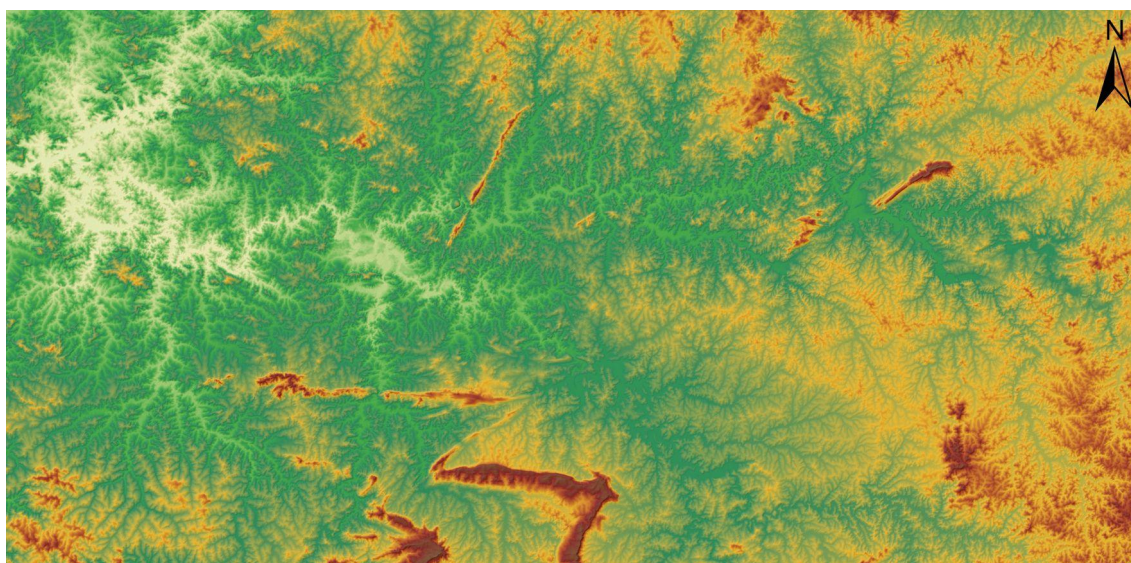


**Distribuição fractal da variabilidade de polimorfismos
em subpopulações mistas com filopatria a sítios de
reprodução segregados: evidência a partir de um
padrão de estrutura genética populacional hierárquica
em um peixe migratório Neotropical**



Gabriel M. Yazbeck

**Universidade Federal de São João Del Rei
Campus Tancredo de Almeida Neves
Departamento de Zootecnia
Laboratório de Recursos Genéticos**

**Distribuição fractal da variabilidade de polimorfismos
em subpopulações mistas com filopatria a sítios de
reprodução segregados: evidência a partir de um
padrão de estrutura genética populacional hierárquica
em um peixe migratório Neotropical**

Tese apresentada como requisito para promoção à classe de Professor Titular da
Universidade Federal de São João Del Rei

Prof. Gabriel M. Yazbeck

**São João Del Rei, MG, Brasil
Abril, 2025**

Ficha catalográfica elaborada pela Divisão de Biblioteca (DIBIB)
e Núcleo de Tecnologia da Informação (NTINF) da UFSJ,
com os dados fornecidos pelo(a) autor(a)

Y35d Yazbeck, Gabriel de Menezes.
Distribuição fractal da variabilidade de
polimorfismos em subpopulações mistas com filopatria
a sítios de reprodução segregados : Evidência a partir
de um padrão de estrutura genética populacional
hierárquica em um peixe migratório Neotropical /
Gabriel de Menezes Yazbeck ; orientador Yazbeck
Gabriel de Menezes . -- São João del-Rei, 2025.
83 p.

Tese (Doutorado - Ciências Biológicas) --
Universidade Federal de São João del-Rei, 2025.

1. Genética de populações. 2. Genômica. 3. Peixes
Neotropicais. 4. Bioinformática. 5. Tese de
Titularidade. I. Gabriel de Menezes , Yazbeck ,
orient. II. Título.



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI
DEPARTAMENTO DE ZOOTECNIA

ATA DE DEFESA DE MEMORIAL ACADÊMICO

Ao dia quinze de abril do ano de 2025, às 8 horas, sob a presidência do Prof. Pedro Manoel Galetti Junior, com a participação do Prof. Paulo De Marco Júnior, do Prof. Fabrício Rodrigues dos Santos, e do Prof. Paulo dos Santos Pompeu, reuniu-se a banca de defesa do Memorial de Tese do **Prof. Gabriel de Menezes Yazbeck**, docente do Departamento de Zootecnia da Universidade Federal de São João del-Rei, para avaliar o atendimento aos requisitos para promoção à Classe E, denominada Titular. Após apurados os resultados, o docente foi considerado **aprovado** para promoção à Classe E, Titular.

São João del-Rei, 15 de abril de 2025

Documento assinado digitalmente
gov.br PEDRO MANOEL GALETTI JUNIOR
Data: 15/04/2025 15:20:48-0300
Verifique em <https://validar.iti.gov.br>

Prof. Pedro Manoel Galetti Junior (UFSCar)
Presidente

Documento assinado digitalmente
gov.br PAULO DE MARCO JUNIOR
Data: 30/04/2025 11:31:56-0300
Verifique em <https://validar.iti.gov.br>

Prof. Paulo De Marco Júnior (UFG)
Membro Externo

Documento assinado digitalmente
gov.br FABRICIO RODRIGUES DOS SANTOS
Data: 30/04/2025 11:39:30-0300
Verifique em <https://validar.iti.gov.br>

Prof. Fabrício Rodrigues dos Santos (UFMG)
Membro Externo

Documento assinado digitalmente
gov.br PAULO DOS SANTOS POMPEU
Data: 30/04/2025 11:20:48-0300
Verifique em <https://validar.iti.gov.br>

Prof. Paulo dos Santos Pompeu (UFLA)
Membro Externo

Dedico este trabalho às mulheres da minha vida: Agata, Melissa, Carol, Elisa e Dalva.

AGRADECIMENTOS

Sinceramente, agradeço a uma imensurável cadeia humana de diferentes pessoas que me ajudaram diretamente em minha jornada, especialmente aos meus pais, às minhas filhas e minha esposa, todos os meus professore(a)s e tutor(a)s, todas as minhas alunas e meus alunos, meus colaboradores de pesquisa, além de alguns insuspeitos e importantes colegas, amigos e amigas, aqui e ao redor do mundo.

Também agradeço aos membros da banca de avaliação pelo expertise, tempo e atenção reservados a este trabalho. Agradecimentos, ainda, à todos na UFSJ nos bastidores, além da PROPE, PPGBiotec, PGE, DEZOO e PROFNIT.

Pelo apoio aos trabalhos do LARGE-UFSJ, gostaria de agradecer ao programa Peixe Vivo da Companhia Energética de Minas Gerais (CEMIG) e à Agência Nacional de Energia Elétrica (ANEEL), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e ao Laboratório de Síntese de Biomoléculas (LSB) do Centro Nacional de Pesquisa em Energia e Materiais (CNPEM).

“Todos modelos estão errados, mas alguns são úteis”

George E.P. Box

*"As nuvens não são esferas, as montanhas não são cones,
os litorais não são círculos e as cascas de árvore não são lisas, nem os raios viajam em linha reta."*

Benoit B. Mandelbrot.

Resumo

Este trabalho propõe um modelo conceitual da distribuição fractal da variabilidade genética em organismos exibindo subpopulações mistas, mas com filopatria a distintos sítios reprodutivos segregados. Eu argumento que isso surgiria sob a influência de atributos físicos da paisagem que seguem aproximadamente uma disposição geral fractal, devido a pistas ou restrições de rotas migratórias. Esta hipótese é indutivamente derivada a partir de uma observação empírica fortuita e notável, no contexto de uma espécie de peixe migratório Neotropical. Aqui eu apresento o caso de um padrão fractal aparente da estrutura populacional genética hierárquica na curimba *Prochilodus lineatus*, um peixe potamódromo amplamente distribuído e de longo alcance migratório. Os objetivos originais desta investigação são testar a presença de estoques mistos, os efeitos do isolamento por barragens e de peixamentos sobre a diversidade genética desta espécie. O trecho superior do rio Grande, MG, Brasil, extremamente fragmentado, dentro da megadiversa bacia continental do rio da Prata é focado. Eu utilizo o sequenciamento de nova-geração, para gerar ~483 milhões de leituras curtas pareadas (43 bilhões de bases de DNA), usadas para a primeira montagem *de novo* do genoma (1,4 Gb) deste peixe. Um painel *in silico* de 37.476 *loci* microsatélites é apresentado e livremente disponibilizado. Quinze microsatélites validados empiricamente são aplicados neste estudo. Dois estoques principais, geneticamente admisturados, são identificados, por meio de estimativa de máxima verossimilhança e desdobrados reiteradamente com mais duas etapas iguais, em uma análise hierárquica da estrutura genética. Esta revela, consistentemente, dois novos grupos internos, a partir dos agrupamentos previamente definidos (2³). Este padrão estratificado e autossimilar de estoques admisturados do canal principal do rio pode muito bem ser atribuído à existência de diferentes áreas de reprodução bem definidas, aninhadas em escalas diferenciais através da bacia, organizadas de acordo com a geometria fractal (rio principal, tributários de 1^a e 2^a ordem). Isso implicaria em comportamento de *homing* para essa espécie, uma possibilidade já aventada e um tema comum em peixes migratórios. Os possíveis efeitos genéticos de peixamentos e das represas também são discutidos, revelando o potencial isolamento de estoques de peixes à montante do último trecho livre do rio Grande, MG, em uma área já considerada prioritária para a conservação de peixes migratórios. Essas descobertas beneficiarão pesquisadores e gestores ambientais no enfrentamento de questões desafiadoras, pois apontam para o papel potencial de tributários de diferentes ordens como abrigo reprodutivo populações geneticamente diversas. Esses resultados têm implicações potenciais para o gerenciamento de pesca de águas interiores e para a conservação de vários peixes migratórios Neotropicais negligenciados da América do Sul. Possivelmente, este padrão fractal da diversidade genética poderá ser confirmado em outros peixes, organismos migratórios aquáticos e não-aquáticos, que mostrem uma mistura temporária, mas com locais de reprodução independentes e bem definidos.

