.44**	-	0.7	0.65	0.65	**0.76**	0.78
Ilhabela	**0.92**	**0.89**	**0.89**	**0.88**	**0.88**	**0.72**	**0.72**	**0.8**	**0.75**	**0.65**	**0.58**	**0.78**	**0.63**	**0.71**	**0.62**	**0.74**	**0.7**	-	0.37	0.38	**0.61**	0.67
Ubatuba	**0.93**	**0.88**	**0.88**	0.86	**0.87**	**0.68**	**0.68**	0.73	**0.69**	**0.6**	**0.49**	**0.75**	**0.58**	**0.68**	**0.56**	0.68	**0.65**	**0.37**	-	0.17	0.55	0.63
Picinguaba	**0.94**	**0.9**	**0.88**	0.87	**0.89**	**0.66**	**0.66**	0.71	0.67	**0.57**	0.45	**0.76**	**0.56**	**0.67**	**0.54**	0.69	**0.65**	**0.38**	0.17	-	0.55	0.66
Toca do Boi	**0.93**	**0.9**	**0.9**	**0.9**	**0.89**	**0.78**	**0.77**	**0.85**	**0.8**	**0.71**	**0.67**	**0.82**	**0.7**	**0.76**	**0.7**	**0.8**	**0.76**	**0.61**	**0.55**	**0.55**	-	**0.68**
Tarituba	**0.95**	**0.93**	**0.92**	**0.93**	**0.93**	**0.78**	**0.78**	**0.87**	**0.81**	**0.72**	**0.67**	**0.85**	**0.71**	**0.79**	**0.71**	**0.84**	**0.78**	**0.67**	**0.63**	**0.66**	**0.68**	-

(d) *Bryconamericus*

	Três Forquilhas	Araranguá	Itajai	Itapocu	Cubatão Norte	Guaratuba	Paranaguá	Ribeira de Iguape	Pereque-Açu	Taquari	Guapimirim	Macacu	São João
Três Forquilhas	-	0.11	0.77	0.48	0.63	0.69	0.40	0.53	0.85	0.82	0.82	0.85	0.81
Araranguá	**0.11**	-	0.78	0.50	0.64	0.69	0.43	0.55	0.85	0.82	0.82	0.85	0.81
Itajai	**0.77**	**0.78**	-	0.69	0.82	0.86	0.61	0.31	0.80	0.73	0.77	0.79	0.75
Itapocu	**0.48**	**0.50**	**0.69**	-	0.28	0.37	0.21	0.52	0.77	**0.76**	0.72	0.77	0.73
Cubatão Norte	**0.63**	**0.64**	**0.82**	**0.28**	-	0.34	0.30	**0.62**	0.88	**0.85**	0.86	0.88	0.85
Guaratuba	**0.69**	**0.69**	**0.86**	**0.37**	**0.34**	-	0.35	**0.64**	0.90	**0.87**	0.89	0.90	0.88
Paranaguá	0.40	**0.43**	**0.61**	0.21	**0.30**	**0.35**	-	0.39	0.71	0.70	0.65	0.72	0.66
Ribeira de Iguape	**0.53**	**0.55**	**0.31**	**0.52**	**0.62**	**0.64**	**0.39**	-	**0.43**	**0.41**	0.43	**0.42**	0.47
Pereque-Açu	**0.85**	**0.85**	**0.80**	**0.77**	**0.88**	**0.90**	**0.71**	**0.43**	-	0.54	0.88	0.72	0.85
Taquari	**0.82**	**0.82**	**0.73**	**0.76**	**0.85**	**0.87**	**0.70**	**0.41**	**0.54**	-	0.79	0.59	0.79
Guapimirim	0.82	**0.82**	**0.77**	**0.72**	**0.86**	**0.89**	0.65	**0.43**	**0.88**	**0.79**	-	0.87	0.80
Macacu	**0.85**	**0.85**	**0.79**	**0.77**	**0.88**	**0.90**	**0.72**	**0.42**	**0.72**	**0.59**	**0.87**	-	0.85
São João	**0.81**	**0.81**	**0.75**	**0.73**	**0.85**	**0.88**	**0.66**	**0.47**	**0.85**	**0.79**	**0.80**	**0.85**	-

