



UFRGS
UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL



INSTITUTO DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA ANIMAL

MATEUS SANTOS DE SOUZA

**PIRATAS DE ÁGUA DOCE: TESTANDO HISTÓRIAS DE CAPTURA DE RIO EM
PEIXES ATRAVÉS DE DADOS GENÉTICOS**

PORTO ALEGRE

2020

MATEUS SANTOS DE SOUZA

**PIRATAS DE ÁGUA DOCE: TESTANDO HISTÓRIAS DE CAPTURA DE RIO EM
PEIXES ATRAVÉS DE DADOS GENÉTICOS**

Dissertação apresentada ao Programa de Pós-Graduação em Biologia Animal, Instituto de Biociências da Universidade Federal do Rio Grande do Sul, como requisito parcial à obtenção do título de Mestre em Biologia Animal.

Área de concentração: Biologia Comparada

Orientador: Prof. Dr. Nelson Jurandi Rosa Fagundes

Coorientadora: Dra. Andréa Tonolli Thomaz

PORTO ALEGRE

2020

MATEUS SANTOS DE SOUZA

**PIRATAS DE ÁGUA DOCE: TESTANDO HISTÓRIAS DE CAPTURA DE RIO EM
PEIXES ATRAVÉS DE DADOS GENÉTICOS**

Aprovada em ____ de _____ de ____.

BANCA EXAMINADORA

Dr. Marcelo Coelho Miguel Gehara (Rutgers University)

Dr. Sandro Luis Bonatto (PUCRS)

Dr. Luiz Roberto Malabarba (UFRGS)

“There is a little river-pirate in eastern Pennsylvania, unsuspected by its rural neighbors. It is in a quiet, well-watered farming district, where the streams, as a rule, are bent only on minding their own business, and not interfering with their fellows; and yet one of them is a confirmed pirate, and goes on unhindered in its robbery.”

Fragmento de *A River-pirate* (Davis, W. M., 1889. *Science* 13: 108 - 109).

AGRADECIMENTOS

Agradeço a todos que contribuíram para o desenvolvimento desta dissertação, em especial aos meus orientadores Nelson e Andréa, aos colegas e professores do Laboratório de Ictiologia e do Laboratório de Genética Médica e Evolução da UFRGS e a minha família. Agradeço também ao CNPq pela concessão de bolsa de mestrado e ao PPGBAN e a UFRGS pela estrutura disponibilizada.

SUMÁRIO

RESUMO	4
ABSTRACT	6
INTRODUÇÃO GERAL	8
CONSIDERAÇÕES FINAIS	15

RESUMO

Os peixes de água doce constituem um grupo diversificado de organismos que abrange cerca de um quarto do total de espécies de vertebrados. Interessantemente, essa grande diversidade ocorre em uma fração pequena do volume de água habitável do planeta. Uma das causas desse fenômeno é o isolamento provocado pela fragmentação dos ambientes de água doce em drenagens isoladas entre si, o qual propicia diferenciação genética entre populações habitando diferentes bacias, possivelmente culminando em especiação alopátrica. No entanto, diversos estudos filogeográficos reportam casos de similaridade genética entre populações de diferentes bacias. Fenômenos climático ou geológicos, como as capturas de rio, são normalmente invocados para explicar esse padrão. Na captura de rio, a proximidade genética entre indivíduos de diferentes bacias é explicada através de um contato secundário entre populações previamente isoladas. No entanto, a manutenção de polimorfismos ancestrais pode ser tão plausível quanto um contato secundário, mas essas hipóteses raramente são testadas explicitamente através de um método estatístico. O objetivo do presente estudo foi avaliar o desempenho da metodologia de Computação Bayesiana Aproximada (ABC) para testar, com base em dados genéticos empíricos e simulados, cenários explícitos de contato secundário causado por capturas de rio contra cenários de polimorfismo compartilhado em peixes de água doce. Inicialmente, foram avaliadas as conclusões de um estudo da literatura sobre os padrões de estrutura genética entre bacias hidrográficas observados em *Cnesterodon decemmaculatus* (Cyprinodontiformes: Poeciliidae) sugeridos como possíveis casos de captura de rio. Posteriormente, foi realizado um estudo de simulação, a fim de avaliar o desempenho da metodologia ABC em distinguir entre cenários em diferentes condições, variando desde o uso de diferentes marcadores moleculares e estratégias de amostragem até diferentes parâmetros demográficos. Para o estudo com *C. decemmaculatus*, os resultados demonstraram que o