Palavras-chave: Alto rio Grande MG, Bioinformática, Microsatélites, Genética de populações, Genômica, Sequenciamento por síntese,

Abstract

This work purposes a conceptual model for the fractal distribution of the genetic variability in organisms exhibiting mixed subpopulations, but with distinct segregated philopatric breeding grounds. I argue this would arise under the influence of physical landscape attributes that roughly follow a general fractal disposition, due to migratory cues or route restrictions. This hypothesis was inductively derived from a fortuitous and remarkable empirical observation, in the context of a Neotropical migratory fish species. Here, I present the case for an apparent fractal pattern of the hierarchical genetic population structure in curimba *Prochilodus lineatus*, a widespread and long-range potamodromous fish. The original aims of this investigation is to test for the presence of mixed stocks, the effects of isolation by dams and broodstocking over the genetic diversity for this species. The upper segment of the extremely fragmented Grande River, MG, Brazil, from the megadiverse continental La Plata Basin is focused. Next-generation sequencing was used, delivering ~483 million paired-end short-reads (43 billion DNA bases) and employed in the first *de novo* genome assembly (1.4 Gb) for this fish. An *in silico* panel of 37,476 microsatellite *loci* is disclosed and hereby made freely available. Fifteen empirically validated microsatellites are employed for this study. Two main co-occurring genetically admixed stocks were identified by maximum likelihood estimation and iteratively unfolded with two more equal steps, through a hierarchical analysis of genetic structure. It consistently unveils two newly defined internal groups, from the previously resolved clusters (2³). This stratified, auto-similar pattern of mixed stocks from the main river channel may very well be attributed to the existence of different, well defined breeding grounds, nested within differential scales throughout this basin, arranged according to fractal geometry (main river, 1st and 2nd order tributaries). This would imply in homing behavior for this species, a possibility already raised a common theme in migratory fishes. The possible genetic effects of broodstock release and dams are also discussed, revealing potential isolation of fish stocks upstream the last free-flowing stretch of the Grande River, MG, in an area already deemed a priority for conservation of migratory fishes. These findings will benefit researchers and environmental managers in tackling challenging issues, as they point to the potential role of tributaries of different orders in harboring the spawning of genetically diverse fish populations. These results have potential implications for inland fisheries management and the conservation of several overlooked Neotropical migratory fishes from South America. Possibly this fractal pattern of genetic diversity can be verified in other migratory fishes and maybe other non-aquatic migratory organisms showing a temporary mixture but independent, well defined breeding sites.

Key words: Bioinformatics, Genomics, Microsatellites, Population genetics, Sequencing by synthesis, Upper Grande River MG

Sumário

Introdução.....	11
Fractais.....	11
Peixes migratórios de água doce Neotropicais: <i>Prochilodus lineatus</i>	14
Marcadores moleculares em peixes migratórios Neotropicais.....	16
Microssatélites de DNA.....	18
Admistura.....	19
Estrutura genética populacional em <i>Prochilodus lineatus</i>	21
Rio Grande, MG e recursos genômicos em <i>Prochilodus</i>	22
Objetivos.....	23
Material e Métodos.....	24
Genômica e bioinformática.....	24
Desenvolvimento e validação de marcadores moleculares.....	26
Genética de populações.....	27
Análise hierárquica da estrutura genética populacional.....	29
Declarações éticas.....	32
Resultados.....	33
Genômica e bioinformática.....	33
Desenvolvimento e validação de marcadores moleculares.....	35
Genética de populações.....	36
Análise hierárquica da estrutura genética populacional.....	38
Discussão.....	43
Genômica e bioinformática.....	43
Desenvolvimento e validação de marcadores moleculares.....	46
Genética de populações.....	48
Análise hierárquica da estrutura genética populacional.....	51
Estrutura genética populacional fractal.....	54
Considerações finais.....	56
Conclusão.....	57
Referências bibliográficas.....	58
Apêndices.....	76
Apêndice A – Arquivos suplementares externos.....	76
Apêndice B - Tabelas Suplementares.....	78
Apêndice C – Figuras suplementares.....	81

Introdução

Fractais

A geometria fractal foi descrita como sendo a ponte entre o simples e o complexo e sua análise foi considerada “uma rebelião contra o cálculo” (3Blue1Brown, 2017). O estudo de fractais foi motivado tanto pela abordagem de problemas puramente matemáticos, como definir o perímetro de uma forma simples e autorrepetitiva sob recursão (*e.g.* Figura 1), quanto por problemas aplicados, como na comparação de formas com curvas extremamente irregulares ou interrompidas (ou seja, não-contínuas, fraturadas), através de uma medida de seu grau contido de aspereza (rugosidade), por exemplo, nos litorais da Grã-Bretanha ou da Noruega (Figura 2).

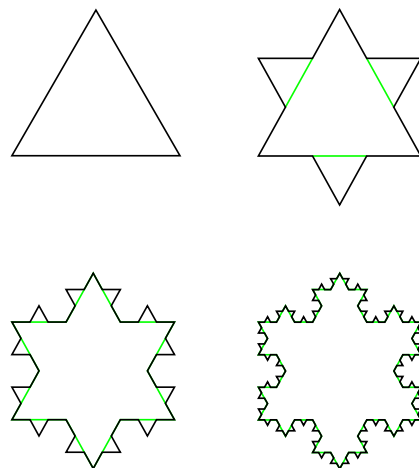


Figura 1: O floco de neve de Koch, um exemplo de fractal simples que surge da projeção da curva angular fundamental, autocontida em um triângulo equilátero, como um padrão recorrente (autossimilaridade), aumentando-se reiteradamente a aspereza de seus limites em um terço, a cada passo iterativo. Observe que, à medida com a qual as iterações continuam *ad nauseam*, o perímetro tenderá a uma medição linear infinita. A curva fechada, porém, conterá uma área estritamente finita (Fonte: *Wikicommons*, 2025, sob licença *Creative Commons BY-SA 3.0*).

Em seu estudo original de objetos fractais, Benoît B. Mandelbrot (o “B.” significa “Benoît B. Mandelbrot”) argumenta sobre as limitações de uma definição geral, precisa e simples para fractais, mas estabelece que o principal atributo de um objeto fractal estaria em sua dimensionalidade fractal D ou fractalidade (Mandelbrot, 1975, p. 1). Por exemplo, considerando

dimensões algébricas, poderíamos dizer que uma dada sequência de DNA tem 1 dimensão, a estrutura secundária planar em forma de trevo de um RNA de transferência (tRNA) tem 2 dimensões e que sua estrutura topológica espacial em forma de L invertido tem 3 dimensões. As formas fractais, no entanto, manifestam dimensões que não são medidas por números inteiros, mas por números reais (dimensões de Housedorff), que expressam a frequência de autossimilaridade e irregularidades da forma, através das escalas, uma medida de seu grau de aspereza. É importante notar que o uso de escalas diferentes produzirá medidas alternativas para o mesmo objeto (“a arbitrariedade dos resultados da medida” - Mandelbrot, 1983, p. 26) .

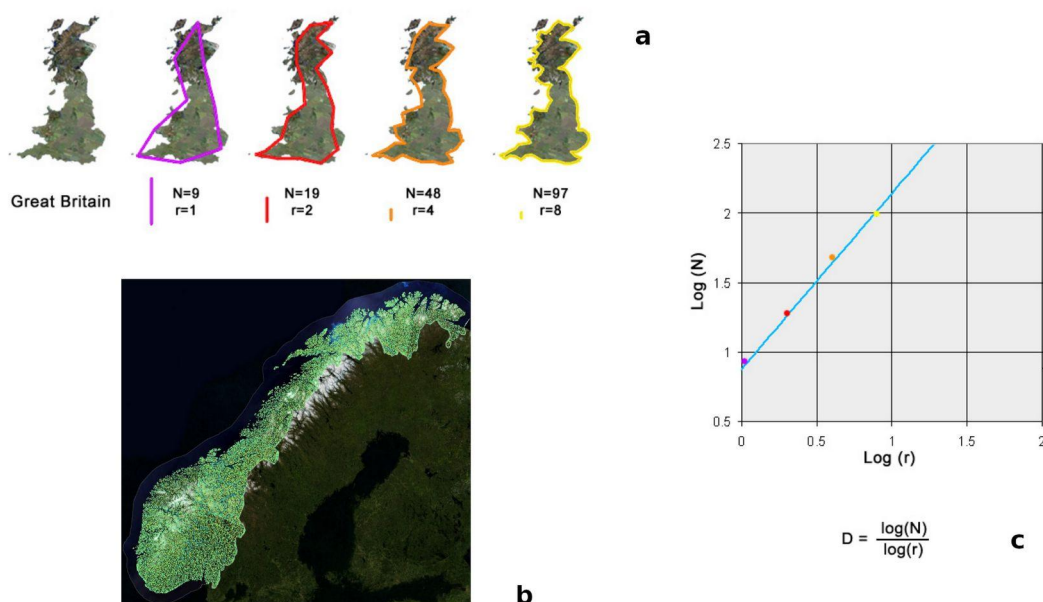


Figura 2: Dimensão fractal D . **a)** A medida do número de “partes” N usadas para se medir o litoral da Grã-Bretanha depende da escala de ampliação da régua r usada. **b)** O litoral da Noruega tem um valor D mais alto ($\sim 1,52$), quando comparado ao da Grã-Bretanha (Imagem: ESA - *Copernicus Sentinel-1*). **c)** A dimensão fractal D é igual à inclinação da curva (a taxa com a qual o perímetro muda de acordo com o fator de magnificação r , da escala utilizada) – Modificado de fractalfoundation.org.