Table D.6. Point estimate of demographic parameters for divergence models per species for each one of the three common identified geographic breaks estimated across 40 runs of FASTSIMCOAL2. Specifically, divergence time, T_{DIV}, ancestral population size, N_{ANC}, and the population size for the northern population of the break, N_2, as well as the number of loci used to calculate the site frequency spectrum (SFS). The 95% confidence intervals are shown between parentheses. Note that the population size of the southern population per breaks (N_1) was not estimated under a divergence model, but instead was calculated directly from the empirical data (i.e., it is a fixed parameter in the model) to improve the accuracy of the other parameters estimated from the SFS (following the recommendations for the program; see Excoffier and Foll, 2011).

Break	Species	# Ind.	Loci	Genome Size (C-value)*	μ	N_{ANC}	N_1	N_2	T_{DIV}
North	*M. microlepis*	12	1035	1.53 *M. microlepis*	2.27E-08	14444 (10097-18154)	41850	18098 (15046-23085)	25937 (21670-32997)
	H. boulengeri	15	2083	1.15 *H. reticulatus*	1.87E-08	96521 (51678-103063)	24064	39876 (35784-44677)	135418 129892-167069)
Middle	*M. microlepis*	15	1235	1.53 *M. microlepis*	2.27E-08	15158 (10997-19055)	70485	24893 (20830-30426)	26453 (23919-33252)
	H. boulengeri	15	4385	1.15 *H. reticulatus*	1.87E-08	66139 (53172-66256)	50802	26644 24679-29591)	21456 (20991-23824)
	Hollandichthys	15	5533	1.5 "Clade C"	2.24E-08	73595 (62392-77854)	98214	139357 (129456-152669)	62845 (61037-67809)
	Bryconamericus	15	4142	1.64 *B. stramineus*	2.38E-08	40127 (33669-44818)	81933	90945 (85812-101269)	62652 (59590-70291)
	Bryconamericus - No admixture	15	4102	1.64 *B. stramineus*	2.38E-08	30975 (21617-34932)	58824	52204 (48687-57493)	76940 (73003-85261)
South	*M. microlepis*	15	3004	1.53 *M. microlepis*	2.27E-08	8798 (7399-9694)	26432	5169 (4629-5933)	6689 (6288-7822)
	Hollandichthys	15	5785	1.5 "Clade C"	2.24E-08	23054 (19614-23181)	42411	6331 (5907-7044)	9709 (9194-11112)
	Bryconamericus	7	6133	1.64 *B. stramineus*	2.38E-08	24873 (18169-28688)	42017	66225 (61494-71749)	74949 (72852-82990)

* Carvalho, M. L., Oliveira, C., Navarrete, M. C., Froehlich, O., & Foresti, F. (2002). Nuclear DNA content determination in Characiformes fish (Teleostei, Ostariophysi) from the Neotropical region. Genetics and Molecular Biology, 25(1), 49-55.