mtDNA tem potencial para distinguir entre os cenários, tornando possível a escolha do cenário mais provável na maior parte dos casos avaliados - embora esse marcador apresente algumas limitações, especialmente quando a divergência entre as populações é recente. A partir do estudo de simulações, foi verificado que o uso de milhares de polimorfismos de nucleotídeo único (SNPs) independentes é uma opção altamente recomendável quando comparado ao uso de um ou poucos loci, ainda que sua análise na perspectiva de ABC demande um maior trabalho computacional. Porém, a identificação de cenários de contato secundário com divergência recente e/ou com baixo número de migrantes é desafiadora, mesmo com o uso de SNPs.

Palavras-chave: *Cnesterodon decemmaculatus*, Computação Bayesiana Aproximada, contato secundário, DNA mitocondrial, Seleção de modelos, SNPs.

ABSTRACT

Freshwater fish constitute a diversified group of organisms that account for about one quarter of the total number of vertebrate species. Interestingly, this great diversity is restricted to a small fraction of the habitable water volume in the planet. One of the causes for this phenomenon is the isolation caused by the fragmentation of freshwater environments into isolated drainages, which promotes genetic differentiation between populations inhabiting different hydrographic basins and possibly culminating in allopatric speciation. However, several phylogeographic studies report instances of great genetic similarity between populations from different basins. Climatic or geological phenomena, such as river capture, are usually invoked for explaining this pattern. Under a river capture, a close genetic relationship between individuals from different basins is due to a secondary contact between populations that were isolated until then. However, retention of ancestral polymorphisms can be as plausible as a secondary contact, but these hypotheses are rarely tested explicitly through the use of a statistical method. The aim of the present study was to evaluate the performance of the ABC for testing, based on empirical and simulated genetic data, hypotheses of secondary contact generated by river captures versus shared polymorphisms in freshwater fish. Initially, we evaluated the conclusions of a study from the literature about the patterns of genetic structure between basins observed in *Cnesterodon decemmaculatus* (Cyprinodontiformes: Poeciliidae), which were suggested as possible cases of river capture. In the following, we performed a simulation study in order to evaluate the performance of the ABC methodology in distinguishing between scenarios under different conditions, ranging from different molecular markers to different demographic parameters. In the study with *C. decemmaculatus*, results showed that mtDNA has potential for distinguishing between the scenarios, which made it possible to choose the most probable scenario in almost all cases evaluated, but also showed

that this marker has some limitations, especially when population divergence is recent. From the simulations study, it was clear that the use of thousands of independent single nucleotide polymorphisms (SNPs) is a highly recommended option when compared to the use of one or few loci, even though its analysis under an ABC framework requires more computational work. However, the identification of secondary contact scenarios with recent divergence and/or with a low number of migrants is challenging, even using SNPs.

Keywords: *Cnesterodon decemmaculatus*, Approximate Bayesian Computation, Secondary Contact, mtDNA, Model Choice, SNPs.

INTRODUÇÃO GERAL

Os peixes de água doce constituem um grupo diversificado e que abrange cerca de 13.000 espécies, correspondendo a aproximadamente um quarto do total de espécies de vertebrados (Lévêque et al., 2008; Cavin, 2017). Essa grande diversidade de espécies ocorre em uma fração pequena da superfície (0,08%) e do volume de água habitável (0,02%) do planeta (Dawson, 2012), um fato notável e que recebeu o nome de *freshwater fish paradox* (Tedesco et al., 2016). Dentre os fatores responsáveis por essa característica está o isolamento provocado pela fragmentação dos ambientes de água doce em rios e lagos (Tedesco et al., 2016), que é favorecido pela limitada capacidade de dispersão de grande parte dos peixes (Myers, 1938; Vari, 1988; Albert e Carvalho, 2011), fatores estes que propiciam taxas altas de divergência e, conseqüentemente, especiação nesses ambientes.

Devido à grande estruturação dos ambientes aquáticos de água doce, como no caso de rios em bacias hidrográficas isoladas, se esperaria que o isolamento também fosse refletido na diversidade genética das populações dos peixes, especialmente em espécies com distribuição ampla, englobando diferentes bacias hidrográficas. No entanto, diversos estudos filogeográficos mostram que, apesar de haver certo nível de estruturação genética populacional, indivíduos de determinadas bacias podem ser mais proximamente relacionados a indivíduos de uma bacia vizinha. Tomando o DNA mitocondrial como marcador genético, por exemplo, isso significa que não há monofilia recíproca entre as bacias, inclusive com alguns casos de compartilhamento de haplótipos entre elas (Lee & Johnson, 2009; Ramos-Fregonezi et al., 2017; Eaton et al., 2018; Schonhuth et al., 2018).