A Figura 2 mostra que o uso de fatores de ampliação (magnificação) r cada vez maiores (inversamente proporcional à escala de comprimento decrescente das respectivas régua) resulta em valores mais e mais altos da medida N do perímetro. Neste contexto, podemos definir a dimensão fractal D como $\log(N)/\log(r)$. A dimensão fractal D do perímetro da Grã-Bretanha é $\sim 1,21$, próximo

à dimensão fractal D do floco de neve de Koch da Figura 1, que é de aproximadamente 1,2618 (Mandelbrot, 1983, p. 36).

O valor D , portanto, fornece um índice estatístico racional do nível de detalhes e complexidade em um determinado padrão. Como entendemos, do ponto de vista estatístico, por aumento de precisão, a minimização gradual entre os limites superiores e inferiores do erro de uma medida, não podemos dizer que o incremento reiterado de magnificação implicará em aumento da precisão, uma vez que a medida obtida com o uso de réguas cada vez menores irá, necessariamente, implicar em um aumento incremental dos valores de N computados. Ou seja, o valor da medida só aumenta com a ampliação, mas não a precisão, pois não há convergência em torno de um valor. Isso é denominado paradoxo do litoral (Richardson, 1961).

Aqui, porém, definirei uma forma fractal simplesmente como sendo composta de padrões aproximadamente repetitivos (autossimilaridade), através de escalas diferenciais, com ênfase na paisagem geográfica natural. Este tema está presente de forma nítida nos principais componentes do meio abiótico, como nas cadeias de montanhas, limites de habitats, bacias hidrográficas, etc., além de litorais (Mandelbrot, 1983; Xu *et al.*, 1993; Andronache *et al.*, 2019).

Estruturas fractais também são prevalentes na natureza viva, desde possíveis arranjos de proteínas (Sendker *et al.*, 2024), ou na ultraestrutura do DNA (Bancaud *et al.*, 2012; Anitas, 2022) e arranjos celulares (*e.g.* Legget *et al.*, 2019), até na morfologia geral de árvores brônquicas, sistemas vasculares, penas de aves e inúmeros outros exemplos (*e.g.* Bassingthwaighth, 1994; Meyer *et al.*, 2020).

Peixes migratórios de água doce Neotropicais: *Prochilodus lineatus*

A fauna de peixes de água doce Neotropical está entre os grupos mais diversos de vertebrados não-marinhos (Tagliacollo *et al.*, 2024). A curimba *Prochilodus lineatus* Valenciennes 1836 (Characiformes: Prochilodontidae), também conhecida como sabálo, curimbatá, curimatã, grumatã, grumatá, papa-terra (entre outros nomes comuns), é uma das espécies migratórias de água doce mais importantes para a pesca continental na América do Sul, representando mais de 50% da biomassa de pescado em partes da grande bacia do rio da Prata (Castro e Vari, 2003; Bowen, 2022a,b), que reúne as bacias dos rios Paraná, Paraguai e Uruguai. Este enorme sistema hídrico é o segundo maior do hemisfério ocidental, possivelmente o quinto maior do mundo e alcança cinco países: Argentina; Bolívia; Brasil; Paraguai; e Uruguai. Estima-se que cerca de 110 milhões de pessoas sejam diretamente beneficiadas pelos seus serviços ambientais (Espindola *et al.*, 2021). Populações de *P. lineatus* também ocorrem naturalmente na bacia do rio Paraíba do Sul, e 12 outras espécies congêneres ocupam outras grandes bacias hidrográficas Neotropicais distintas (Melo *et al.*, 2018), revelando uma história evolutiva complexa (Santos *et al.*, 2021, Frable *et al.*, 2022). *Prochilodus lineatus* também foi registrada como uma espécie invasora, uma questão de conservação global cada vez mais premente (Xu *et al.*, 2024), sendo introduzida na China (Endruweit, 2014) e no Vietnã (Kalous *et al.*, 2012).

Prochilodus lineatus apresenta uma boca conspícua e especializada em forma de funil invertido (Figura 3), uma adaptação de seus hábitos iliófagos (*i.e.* detritívoros - Bowen, 2022b). A espécie é um exemplo Neotropical representativo da história de vida potamódroma (*i.e.* de migração exclusiva em água doce). Sua estratégia reprodutiva foi primeiramente estudada sistematicamente por Manuel Pereira de Godoy (*e.g.* Godoy, 1970) e envolve migrações anuais notoriamente longas (de centenas a milhares de quilômetros), desencadeadas por gatilhos sazonais ambientais. Encontros de volumosos cardumes acontecem antes do deslocamento massivo de peixes, rio acima,

denominado de piracema (nome também usado para caracterizar o conjunto de diferentes espécies de peixes migratórios). A piracema começa aproximadamente no final da primavera e no início do verão, por volta de dezembro, durando normalmente até março (Carolsfeld *et al.*, 2003) sendo, portanto, um processo sensível diante das mudanças climáticas em curso (Comte e Olden, 2017; Alix *et al.*, 2020, Ilosvay *et al.*, 2022).

As subidas migratórias potencialmente colocam indivíduos de diferentes subpopulações de reprodução aleatória (“demes”) bem definidas no mesmo estoque local de peixes, ao longo de rotas compartilhadas (Collins *et al.*, 2013). Os peixes piracema desovam naturalmente ao atingir sítios de reprodução à montante (*e.g.* Riboli *et al.*, 2020). Ovos, larvas e alevinos, então, derivam passivamente à jusante, até alcançar baixadas e lagoas marginais, onde crescem e amadurecem (Pettry *et al.*, 2003). Adultos desovam vários anos consecutivos durante seu ciclo de vida (Carolsfeld *et al.*, 2003).

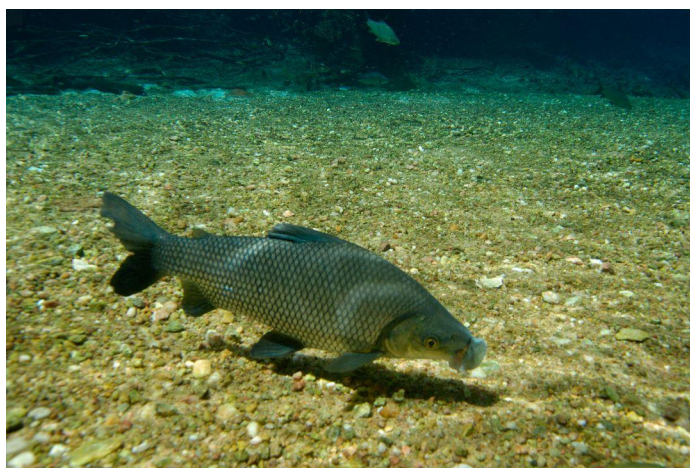


Figura 3: A curimba *Prochilodus lineatus* da bacia do rio da Prata em seu ambiente natural, Bonito, MS, Brasil (Fotografia gentilmente cedida por André Seale).

Espécies do gênero *Prochilodus* estiveram dentre os primeiros peixes migratórios Neotropicais a ter dominada sua desova *ex situ*, motivada principalmente pela exploração intensiva da geração de energia hidrelétrica e seu impacto direto sobre os peixes de piracema nas regiões afetadas. O principal desafio para a reprodução *ex situ* está na patente ausência de estímulos ambientais naturais nas pisciculturas. Isso tem sido normalmente contornado por meio de indução

hormonal, relatada desde pelo menos 1946, de acordo com Fenerich-Verani *et al.* (1984), e usada para dar suporte a operações de soltura de peixes produzidos em piscicultura (peixamentos), visando neutralizar a diminuição da pesca em torno de regiões impactadas por represas.

Essas práticas de estocagem artificial de peixes foram amplamente abandonadas em alguns lugares relevantes (*e.g.* Minas Gerais), durante a última década, em parte devido ao relaxamento geral das normas orientadas à proteção ambiental (Zellhuber, 2016), embora ainda falte uma avaliação rigorosa, sistemática e holística (ou seja, ecológica, genética e antropológica) dos potenciais benefícios ou problemas da atividade. Dado seu status histórico como o principal alvo para práticas de piscicultura e peixamentos (*e.g.* Saraiva e Pompeu, 2016), nos últimos 50 anos ou mais, *P. lineatus* é uma espécie candidata natural a organismo-modelo para a investigação dos impactos de barragens e a eficiência de possíveis medidas de mitigação, em relação aos peixes migratórios Neotropicais em geral.