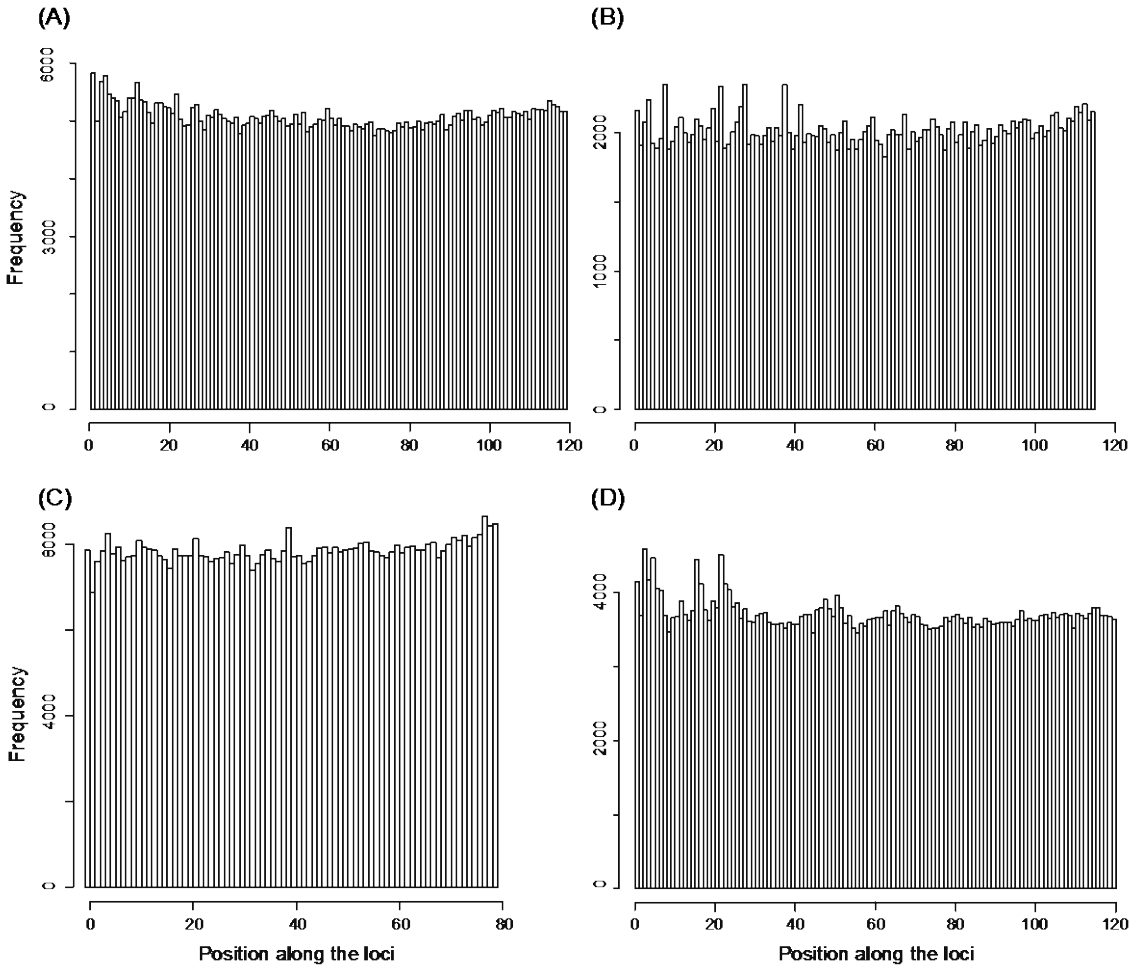


Figure D.1. Summary of the frequency of segregating sites for each base-pair position of a locus per species: *M. microlepis* (A), *H. boulengeri* (B), *Hollandichthys* (C), *Bryconamericus* (D).