Esse tipo de incongruência entre a atual configuração das bacias hidrográficas - virtualmente isoladas entre si - e a proximidade genética entre alguns dos indivíduos de suas populações é normalmente interpretado como o resultado de um fenômeno climático, como

recuos no nível do mar durante períodos glaciais (Weitzman et al., 1988; Thomaz & Knowles, 2018) ou geológico, como as capturas de cabeceira ou de rio (Bishop, 1995), que alterou a conectividade entre bacias após a divergência dessas populações.

As capturas de cabeceira podem ocorrer por diversas causas, desde processos erosivos locais a grandes eventos de tectônica de placas, mas resultam basicamente na transferência de segmentos de rios entre bacias adjacentes. Quando se assume uma captura de rio, a proximidade genética entre indivíduos de diferentes bacias se deve ao fato de que alguns indivíduos foram levados de uma bacia a outra junto com o segmento de rio capturado, o que proporciona um contato secundário entre populações que estavam isoladas até então.

No entanto, um evento de contato secundário não é a única explicação possível para a ausência de monofilia recíproca entre bacias hidrográficas. A manutenção de polimorfismo ancestral (i.e., que já estava presente na população ancestral que originou as populações atuais) pode ser tão plausível quanto um contato secundário. Ictiólogos geralmente reconhecem essa possibilidade, mas não confrontam explicitamente as hipóteses alternativas de história populacional. De fato, comparar essas histórias populacionais realmente não é uma tarefa simples. A genealogia das amostras inferida nos estudos filogeográficos possui certo nível de incerteza associado (i.e., dados faltantes, processos estocásticos e dificuldades de reconstrução), o que é inerente aos métodos aplicados para reconstruí-las, e, mesmo assim, a mesma genealogia pode ser plausível em diferentes cenários demográficos (Nielsen & Beaumont, 2009), sejam eles com ou sem captura de rio. Por exemplo, um cenário de divergência recente entre populações com grande tamanho efetivo favorece o compartilhamento de haplótipos (no caso, a manutenção de polimorfismo ancestral observável) mesmo que as populações estejam completamente isoladas. Porém, um cenário no qual o compartilhamento é devido a um contato secundário pode ser igualmente plausível. Claramente, um dos grandes desafios ao se fazer inferências sobre os processos responsáveis

por gerar um determinado padrão de diversidade genética é levar em conta o “espaço” genealógico e o “espaço” de parâmetros demográficos possíveis.

Um método estatístico que tem sido utilizado para enfrentar esse desafio é a Computação Bayesiana Aproximada (ABC; Bertorelle et al., 2010; Csilléry et al., 2010). A ideia por trás da metodologia é utilizar um grande número de simulações de dados genéticos para cenários demográficos específicos com as mesmas características do dado observado (i.e., mesmo número de sítios ou loci e mesmo tamanho amostral) e comparar estatísticas sumário das simulações e do dado observado. Ao reter uma pequena quantidade de simulações que produziram estatísticas sumário mais similares às daquelas do dado observado, uma aproximação da probabilidade de obter o dado observado condicional ao modelo demográfico (i.e., a função de verossimilhança, ou *likelihood*) é produzida. Essa aproximação da verossimilhança possibilita a inferência de parâmetros demográficos e da probabilidade relativa para cada cenário alternativo em uma perspectiva bayesiana (Nielsen & Beaumont, 2009).

Recentemente, Machado et al. (2018) usaram essa metodologia para testar hipóteses de vicariância e geodispersão (envolvendo captura de rio) antigas em um grupo de peixes neotropicais usando dados de sequência de DNA. Esses autores caracterizaram seu cenário de geodispersão através de um efeito fundador durante o evento de captura de rio e, em geral, encontraram probabilidades relativamente pequenas para os cenários mais compatíveis com os dados (Machado et al., 2018). Porém, esse trabalho estava mais interessado no primeiro momento de dispersão de uma bacia para outra, e não na possibilidade de haver contato secundário entre populações previamente isoladas. Sousa et al. (2012) propuseram a utilização de ABC para estimar a probabilidade relativa de cenários alternativos para confrontar hipóteses de contato secundário e de retenção de polimorfismo ancestral em peixes de água doce. Para isso, eles avaliaram o desempenho dos marcadores de microssatélites utilizando simulações e

mostraram que, com um número razoável de loci (~20), é possível identificar com grande confiança em qual dos dois cenários um dado foi gerado.