Marcadores moleculares em peixes migratórios Neotropicais

O uso de marcadores moleculares e citogenéticos para a compreensão da biologia e variação em peixes migratórios de água doce Neotropicais têm revelado diferentes níveis de estrutura genética (Piorski *et al.*, 2008; Oliveira *et al.*, 2009). Alguns estudos apontam a ocorrência de isolamento por tempo entre estoques capturados nas mesmas localidades (indicando diferentes subpopulações ao longo do ano, *e.g.* Braga-Silva e Galetti Jr., 2016; Ribolli *et al.*, 2017) e diversas classes de marcadores moleculares têm sido aplicadas na elucidação da influência de barragens e peixamentos (*e.g.*, Hatanaka *et al.*, 2006; Landínez-García *et al.*, 2020; Pimentel *et al.*, 2020; Oliveira-Farias *et al.*, 2022; García-Castro e Márquez, 2024), ou para acessar a estrutura e delimitação local ou regional de estoques (*e.g.*, Sivasundar *et al.*, 2001; Collins *et al.*, 2013; Landínez-García e Márquez, 2018). Alguns estudos com marcadores genéticos apontaram para a ocorrência de comportamento reprodutivo de *homing* (ou seja, reprodução filopátrica, a tendência

de indivíduos regressarem aos sítios de origem) em espécies de peixes-de-couro da Amazônia, como *Brachyplatystoma rousseauxii* (e.g., Batista e Alves-Gomes, 2006; Carvajal-Vallejos *et al.*, 2014; Barthem *et al.*, 2017) e da bacia do Prata, como *Pseudoplatystoma corruscans* (Pereira *et al.*, 2009).

Questões específicas que merecem uma avaliação mais aprofundada na América do Sul incluem a taxa e a intensidade do isolamento genético por barragens para peixes (e.g. Ruzich *et al.*, 2019); a eficiência de mecanismos de transposição na sustentação do fluxo gênico (e.g. Wilkes *et al.*, 2019); se as operações de repovoamento contribuem para a sustentabilidade da biomassa pesqueira (e.g. Osathanunkul e Suwannapoom, 2023); se há efeitos genéticos relevantes das práticas de piscicultura, como gargalos ou domesticação de estoques (e.g. Howe *et al.*, 2024); se os peixamentos contribuem para efeitos importantes de introgressão e alteração (e.g. Chang *et al.*, 2022), ou se estão comprometendo adaptações locais (e.g. Shedd *et al.*, 2022), etc.

Para se abordar essas questões de forma segura e informativa, são necessários planejamento adequado e consideração cuidadosa da complexidade destes sistemas. Também é necessária a disponibilidade suficiente de ferramentas práticas e metodológicas, como número e variedade adequados de diferentes marcadores moleculares, além de modelos conceituais/quantitativos eficientes para a descrição de fenômenos de genética e ecologia populacional dentro do contexto dos impactos de geração de energia hidrelétrica (e.g. Telles *et al.*, 2011; Baggio *et al.*, 2018; Davis *et al.*, 2018; Nguyen *et al.*, 2018).

Estratégias de sequenciamento de DNA de segunda e terceira geração (Shendure *et al.*, 2017), também conhecidas como sequenciamento de próxima geração (NGS), podem ajudar a promover o desembaraço sistemático dessas questões desafiadoras e concorrentes. NGS permite, por exemplo, a caracterização completa do genoma e abordagem de ressequenciamento, genotipagem de alto rendimento por sequenciamento de milhares de amostras ou criando pequenas janelas diretas de dados genômicos, permitindo a detecção rápida de uma ampla paleta de diferentes

classes de marcadores moleculares (Bernos *et al.*, 2020), aplicáveis à investigação abrangente da delimitação de estoques naturais (*e.g.* Layton *et al.*, 2020). Essas informações são particularmente úteis para a piscicultura, gerenciamento da pesca, práticas de gestão ambiental (Allendorf, 2017) e tomada de decisões para o desenvolvimento do uso de recursos hídricos.

Microssatélites de DNA

Peixes migratórios sob impacto da produção hidrelétrica podem se beneficiar de medidas guiadas pelas informações objetivas levantadas por meio da aplicação de marcadores moleculares, como os microssatélites. Esses são *loci* genéticos abundantes e espalhados pelos genomas, caracterizados por temas de sequência de DNA curtas e repetitivas e que desempenham uma diversidade de papéis funcionais e estruturais nos cromossomos (Srivastava *et al.*, 2019; Lei *et al.*, 2021).

Como os *loci* genéticos que carregam essas sequências têm a propensão a variar em número de repetições dentro e entre populações (alta taxa de mutação) e devido à sua relativa facilidade de uso para registrar polimorfismos genéticos (variação do comprimento do fragmento de DNA), os microssatélites têm sido amplamente aplicados como marcadores moleculares para estudos de diversidade genética (Schlötterer, 2004; Budowle e Sajantila, 2024).

Um estudo recente de McComish *et al.* (2024) lançou luz sobre a dinâmica de mutação de microssatélites de vertebrados, revelando estabilidade evolutiva na distribuição da variação do comprimento de fragmentos, de milhares a milhões de anos, nesses *loci*. Eles sugerem que, em termos de mutação, alelos mais curtos tendem a se expandir, enquanto alelos mais longos tendem a diminuir em número de repetições (possível seleção estabilizadora), destacando também a provável associação da variação do comprimento de microssatélites e rápida mudança fenotípica (McComish *et al.*, 2024). Isso é importante, pois muitas aplicações, como estimar cenários demográficos e inferir neutralidade seletiva, podem ser problemáticas se assumirmos modelos de mutação imprecisos para microssatélites (Ellegren, 2004; Putman e Carbone, 2016).

De forma aplicada importante, os marcadores de DNA de microssatélites são passíveis de rotinas de genotipagem relativamente simples, adequadas para operações diárias de monitoramento de pisciculturas ou peixes selvagens, o que viabiliza aplicações práticas na produção de peixes para fins ambientais ou comerciais, que necessitem de informações sobre a variação genética (Wenne, 2023).

Um total de 18 *loci* microssatélites específicos validados para *P. lineatus* estão atualmente disponíveis, cinco marcadores originalmente descritos em Yazbeck e Kalapothakis (2007) e mais 13 apresentados em Rueda *et al.* (2011). Isso é um número limitante para cobrir um genoma segmentado em pelo menos 27 cromossomos (há pelo menos nove cromossomos supranumerários já descritos), com $2N=54$ (Pauls e Bertollo, 1983; Stornioli *et al.*, 2021).

Admistura

Desde o trabalho seminal de Smouse *et al.* (1990), cientistas e gestores têm buscado a aplicação da inferência estatística da estrutura genética para caracterizar a presença de estoques mistos em amostras de peixes, por meio do uso de marcadores de DNA (Hallerman, 2003; Valenzuela-Quíñonez, 2016). Isso é útil quando indivíduos capturados no mesmo momento e local pertencem a diferentes populações de origem ancestral (Pritchard *et al.*, 2000), que podem ser detectadas analiticamente por meio de modelos da diversidade genética, por meio de inferência Bayesiana (Beaumont e Rannala, 2004) ou estimativa de máxima verossimilhança (Wang, 2003) de admistura e estrutura genética da população.

Estoques mistos de peixes podem seguir um modelo populacional de mistura genética, onde cada genoma (cada peixe) é originário exclusivamente de uma única fonte populacional ancestral (subpopulação), dentre um número K de alternativas possíveis, e onde esses diferentes demes não compartilham eventos reprodutivos. Isso poderia ser teoricamente esperado em, digamos, uma amostra populacional de espécimes, de uma dada espécie hipotética de salmão migratório anádromo absolutamente filopátrica, amostrados no mar, pouco antes de sua migração reprodutiva, em rios

segregados distintos. Essa amostra de genótipos de peixes teria indivíduos cujo genoma descende exclusivamente de uma das possíveis fontes genéticas ancestrais, porque cada população verdadeira (*i.e.* subpopulação ou deme) está dentro das diferentes bacias de água doce isoladas, utilizadas como sítios de reprodução, e dado seu comportamento de retorno ao sítio natal.

Outra situação seria ilustrada pela presença de estoques mistos que seguem um modelo de admistura genética (*i.e.* miscigenação). Aqui, o genoma de cada peixe pode conter diferentes proporções de herança genética ancestral, a partir de demes distintos, onde gerações anteriores potencialmente participaram de eventos reprodutivos comuns (Pritchard *et al.*, 2000, Wang, 2022).

O modelo de admistura visa caracterizar grupos de genótipos, a partir da amostra total, que tendem a um comportamento matemático esperado (equilíbrio de Hardy-Weinberg ou EHW) diante de panmixia - reprodução aleatória - com base na maximização da variância entre os *clusters*. O modelo assume ausência de ligação entre os *loci*. O procedimento, portanto, permite a designação de genótipos amostrados a um dos K possíveis agrupamentos genéticos definidos (*clusters*). O modelo de admistura, portanto, fornece uma moldura teórica conveniente para a investigação da estrutura genética de peixes migratórios Neotropicais, sob o impacto da implementação de barragens e décadas de operações de peixamento, em grande parte conduzidas sem considerações explícitas sobre questões genéticas.

Uma ampla discussão e diferentes ferramentas computacionais (e boas práticas de uso) têm sido apresentadas sobre a inferência da estrutura genética de populações e admistura, bem como seus modelos relacionados, desde a implementação fundamental de Pritchard *et al.* (2000) de seu *software* STRUCTURE (*e.g.* Falush *et al.*, 2003; Wang, 2003; Evanno *et al.*, 2005; Falush *et al.*, 2007; Alexander *et al.*, 2009; Verity and Nichols, 2016; Janes *et al.*, 2017; Lawson *et al.*, 2018; Liu *et al.*, 2020; Wang, 2022).