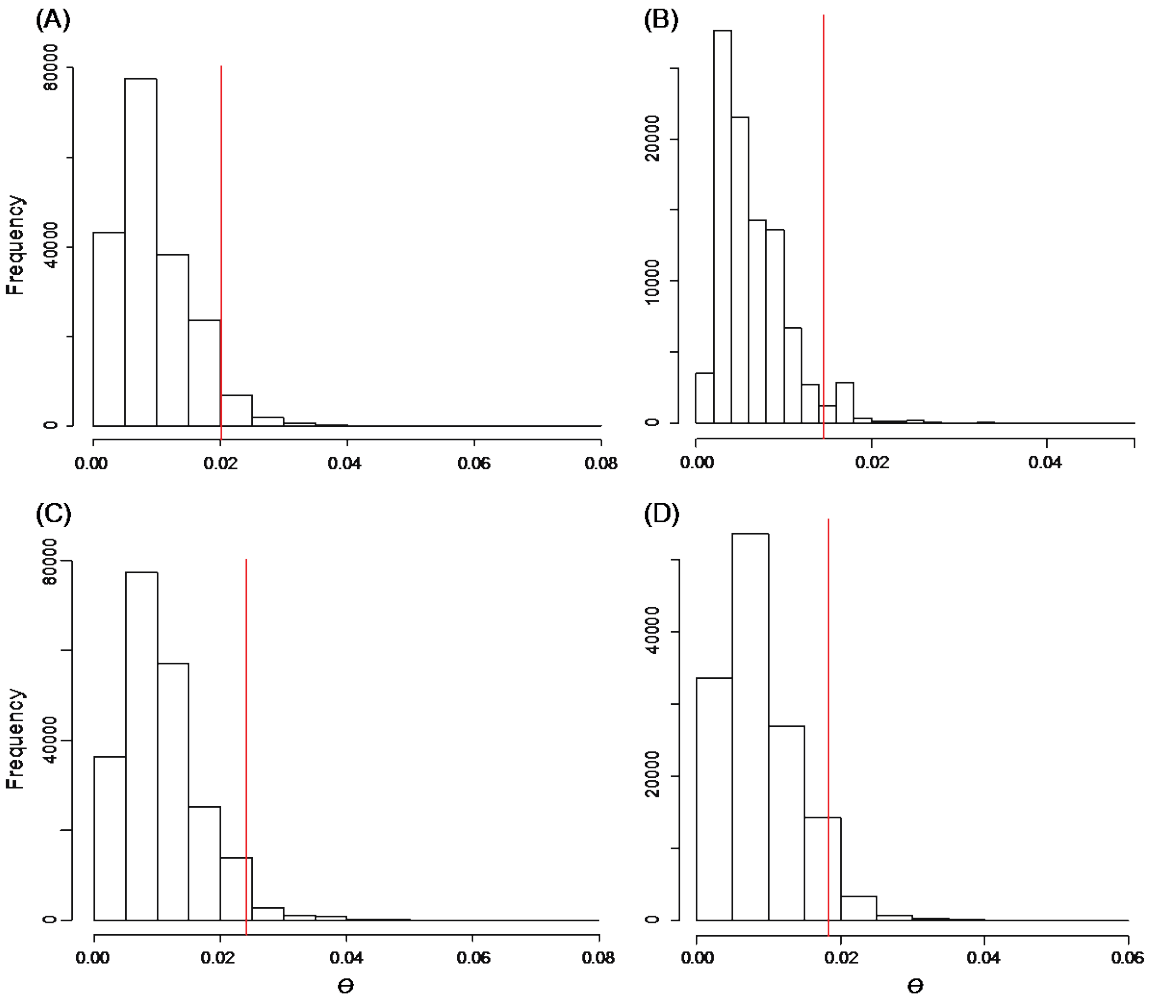


Figure D.2. Distribution of theta, θ, per loci, with the red line marking the θ-values in the 95 percentile that were excluded from analyses to avoid including variation likely reflective of sequencing and/or assembly errors in each species: *M. microlepis* (A), *H. boulengeri* (B), *Hollandichthys* (C), *Bryconamericus* (D).