Porém, embora esses trabalhos tenham demonstrado o potencial da metodologia ABC para avaliar cenários de capturas de rio, várias questões permanecem em aberto, tais como uma avaliação do desempenho da metodologia ABC quando são usados marcadores moleculares, como o DNA mitocondrial (mtDNA), amplamente utilizado em estudos de peixes de água doce, e de polimorfismos de nucleotídeo único (*Single Nucleotide Polymorphisms* - SNPs) para diferentes conjuntos de parâmetros demográficos (i.e. divergência antiga vs. recente; data do contato secundário; intensidade do contato secundário; impacto do tamanho amostral, etc).

OBJETIVOS

O objetivo geral do presente estudo foi avaliar o desempenho da metodologia de Computação Bayesiana Aproximada (ABC) para testar, com base em dados genéticos, hipóteses de contato secundário causado por capturas de rio contra hipóteses de polimorfismo compartilhado em peixes de água doce. Para tanto, o trabalho foi dividido em duas etapas: na primeira, escolhemos um trabalho da literatura, Ramos-Fregonezi et al. (2017), no qual os padrões de diversidade genética exibidos no mtDNA de *Cnesterodon decemmaculatus* (Jenyns, 1842) foram sugeridos como sendo o resultado de possíveis casos de capturas de rio. O objetivo específico foi avaliar, baseado em um caso empírico, a capacidade de distinguir entre ambos os cenários utilizando ABC. Na segunda etapa, foi realizado um estudo de simulação, cujo objetivo específico foi avaliar como o uso de diferentes marcadores (mtDNA ou SNPs) de diferentes tamanhos amostrais e diferentes parâmetros demográficos (tempo desde a divergência entre populações, tempo desde o contato secundário e proporção de linhagens migrantes) afeta o poder da metodologia ABC em recuperar o cenário correto.

REFERÊNCIAS

- Albert JS, Carvalho TP. 2011.** Neogene Assembly of Modern Faunas. In: Albert JS, Reis RE, eds. *Historical biogeography of neotropical freshwater fishes*. Berkeley and Los Angeles: University of California Press, 118-136.
- Bertorelle G, Benazzo A, Mona S. 2010.** ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology* 19: 2609-2625.
- Bishop P. 1995.** Drainage rearrangement by river capture, beheading and diversion. *Progress in Physical Geography* 19: 449-473.
- Cavin L. 2017.** Freshwater fish: 250 million years of evolutionary history. London: ISTE.
- Csilléry K, Blum MGB, Gaggiotti OE, François A. 2010.** Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology and Evolution* 25: 410-418.
- Dawson MN. 2012.** Species richness, habitable volume, and species densities in freshwater, the sea, and on land. *Frontiers of biogeography* 4: 105-116.
- Eaton KR, Loxterman JL, Keeley ER. 2018.** Connections and containers: Using genetic data to understand how watershed evolution and human activities influence cutthroat trout biogeography. *Plos One* 13: 1-30. <https://doi.org/10.1371/journal.pone.0202043>.
- Lee JB, Johnson JB. 2009.** Biogeography of the livebearing fish *Poecilia gillii* in Costa Rica: are phylogeographical breaks congruent with fish community boundaries? *Molecular Ecology* 18: 4088-4101.
- Lévêque C, Oberdorff T, Paugy D, Stiassny MLJ, Tedesco PA. 2008.** Global diversity of fish (Pisces) in freshwater. In: *Freshwater Animal Diversity Assessment*. Dordrecht: Springer, 545-567.
- Machado CB, Galleti Jr PM, Carnaval AC. 2018.** Bayesian analyses detect a history of both vicariance and geodispersal in Neotropical freshwater fishes. *Journal of Biogeography* 45: 1313-1325.