Estrutura genética populacional em *Prochilodus lineatus*

Em relação a *P. lineatus*, investigações de estrutura genética que aplicaram microssatélites para verificação da admistura foram conduzidas exclusivamente por meio de inferência Bayesiana. O primeiro estudo deste tipo foi apresentado em Rueda *et al.* (2013), contando com 12 *loci* e revelou evidências genéticas para estoques distintos, sazonalmente segregados, no mesmo local do baixo rio Uruguai. Esses autores concluíram que essa espécie deve ser tratada como localmente exibindo estoques mistos para fins de manejo pesqueiro. Perini *et al.* (2021) revelaram uma dinâmica genética fonte-dreno de metapopulações, envolvendo outro tributário importante, o rio Pardo, que alimenta o baixo rio Grande, aplicando cinco microssatélites específicos da espécie e cinco heterólogos (*i.e.* isolado originalmente em espécie próxima).

Ambos, Ferreira *et al.* (2017) e Rosa *et al.* (2022) também combinaram dados de sequência de haplótipos mitocondriais (Alça D), com sete microssatélites, para a investigação de polimorfismos genéticos. Aqueles estudaram peixes da bacia do médio e alto rio Paraná e estes, incluindo um marcador heterólogo, avaliou a estrutura genética temporal de uma localização específica do rio Mogi-Guaçu (bacia do rio Pardo). A mesma quantidade de marcadores foi ainda aplicada por Ferreira *et al.* (2023) para investigar a estrutura genética de *P. lineatus* da bacia do rio Paranapanema, um importante tributário do alto rio Paraná, visando trechos livres de rios ao redor do canal principal, interrompido por usinas hidrelétricas.

Em comum, esses estudos se referem a regiões mais baixas desse sistema intensamente fragmentado por barragens. Seus resultados apontam para altos níveis de variabilidade genética, comumente relatando a ausência de estrutura populacional associada a localidades distintas, revelando evidências de subestruturação genética local ou temporal, resolvendo dois ou três *clusters* genéticos. O presente trabalho visa a análises de estrutura populacional genética para *P. lineatus* do alto rio Grande, a região terminal mais alta do flanco oriental da bacia do rio Paraná, dentro da vasta bacia do rio da Prata, por meio de abordagem alternativa ao método Bayesiano.

Rio Grande, MG e recursos genômicos em *Prochilodus*

O rio Grande nasce a uma altitude aproximadamente 2.000 m em Minas Gerais e corre por 1.360 km até sua confluência com o rio Paranaíba, onde ambos coalescem formando o rio Paraná (um importante componente da megadiversa bacia do rio da Prata - Casemiro *et al.*, 2023). O rio Grande um exemplo ilustrativo de rio intensamente represado, com 12 grandes barragens, sendo a usina hidrelétrica de Camargos, a última barragem à montante, instalada por volta de 1960. Pouco abaixo, à jusante, está a barragem da usina de Itutinga, operacional desde 1955 e que delimita um reservatório relativamente pequeno e fechado (sem tributários), a “fio d’água”. Lá, uma estação de piscicultura ambiental operou ativamente, durante décadas, na produção de peixes para estocagem, principalmente de *P. lineatus*, através da região, até o entorno de 2018. Em seguida, mais a jusante, fica a barragem da usina hidrelétrica do Funil, relativamente nova (construída por volta do ano 2000), sendo esta a única com um mecanismo de transposição de peixes, um elevador (Suzuki *et al.*, 2011), com eficiência possivelmente limitada para a passagem de *P. lineatus* (observação pessoal).

O grau com que essas barragens interferem na estrutura genética de peixes migratórios de água doce no alto rio Grande e quais eventuais problemas ou benefícios, do ponto de vista genético, resultam das operações de peixamento e transposição de peixes ali, ainda permanecem em grande parte desconhecidos.

Assim, valer-se de tecnologias NGS e disponibilizar recursos genômicos robustos e novos marcadores moleculares é relevante para facilitar e incrementar estudos ecológicos e iniciativas de conservação de peixes migratórios em ecossistemas impactados. Esse é um passo necessário para se alcançar a pesca sustentável e o gerenciamento de recursos hídricos na bacia do Prata, intensamente alterada pela atividade humana (Agostinho *et al.*, 2007), particularmente pela geração de energia hidrelétrica, uso da terra, pesca irregular, poluição industrial, urbana e rural, além da introdução de espécies invasoras. (Reis *et al.*, 2016).

Podemos enumerar nos dedos de uma mão a quantidade de recursos genômicos disponíveis atualmente para espécies de peixes do gênero *Prochilodus*. A primeira sequência completa de DNA mitocondrial para *P. lineatus* foi revelada por Carmo *et al.* (2016), enquanto outros mitogenomas de outras espécies de congêneres foram apresentados por Chagas *et al.* (2016) e Santos *et al.* (2021). Em relação ao genoma nuclear, Stornioli *et al.* (2021) divulgaram 3,6 Gb (gigabases) de dados FASTAQ de sequenciamento de DNA de segunda geração (leituras curtas, associadas a informações de qualidade, base a base), no contexto do mapeamento de DNA satélite para *P. lineatus*. Mais recentemente, Yepes-Blandón *et al.* (2023), por sua vez, depositaram o primeiro rascunho de um genoma de referência de alta contiguidade para *Prochilodus magdalenae*, uma espécie colombiana, resultando em uma montagem *de novo* com ~1,3 Gb (com a mediana do tamanho de fragmentos obtidos, ou N₅₀ de 348.313). Ali também foi fornecida a anotação para 34.724 genes potenciais naquela espécie. No entanto, o conjunto de dados de leituras curtas e longas de DNA utilizadas nesta montagem não foram disponibilizados em bancos de dados públicos.

Objetivos

O objetivo original deste trabalho foi produzir novos recursos genômicos para *Prochilodus lineatus*, um organismo-modelo para peixes migratórios Neotropicais de água doce, e aplicá-los na análise da diversidade genética de peixes separados pelas três últimas barragens a montante do alto rio Grande, MG, Brasil.

De forma particular e original, visei os seguintes problemas:

- Testar a hipótese da ocorrência de estoques mistos entre populações de *P. lineatus* segregadas pelas três últimas barragens à montante do alto rio Grande;
- Testar a hipótese de que as barragens têm isolando geneticamente peixes migratórios na região; e

- Testar a hipótese de que práticas de peixamentos com peixes de piscicultura alteraram a diversidade genética em populações de peixes aprisionadas entre barragens.

Para tanto, técnicas NGS foram aplicadas em *P. lineatus* para:

- A realização da montagem do genoma deste peixe;
- A busca de novos *loci* potenciais para o desenvolvimento rápido e de baixo custo de novos marcadores microsatélites nesta espécie; e
- Disponibilizar publicamente os novos recursos genômicos produzidos em *P. lineatus*.

Material e Métodos

Genômica e bioinformática

Bibliotecas genômicas foram construídas a partir de um único espécime, adulto, fêmea coletado na Estação da Piscicultura de Itutinga (21°19'20.7"S 44°36'56.9"W) em 2014. Este indivíduo é mantido como um voucher no LARGE-UFSJ (www.ufsj.edu.br/large – registro 140000). O DNA genômico completo foi extraído do tecido muscular com o kit de extração Wizard DNA (Promega) e verificado quanto à integridade e qualidade por eletroforese em gel de agarose. Um prestador de serviço (PS) externo avaliou a pureza da extração por espectrofotometria e quantificou o material por fluorometria. Essa extração de DNA genômica foi empregada pelo PS na produção de três bibliotecas genômicas integrais primárias (*whole genome shotgun*), com seleção de fragmentos de 500 pb, submetidos ao sequenciamento-por-síntese (Illumina). Foram geradas leituras pareadas de DNA (PE) de 90 bases em um equipamento Illumina HiSeq 2000 por PS, visando uma profundidade (redundância) de sequenciamento estimada de ~15 vezes (15X) em relação ao genoma modelo de *Danio rerio*. As leituras PE resultantes foram filtradas para remoção de sequências adaptadoras e quando continham mais de 50% de bases com valor $Q \leq 5$ na escala Phred. Os dados de sequências curtas pareadas foram então respectivamente colocados em três pares de arquivos FASTQ gêmeos, *.fq1 e *.fq2. O SP realizou uma montagem genômica *de novo*

(não disponibilizada), busca de microssatélites usando SSRIT (Temnykh *et al.*, 2001) e design de primer com Primer3 (Untergasser *et al.*, 2012).

Verifiquei a qualidade das leituras curtas de DNA resultantes usando FastQC (Andrews, 2010) via MultiQC (Ewels *et al.*, 2016). Subsequências de DNA de comprimento k fixo (k-mero), das leituras curtas FASTQ, foram computadas para os 21-meros com o auxílio do programa Jellyfish (Marçais e Kingford, 2011), que eu usei para para estimar o tamanho esperado do genoma, usando um histograma de frequência, no ambiente computacional R (versão 4.4.1 – R Core Team).

Em colaboração com o Prof. Rafael Sachetto Oliveira (DECOMP-UFSJ) usamos essas leituras curtas, junto com um conjunto contigs (fasta) resolvidos pelo SP, para executar uma montagem genômica *de novo* usando MEGAHIT (Li *et al.*, 2015), com k-mero=47 para resolver novos *contigs* e, em seguida, submeter os resultados para o módulo SOAPdenovo-fusion do SOAPdenovo2 (Luo *et al.*, 2012), para resolver *scaffolds*. Esta montagem foi processada no *cluster* de computação de alto desempenho Sagarana (CEPAD-ICB-UFMG, Belo Horizonte, Brasil).