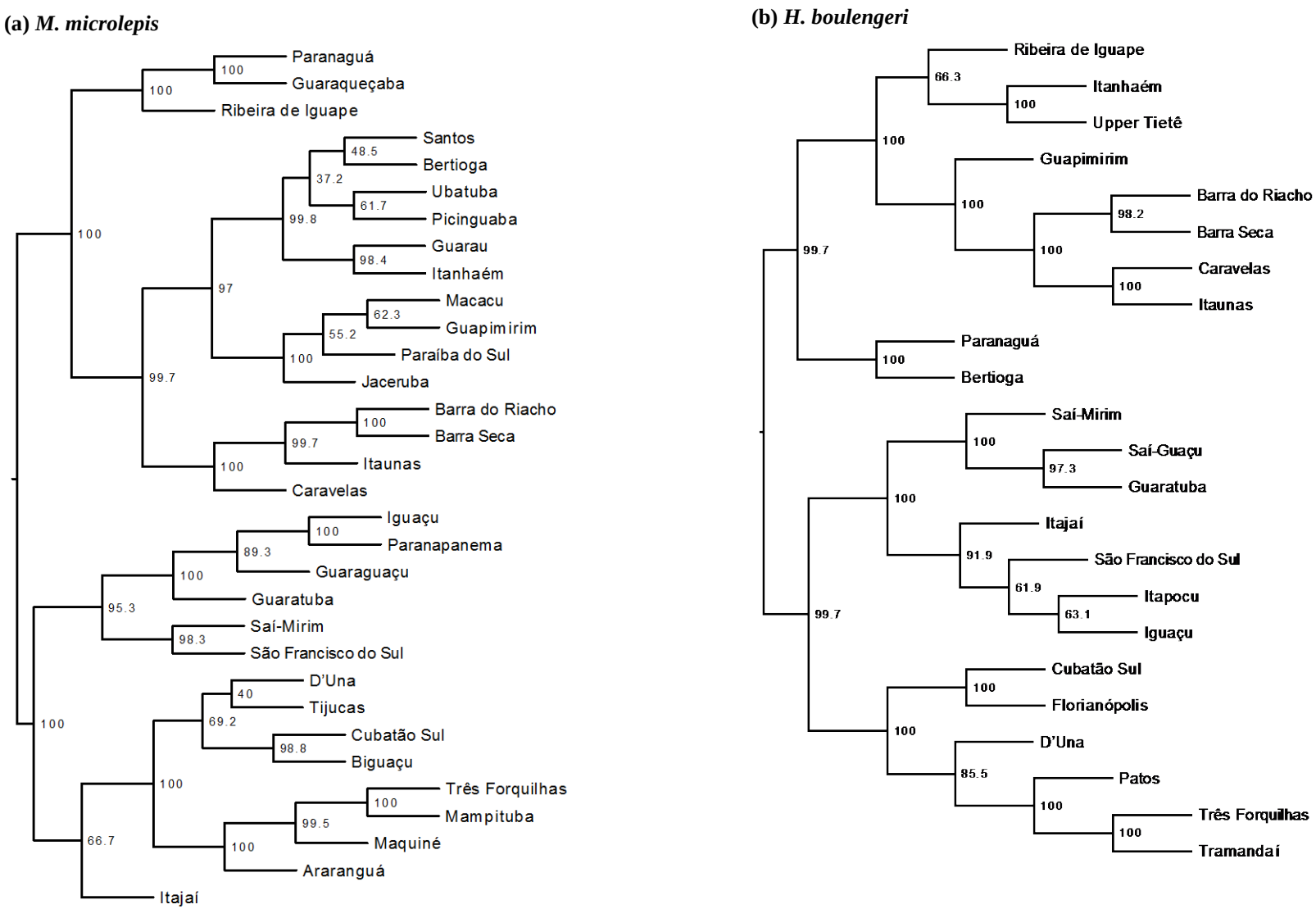
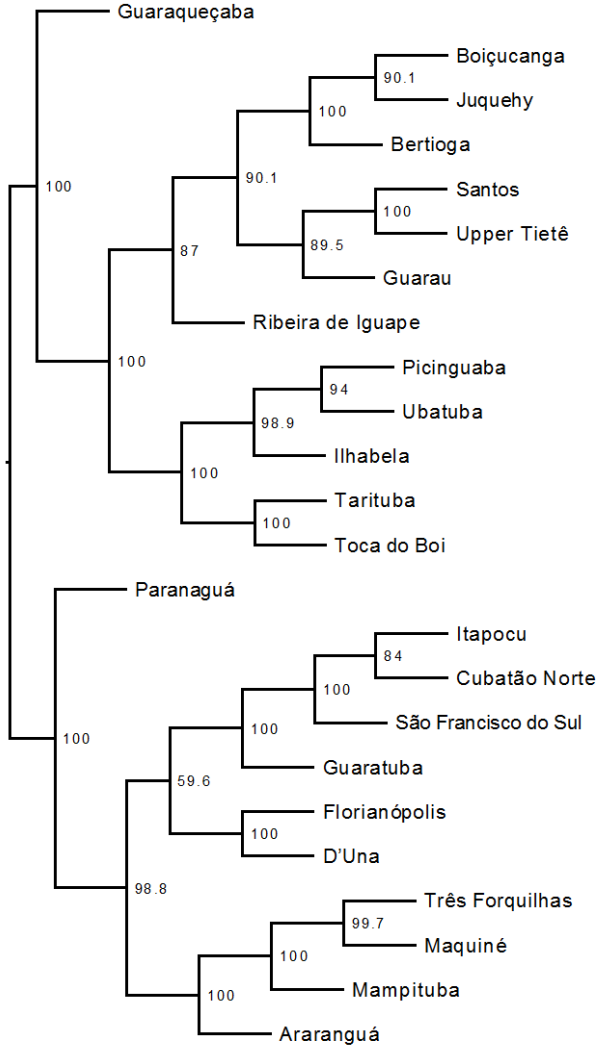