- Myers GS. 1938.** Fresh water Fishes and West Indian Zoogeography. *Annual Report of the Board of Regents of the Smithsonian Institution* **92**: 339-364.
- Nielsen R, Beaumont MA. 2009.** Statistical inferences in phylogeography. *Molecular Ecology* **18**: 1034–1047.
- Ramos-Fregonezi AMC, Malabarba LR, Fagundes NJR. 2017.** Population genetic structure of *Cnesterodon decemmaculatus* (Poeciliidae): a freshwater look at the Pampa Biome in southern South America. *Frontiers in Genetics* **8**: 1-10.
- Sousa VC, Beaumont MA, Fernandes P, Coelho MM, Chikhi L. 2012.** Population divergence with or without admixture: selecting models using an ABC approach. *Heredity* **108**: 521-530.
- Schönhuth S, Gagne RB, Alda F, Neely DA, Mayden RL, Blum MJ. 2018.** Phylogeography of the widespread creek chub *Semotilus atromaculatus* (Cypriniformes: Leuciscidae). *Journal of Fish Biology* **93**: 778-791.
- Tedesco PA, Paradis E, Lévêque C, Hugueny B. 2017.** Explaining global-scale diversification patterns in actinopterygian fishes. *Journal of Biogeography* **44**: 773-783.
- Thomaz AT, Knowles LL. 2018.** Flowing into the unknown: inferred paleodrainages for studying the ichthyofauna of Brazilian coastal rivers. *Neotropical Ichthyology* **16**: 1-13.
- Vari RP. 1988.** The Curimatidae, a Lowland Neotropical Fish Family (Pisces: Characiformes); Distribution, Endemism and Phylogenetic Biogeography. In: Vanzolini PE, Heyer WR, eds. *Proceedings of a workshop on neotropical distribution patterns*. Rio de Janeiro: Academia Brasileira de Ciências, 343-377.
- 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. In: Vanzolini PE, Heyer WR,

eds. *Proceedings of a workshop on neotropical distribution patterns*. Rio de Janeiro:
Academia Brasileira de Ciências, 379-427.

CONSIDERAÇÕES FINAIS

Os resultados do presente trabalho contribuem para um maior entendimento sobre como cenários alternativos para a história demográfica de peixes de água doce podem ser testados de forma quantitativa. O DNA mitocondrial demonstrou ser um marcador com potencial, embora limitado, para ser utilizado pela metodologia ABC para confrontar hipóteses de retenção de polimorfismo ancestral e de contato secundário em peixes de água doce e, portanto, elucidar a probabilidade de fenômenos como a captura de rio. No estudo de caso avaliado (Capítulo 1), foi possível mostrar que a ausência de monofilia entre as populações de *Cnesterodon decemmaculatus* das bacias do Rio Uruguai e Rio Negro, das bacias do Rio Negro e Lagoa Mirim e das bacias costeiras do Sul do Uruguai e Lagoa Mirim são melhor representadas por um cenário de contato secundário. Por outro lado, não foi possível apontar com confiança o melhor cenário para a relação entre Rio Uruguai e as bacias costeiras do sul do Uruguai demonstrando o poder limitado do mtDNA especialmente em casos onde provavelmente não há mutações ou linhagens específicas em diferentes drenagens, o que faz com os padrões de diversidade observados sejam muito semelhantes independentemente do cenário correto.

Através do estudo de simulações (Capítulo 2) foi demonstrado que o uso de muitos SNPs supera em grande parte as dificuldades apresentadas pelo uso de um único locus para distinguir entre os cenários avaliados e, portanto, a utilização de SNPs é altamente recomendada quando se planeja inferir a história de demográfica de uma espécie utilizando ABC. No entanto, existe uma maior demanda de trabalho computacional seja para montar e filtrar os dados empíricos seja para analisá-los através de ABC, além de situações que são problemáticas para esse marcador, especialmente quando a divergência entre populações é recente e, novamente, padrões semelhantes de diversidade podem ser gerados em quaisquer dos cenários concorrentes. Interessantemente, o número de amostras tomadas das populações de cada bacia

demonstrou ser uma variável de pouca importância, ou seja, o desempenho da escolha de cenários em ABC é muito similar com uma amostragem relativamente grande ou pequena, o que pode ser interessante para a montagem de desenhos experimentais que minimizem o custo por amostra/populações maximizando o número de marcadores.

Através dos resultados aqui apresentados, espera-se contribuir para um aumento no uso de metodologias que utilizem simulações e comparação entre cenários para inferir os processos que geraram os padrões genéticos observados na ictiofauna de água doce, com especial ênfase na rica diversidade Neotropical.