A ferramenta de triagem e filtragem de contaminação FCS (Astashyn *et al.*, 2024) do *National Center for Biotechnology Information* (NCBI – Sayers *et al.*, 2022) foi usada por membros da equipe do NCBI e por mim para filtrar a montagem mitocondrial, sequências de microrganismos e adaptadores da biblioteca encontrados na sequência final no nível de *scaffold*. A seleção da montagem genômica para uso posterior foi baseada principalmente no tamanho e no N₅₀ da montagem. As estatísticas de montagem foram obtidas por meio do assemblyStats (<http://gif.biotech.iastate.edu>) e sua visualização, pelo assembly-stats 17.02 (Chalis, 2017).

Para mapear as dezenas de milhares de *loci* potencias de microssatélites preditos pelo SP na nova montagem do genoma, o seguinte ensaio *in silico* foi executado, com o auxílio do Prof. Sachetto: realizamos uma busca local BLAST (Altschul *et al.*, 1990) na nova montagem selecionada, usando as sequências de produtos de PCR preditas na busca de microssatélites. Quando estes tiveram resultados de busca BLAST nulos, usamos as sequências de seus pares de *primers*

para buscas por algoritmo de Smith-Waterman, usando SWIPE (Rognes, 2011) e retendo resultados com correspondência de ambos, *primers* direto e reverso, no mesmo contig ou *scaffold*. Um mapeamento de alinhamento das sequências curtas (SAM) foi feito sobre a nova montagem genômica com BBmap (Bushnell, 2014) e, em seguida, convertido em um arquivo binário mais compacto (BAM) com SAMtools (Li *et al.*, 2009).

Desenvolvimento e validação de marcadores moleculares

A partir da caracterização do painel genômico dos *loci* microssatélites potenciais de *P. lineatus*, candidatos a novos marcadores foram selecionados para desenvolvimento empírico, a partir da escolha dos 12 primeiros resultados da classificação decrescente do tamanho do tema repetitivo e número do número de repetições resolvidas, a fim de aumentar as chances de encontrar *loci* polimórficos (Brandström e Ellegren, 2008; Srivastava *et al.*, 2019), além escolha aleatória de outros 10 candidatos.

A descrição dos passos de validação empírica e aplicação populacional dos novos microssatélites estão melhor detalhados em Nascimento (2024). Em suma, 22 pares de *primers* candidatos foram testados por duas discentes de pós-graduação do LARGE-UFSJ (PGE e PPBiotec-UFSJ). Inicialmente, foram aplicadas quatro amostras individuais de DNA em reações em cadeia da polimerase (PCR), com volume final de 10 µl, com 10 pmols de cada *primer* e usando *Taq* DNA polimerase (Phoneutria, Belo Horizonte). Os programas foram padronizados para 30 ciclos de termociclagem.

Os produtos de PCR resultantes foram visualizados após eletroforese 15 V/cm, em gel de poliacrilamida (PAGE) a 10%, corados com brometo de etídio, fotografados sob luz UVB com referência de tamanho de fragmento de DNA (*ladder*) de 25 pb e usando o software GelAnalyzer. As amplificações que produziram padrões polimórficos foram então analisadas em uma amostra de 20 peixes, para otimização das temperaturas de anelamento e condições de PCR. Os fragmentos de DNA foram então convertidos em classes nominais de alelos (01, 02, 03 e etc.) por meio de uma

abordagem de discretização (*binning*) por análise de histogramas de frequência em Excel (Microsoft).

Genética de populações

Foram utilizadas 100 amostras individuais de tecido de *P. lineatus* (fragmentos de nadadeiras caudais preservadas em etanol) do banco de DNA do LARGE. Estas foram coletadas anteriormente em três locais no último trecho barrado à montante, no alto rio Grande, MG. Especificamente, foram coletadas amostras de tecidos de peixes à jusante da barragem do Funil, à montante da barragem de Itutinga e à montante da barragem de Camargos (ver Suzuki *et al.*, 2011 para uma caracterização mais detalhada desses locais). O reservatório da barragem de Itutinga teve outros 40 peixes amostrado, mais recentemente, com a ajuda de um casal de pescadores profissionais. Essas informações, juntamente com os locais das amostras populacionais, estão resumidas na Tabela 1. Os locais são mostrados na Figura 4.

Tabela 1: Amostragem de indivíduos de *Prochilodus lineatus*, provenientes das três últimas represas do alto Rio Grande, MG, Brasil.

Amostra	ID	Local	Ano	N
Camargos	1-20	21°20'27.1"S 44°36'34.3"W	2013	20
Itutinga 1	21-60	21°17'39.9"S 44°37'24.1"W	2013	40
Itutinga 2	61-100	21°17'30.4"S 44°37'11.4"W	2023	40
Funil	101-140	21°08'39.4"S 45°02'19.1"W	2013	40
Total				140

ID=Identificação de indivíduos; N=Tamanho amostral.

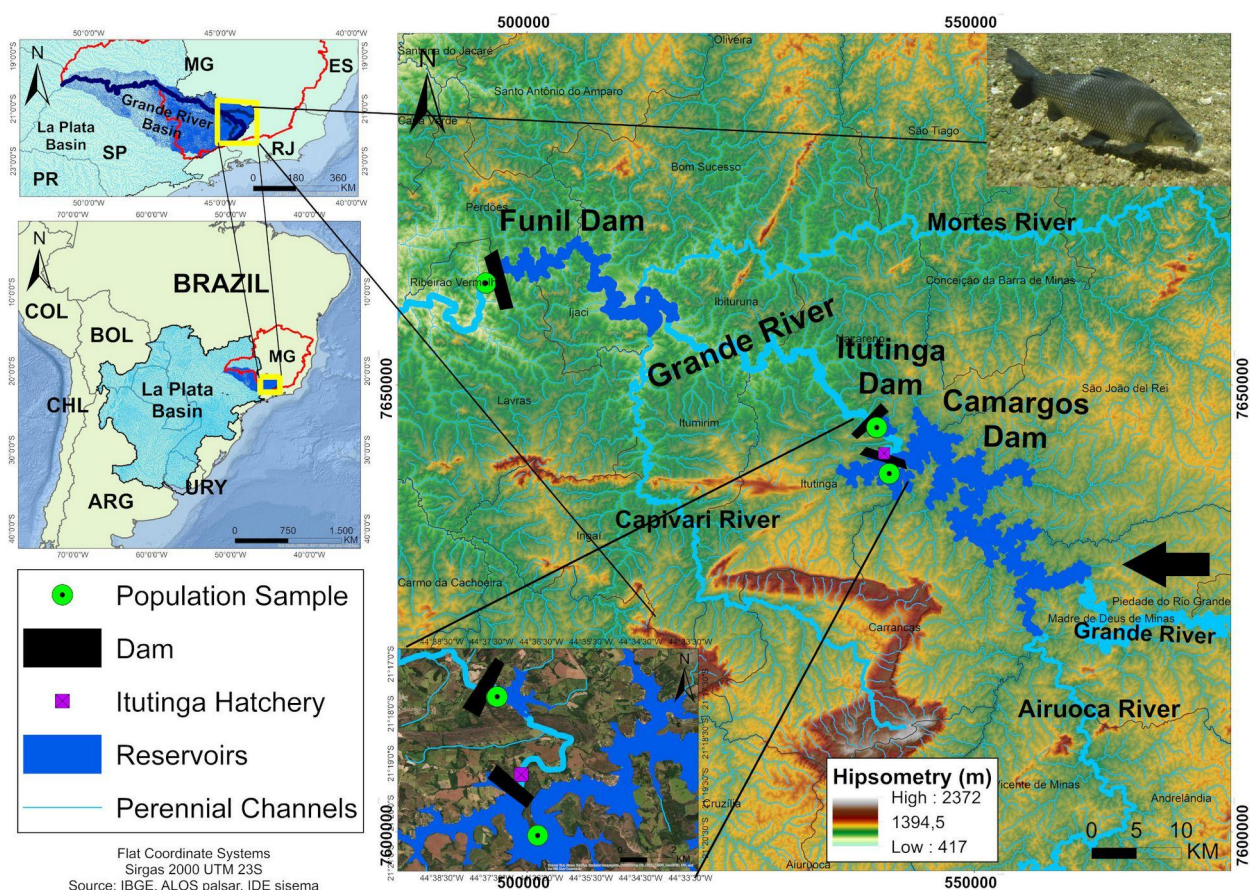


Figura 4: Localização geográfica das amostras populacionais estudadas para curimba *Prochilodus lineatus* do alto rio Grande , MG, Brasil.

Um total de 15 *loci* microssatélites validados apresentados em Nascimento (2024) foram aplicados nos 140 indivíduos de *P. lineatus*. Após duas rodadas de controle de qualidade visual da análise de comprimento de fragmentos, os dados foram formatados para genepop. O EHW e o desequilíbrio de ligação (DL) foram testados por meio dos métodos implementados no genepop 007 (Rousset, 2008), para cada amostra populacional, com opções padrão para parâmetros de Monte Carlo com cadeia de Markov (MCMC). Eles foram analisados sob novos valores de significância para testes múltiplos. Isso foi feito por meio de correções sequenciais de Bonferroni (também conhecida como correção de Holm-Bonferroni), controlando-se a taxa de descobertas falsas, a probabilidade de se encontrar, sob a hipótese nula, meramente ao acaso, uma distribuição de dados tão extrema quanto ou pior que aquela observada na amostra (ou seja, significância). O programa Microchecker (van Oosterhaut *et al.*, 2004) foi usado para controlar erros de

genotipagem (gagueira, *drop out* de alelos e alelos nulos). Visualizações e manipulação e conversão de dados e tabelas foram conduzidas principalmente com o auxílio do pacote adegenet para R (Jombart, 2008), a partir da importação inicial dos dados genepop.