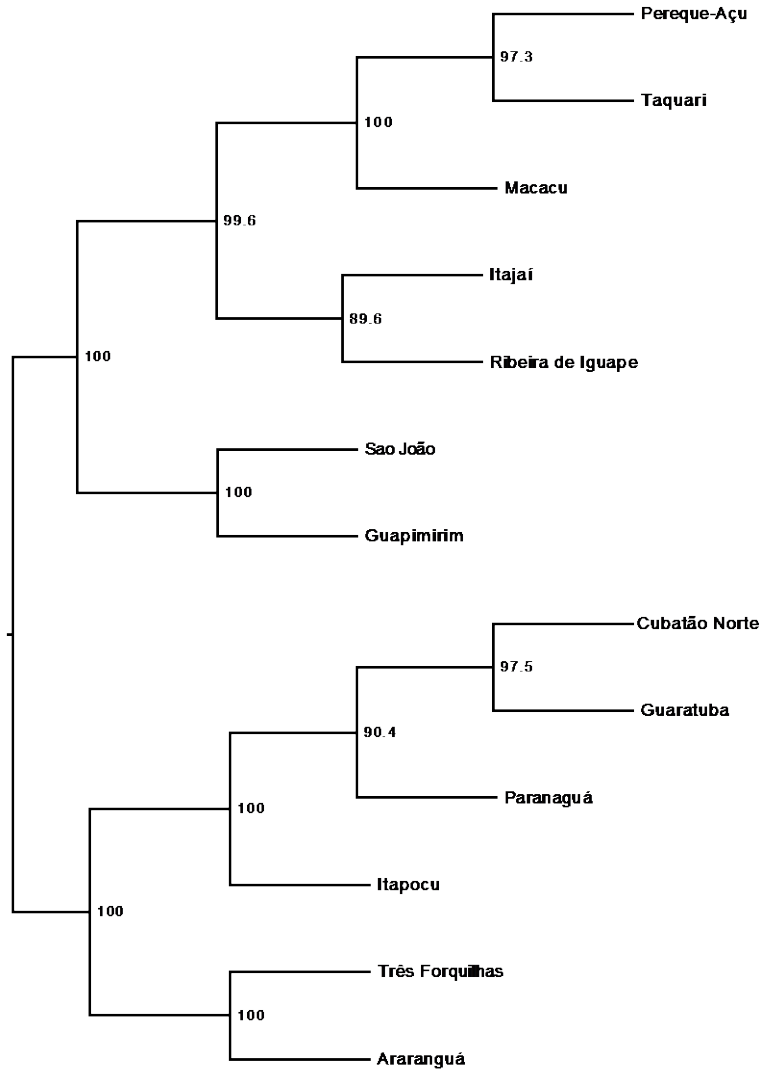