Outras estimativas básica de diversidade e estrutura genética das amostras populacionais foram obtidas com GenAlex (Peakall e Smouse, 2012), incluindo heterozigosidades esperada e observada (H_E e H_O , respectivamente); número de alelos (ou riqueza alélica - N_A); D de Jost; índice de fixação desenviesado e padronizado, G''_{ST} ; e análise de variância molecular (AMOVA), cada uma com 999 permutas aleatórias para avaliar valores de probabilidade. Para D e G''_{ST} , erros-padrão (EP) foram calculados por *jackknife* sobre os *loci* e intervalos de confiança a 95% (IC) foram obtidos por *bootstrap* sobre os *loci* (Peakall e Smouse, 2012). Uma análise discriminante de componentes principais (DAPC - Jombart *et al.*, 2010) foi conduzida através do adegenet, por agrupamento *de novo*; retendo-se inicialmente 80% da variação total capturada por componentes principais (PC) e com k ótimo escolhido de acordo com o critério de informação Bayesiano (BIC). Um procedimento de validação cruzada, com 100 réplicas, foi executado e usado para avaliar a confiabilidade da análise DAPC, com o melhor número de PC encontrado. Finalmente, a DAPC também foi conduzida com informações de amostra populacional, com o mesmo número ótimo de PC retidos na validação cruzada, a fim de registrar visualmente a diversidade genética dentro e entre os grupos.

Análise hierárquica da estrutura genética populacional

Os dados de polimorfismos de microssatélites também foram meticulosamente processados com as etapas e métodos analíticos implementadas no programa PopCluster (Wang, 2022, 2024), principalmente pela versão gráfica para sistema operacional Windows (Microsoft), devido às limitações de produção de resultados visuais na versão de interface de linha de comando (Wang, 2024).

Optei pela execução adicional de um agrupamento exploratório (*clustering*) *k-means*, além do modelo nulo inicial de admistura, ou seja um modelo estrito de mistura genética (genomas individuais oriundos apenas de uma das K populações ancestrais possíveis). Os resultados deste “rascunho” são então usados para alimentar a estimativa de máxima verossimilhança (MLE) dos coeficientes individuais de admistura na amostra (q) - genomas individuais miscigenados – com base nas frequências alélicas definidas no agrupamento preliminar, o que contribui para parcialmente contornar o viés devido à amostragem desbalanceada de populações ancestrais. O programa PopCluster se vale da implementação de um algoritmo de arrefecimento simulado (*annealing*), que contorna a armadilha do aprisionamento de estimativas em torno de soluções sub-ótimas na função de verossimilhança, aumentando a chance de visita de todos picos de valores delineados nos modelos avaliados.

Depois de exaustivos ensaios exploratórios do programa PopCluster (*e.g.*, fator de escala fraco, médio, forte e nulo, com e sem agrupamento *k-means*, etc), segundo recomendações do autor, Jinliang Wang, apliquei a opção de "fator de escala fraco" (Wang, 2022). A intenção deste fator de escala também é lidar com o viés de eventual de amostragem desigual de indivíduos oriundos de diferentes demes ancestrais. Isso é normalmente mitigado na análise tradicional Bayesiana, assumindo-se priores desiguais para cada *cluster*, ver abaixo – Wang, 2017). Avaliei, portanto a MLE de admistura *de novo* (*i.e.* sem usar informações de amostras populacionais dos indivíduos) variando valores de K entre 1 e 14, com 100 réplicas independentes cada.

A escolha do valor K ideal que melhor captura a subestrutura da amostra foi feita principalmente pelo estimador de taxa de mudança (delta) da log-verossimilhança de 2ª ordem, D_{LK2} (Wang, 2022), cuja premissa é comparável ao estimador ΔK de Evanno *et al.* (2005). Para o mesmo fim, alternativamente, foram computados valores do chamado F_{STIS} , daquele autor, que visa escolher o cenário de K ideal com base na maximização do F_{ST} entre *clusters* e na minimização do F_{IS} dentro

de *clusters*. Este procedimento inicial foi rotulado como uma análise de 1ª ordem da estrutura genética populacional.

Então, para explorar possíveis estratos hierárquicos escondidos da estrutura genética, seguindo os passos de Vähä *et al.* (2007) e defendidos em Janes *et al.* (2017): os melhores resultados de réplicas para o valor K escolhido definido na rodada de estimativa de 1ª ordem, foram usados para se reorganizar amostras em K novos grupos. Cada grupo resultante foi, por sua vez, submetido independentemente a uma análise de mistura de 2ª ordem, buscando novas subdivisões, variando K=1 a 7. Finalmente, este procedimento foi reiterado mais uma vez, com os novos grupos definidos na etapa anterior, como uma análise de 3ª ordem da estrutura populacional, agora avaliando K de 1 a 4. Isso permitiu uma análise hierárquica da estrutura genética de populações em três níveis. Os resultados para cada rearranjo sugerido de *clusters* foram examinados em relação à diversidade, estrutura e variação genética, conforme conduzido para as amostras populacionais originais, com o auxílio do GenAlex.

Ainda de seguindo a discussão apresentada por Wang (2022), devido ao número limitado de marcadores e possível amostragem desbalanceada entre as populações de origem, também realizamos uma análise da admistura por meio de inferência Bayesiana, em paralelo, com a suíte STRUCTURE (Pritchard *et al.*, 2000). Para isso, adotei diferentes priores para o parâmetro de admistura (q), para cada população, onde o parâmetro α da distribuição de Dirichlet difere para cada *cluster* (ou seja, parâmetro POPALPHAS definido como 1). Isso equivale a não se assumir de que cada população de origem contribui igualmente para a amostra, melhorando a acurácia da análise, segundo Wang (2017), uma vez que a escolha de *prior* é um passo essencial da análise.

Um total de 5×10^5 passos da cadeia foram ignoradas antes de computar 1×10^6 iterações do MCMC, com K variando de 1 a 8 e 20 réplicas cada. As estimativas do melhor valor K para a análise do STRUCTURE foram realizadas principalmente com os estimadores implementados no

Kfinder (Wang, 2019), Structure Selector (Li e Liu, 2018) e visualizações foram conduzidas com a ajuda do clumpak (Kopelman *et al.*, 2015).

Declarações éticas

Todos os procedimentos com animais foram conduzidos sob a as normas e licenças 27/2011 e 6991130223 do Comitê de Ética no Uso Animal (CEUA-UFSJ); além das licenças SISGEN A9D0E5 e AC5A784; e licenças SISBIO 37222-2 e 85497-1.

Alguns dados utilizados nesta tese, bem como muitos de seus resultados analíticos associados, foram anteriormente apresentados em Nascimento (2024), particularmente aqueles relativos ao desenvolvimento empírico de marcadores microsatélites e sua aplicação na detecção de polimorfismos genéticos das amostras de *P. lineatus*. Este trabalho foi apresentado no contexto de defesa de dissertação de Mestrado do Programa de Pós Graduação em Ecologia da UFSJ pela discente Gésica A. S. Nascimento, minha orientada no LARGE-UFSJ, em Julho de 2024. No presente estudo, estes resultados são revisitados, analisados e discutidos sob a ótica dos objetivos diferenciais listados neste presente estudo, além dos novos elementos (dados e resultados) aqui apresentados.

Sem considerar técnicas de aprendizado de máquinas intrínsecas a métodos listados anteriormente (*e.g.* validação cruzada da DAPC), o uso de inteligência artificial (IA) e agentes de modelos grandes de linguagem (LLM) de IA generativa, limitou-se principalmente a:

- Perplexity (*Online*), Hermes2 e llama3.2 (ambos *via* Ollama): aplicação incidental durante o refinamento de código manual na linguagem R, para apresentação gráfica da análise de estimativa *jackknife* de riqueza alélica; código para concatenação de pacotes R, para processamento de matrizes de dados genéticos S4 previamente preparadas para adgenet, na análise e apresentação gráfica de rarefação desta mesma variável (N_A); Avaliações de conteúdo textual escrito manualmente por mim na língua Inglesa;

- llama 3.2, Hermes3, e sobretudo deepseek-r1:14b (todos *via* Ollama): Sondagens e reflexões voláteis (*i.e.* não registradas ou diretamente trazidas para o texto) sobre a razoabilidade e bases teóricas (*e.g.* fractais, autocorrelação, estrutura genética e admistura) e metodológicas (*e.g.* modelagem, inferência por MLE vs Bayesiana; limitações de PAGE vs eletroforese capilar para análise de fragmentos), no contexto dos resultados e discussões oriundos do estudo; consultas gerais sobre estatísticas, computação e filosofia e história da ciência (sempre aprofundado e controlado por meio de verificação cruzada e consulta direta à literatura primária); além de
- Uso do Google Tradutor em tradução textual reversa do Inglês do arcabouço de um manuscrito original (Apêndice A – Arquivo suplementar extra) para o Português (sob supervisão estrita, livre editoração corretiva, por mim).