Figure D.3. Phylogenetic trees estimated with SVDquartets at the population level for each species (A = *M. microlepis*, B = *H. boulengeri*, C = *Hollandichthys* and D = *Bryconamericus*). Bootstraps support values are shown on each node – *continue next page.*

(c) *Hollandichthys*



(d) *Bryconamericus*



CHAPTER VI

Conclusions and Future Directions

The work presented here advances our knowledge on the evolution of freshwater fishes, with special focus on the diverse Neotropical ichthyofauna. Overall, by exploring relevant hypotheses to identify processes that structure genetic variation within and between basins, and across species, I was able to distinguish among evolutionary processes operating at different spatial and temporal scales in freshwater fishes along coastal Brazilian basins.

After almost 30 years of the first preposition that sea level changes may have influenced speciation and diversification processes of freshwater fishes along the Brazilian coastal basins (Weitzman *et al.*, 1988), this dissertation presented an extensive test of the effects of Pleistocene glacial cycles in promoting divergence in the study area. The findings indicate that diversification during Pleistocene is more conspicuous than originally thought and unveil the strong genetic structure observed within species. These results will impact the way researchers perceive the ichthyofauna in the area.

The system of drainages along the Brazilian coast is demonstrated here to be a model region to study evolution of the diverse Neotropical fishes given its isolation, and a general good understanding of the geological and climatic history. Likewise, freshwater fishes are ideal organisms to study evolutionary consequences of climatic changes during the Pleistocene given the strong constraint imposed by the environment they inhabit, as opposed to the terrestrial fauna. The coastal region has been largely studied and receives large conservation efforts for terrestrial organisms because it is considered one of the hotspots of diversity – the Atlantic Rainforest biome. Freshwater fishes, on the other hand, still lack genetic studies. The strong latitudinal structure assessed in this dissertation, along with several species being considered by taxonomists as a species complex, demonstrate that the diversity of freshwater fishes may still be underestimated in the area, suggesting even higher levels of endemism in the region. This region is also highly urbanized, which means the ichthyofauna is very much at risk. By proposing a foundational scenario of diversification for this area, my work advocates for the expansion of conservation measures and biodiversity sensing to include bodies of water and their organisms,

which unfortunately have received considerably less attention than terrestrial organisms.

Joining the knowledge available for the region with newly developed techniques that combine simulations and empirical data allowed me to create scenarios (i.e., GIS and simulations) and estimate parameters (i.e., divergence estimates and population sizes) that wouldn't have been possible a few years ago. This approach can generally inform conservation measures in riverine environments and how strategies might be effective in the long term. Furthermore, with further developments of the approach, such as the incorporation of a more expansive suite of riverine properties and ecological information, one of my future goals is to expand its utility as a general tool (i.e., it can be applied to more taxa or other ecosystems than studied in this dissertation). For example, I would like to expand the analysis performed here to be broadly applicable to riverine environments in other world regions (and especially in the Neotropics), providing insights into the reasons there is a large number of freshwater fish species on the planet and understand which rivers and/or parts of rivers are critical to species and genetic diversity because of their impact on population persistence and movement patterns of organisms living in these environments. By filling this knowledge gap I believe researchers and the general scientific community (i.e., government and environmental institutions) will be able to think more strategically for a sustainable path forward.

The project developed here started to be built ten years ago, towards the end of my undergraduate studies. Then, when I started my Ph.D. back in 2011, I naively expected that the Ph.D. would be the last step as a student. Today, as I arrive in the end, I realize that this is just the beginning of a much longer journey. Most of the work I envisioned in the beginning is still under development, many new questions arose and different directions were taken. From today, many years of learning, researching and dedication will still come until a more complete picture of the processes driving the diversification of Neotropical freshwater organisms can be developed. I am eager and enthusiastic to discover the path that this research will take. As a scientist in a delicate time for scientific advances and sustainable efforts over the world, I am committed to diversity and the variety of life.

References

Weitzman SH, Menezes NA, Weitzman MJ (1988). Phylogenetic biogeography of the Glandulocaudini (Teleostei: Characiformes, Characidae) with comments on the distributions

of other freshwater fishes in eastern and southeastern Brazil. *Proceedings of a workshop on neotropical distribution patterns*. Academia Brasileira de Ciências, Rio de Janeiro, pp. 379–427.