Resultados

Genômica e bioinformática

Um total de aproximadamente 483,7 milhões de leituras curtas de pares de 90 bases foram geradas a partir das três bibliotecas genômicas integrais *shotgun* para *P. lineatus*. Isso totaliza cerca de 43,53 Gb (gigabases) armazenadas no formato FASTAQ. Estes dados estão disponibilizados no NCBI, no *Sequence Read Archive* (SRA) e podem ser rastreadas pelo código de acesso do Bioproject PRJNA1120875. O conteúdo de CG dessas leituras FASTQ foi de 42%, cada biblioteca teve de 13,4 a 14,4% de leituras duplicadas e nenhuma sequência sobre-representada foi observada. O valor de qualidade Q por base variou de 31 a 39, com uma média de 38 por leitura, na escala Phred (média de 1 erro por cada 6.250 pb). A análise do tamanho do genoma, com base na contagem de 21-meros das leituras curtas, resultou na estimativa de ~1,65 Gb para *P. lineatus*.

O procedimento de montagem *de novo* baseado em grafos de Bruijn resultou em uma montagem final do genoma com um conteúdo de CG de 42,18%, um tamanho acumulado total de

1.404.931.824 pb, para um complemento haploide colapsado (ou 1,4 Gb, sendo ~1,4% dos nucleotídeos atribuições de base indefinidas – N – fornecendo, porém, informação posicional. Com base no tamanho do genoma e o número de bases obtidas, a cobertura média de sequenciamento de foi de cerca 31X (ou seja, a redundância média de cada par de bases do genoma).

Este projeto *Whole Genome Shotgun* foi depositado no DDBJ/ENA/GenBank sob o acesso JBGFTF0000000000. A versão descrita neste trabalho é a JBGFTF0100000000. No entanto, esta é extensivamente fragmentária, com um total de 738.832 sequências, contendo 370.341 lacunas internas (*gaps* posicionais de bases N indefinidas). Seu N_{50} =5.503 (mediana do tamanho dos fragmentos da montagem), com 686.224 sequências menores que este valor em pares de bases (ou seja, L_{50} =52.608) e com comprimento médio de contig/*scaffold* de 1.901,56, sendo o mais longo de 192.837 pb. Esses resultados são resumidos visualmente na Figura 5.

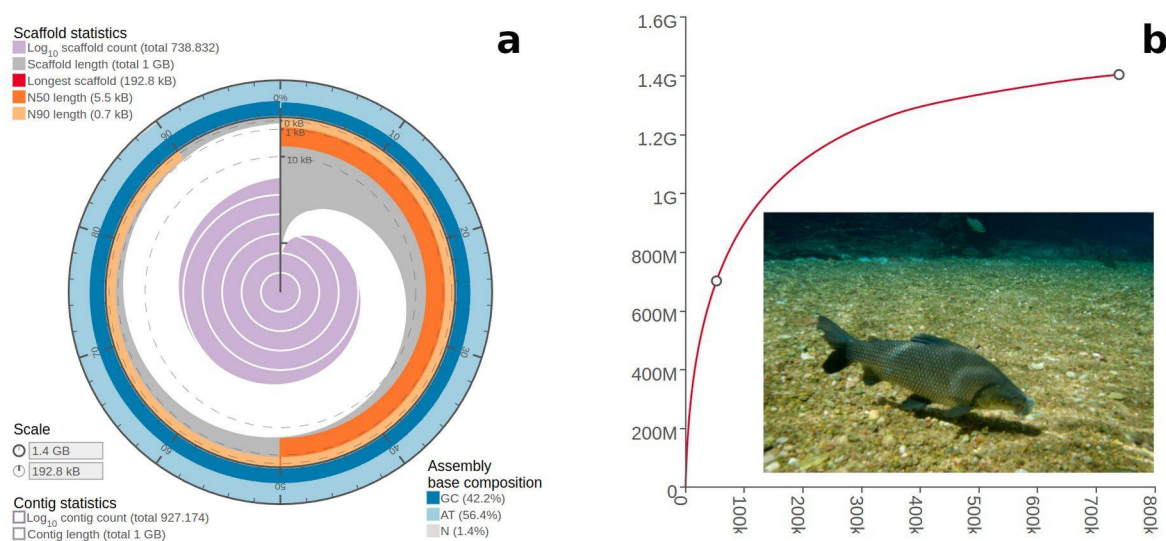


Figura 5: Características estatísticas da montagem do genoma de *Prochilodus lineatus*. **a)** O raio dentro do gráfico circular marca o comprimento do maior *scaffold* resolvido e os segmentos cinza representam a porcentagem cumulativa da montagem contida em *scaffolds* iguais ou maiores que um comprimento específico. **b)** Comprimento cumulativo do contig/*scaffold* da montagem genômica, como uma função do número de sequências resolvidas. O ponto interior da curva é o marco L_{50} , o menor número de sequências, cuja soma ultrapassa a metade do comprimento total da montagem do genoma (ponto no final da curva). G = Gigabases e k = milhares de sequências.

O arquivo BAM de mapeamento de alinhamentos das leituras curtas sobre essa montagem genômica também está disponível no SRA do NCBI, com cobertura média de 9,81X com ~95% de

scaffolds e ~94% de bases cobertas. A sequência completa do mitogenoma deste espécime foi apresentada anteriormente em Santos *et al.* (2021).

Um total de 37.476 *loci* microssatélites di- a hexa-nucleotídeos para *P. lineatus* foram caracterizados *in silico* e compilados no Apêndice A (Arquivos Suplementares 1 e 2). Estes foram mapeados na montagem do genoma apresentado aqui, mas cerca de 12% dos *loci* potenciais não puderam ser encontrados nesta versão.

Desenvolvimento e validação de marcadores moleculares

De 22 possíveis marcadores microssatélites analisados empiricamente em ensaios PCR, cinco (Prol14, Prol15, Prol18, Prol31 e Prol52) resultaram em padrões de amplificação complexos e foram abandonados. Dois, (Prol56 e Prol28) não resultaram em amplificações observáveis sob condições testadas. Um total de 15 marcadores moleculares microssatélites foram validados e são apresentados no Apêndice B (Tabela suplementar 1), com informações de sequência de *primers*, temperatura de anelamento e resultados de polimorfismo básicos para a amostra combinada de 140 peixes. As condições finais de PCR foram as mesmas das inicialmente testadas, exceto para Prol02 e Prol19, que usaram ~6,7 pmol de cada *primer* por reação, sendo este com 35 ciclos de PCR, além do marcador Prol54, que também usou 35 ciclos ao invés de 30. O comprimento de fragmentos de DNA observados variou entre 87 a 389 pb. O número de alelos observados nos 15 *loci* variou de 13 a 31, com média de 22,7 ($\pm 1,3$ EP).

A taxa de falha de amplificação/genotipagem variou de 2,1% (Prol53) a 32,8% (Prol07), com média de 18,6% ($\pm 7,9$ EP), incluindo casos em que um alelo foi atribuído e o outro considerado duvidoso (*i.e.* dado ausente, *e.g.* 02 00) para o mesmo genótipo diploide. Seis amostras de DNA (quatro de Camargos, uma de Itutinga 1 e uma de Funil) falharam em amplificar entre cinco e dez *loci*. Elas foram brevemente removidas do conjunto de dados e a maioria das análises posteriores foram realizadas com e sem as mesmas, em paralelo. No entanto, diferenças marginais ou nulas foram observadas ao comparar-se os resultados e, portanto, essas amostras foram retidas.

Genética de populações

A significância reajustada para 60 testes múltiplos de EHW foi definida como $p < 0,008$.

Apenas seis *loci* foram considerados em EHW (Apêndice B - Tabela suplementar 2). O único *locus* encontrado em conformidade com EHW dentro de cada amostra, Prol07, não estava, porém, em equilíbrio no total das amostras populacionais ($p = 0,002$). Todos os desvios foram na direção de déficits de heterozigotos.

Os testes para DL dentro de amostras populacionais tiveram um nível de significância ajustado para $\alpha = 0,0001$. Desequilíbrio de Ligação foi observado em seis situações: Prol06 e Prol54 em Camargos; Prol01 e Prol19, Prol05 e Prol57 para Itutinga 1; Prol01 e Prol06, Prol04 e Prol07 para Itutinga 2; e finalmente, Prol53 e Prol57 no Funil. Os seguintes pares de *loci* estão em DL para a amostra combinada de peixes: Prol01 e Prol06; Prol04 e Prol07; Prol01 e Prol19; Prol06 e Prol54; Prol05 e Prol57; e, por último, Prol53 e Prol57.

Não houve evidência de erros de genotipagem devido à gagueira ou *drop out* de alelos, mas a presença de alelos nulos foi sugerida como significativa, devido ao excesso geral de homozigotos para a maioria das classes de tamanho alélicos, para todos os *loci*, em todas as amostras populacionais, exceto para Prol07.

A riqueza alélica variou de 10,8 ($\pm 0,8$ EP) alelos para Camargos a 16,9 ($\pm 0,9$ EP) alelos para e Itutinga 2 (Apêndice B - Tabela suplementar 2). Essa tendência foi confirmada pelas estimativas *jackknife* de riqueza alélica (Figura 6), juntamente com os resultados da análise de rarefação (Figura 7). Camargos também teve a menor heterozigosidade esperada, enquanto exibiu o maior valor de coeficiente de endogamia, junto com a amostra Itutinga 2 (Apêndice B - Tabela suplementar 2). Todas as amostras populacionais exibiram alelos privados, variando de 7 alelos para Camargos e 37 alelos para Itutinga 2 (média de $19,7 \pm 12,7$ EP).

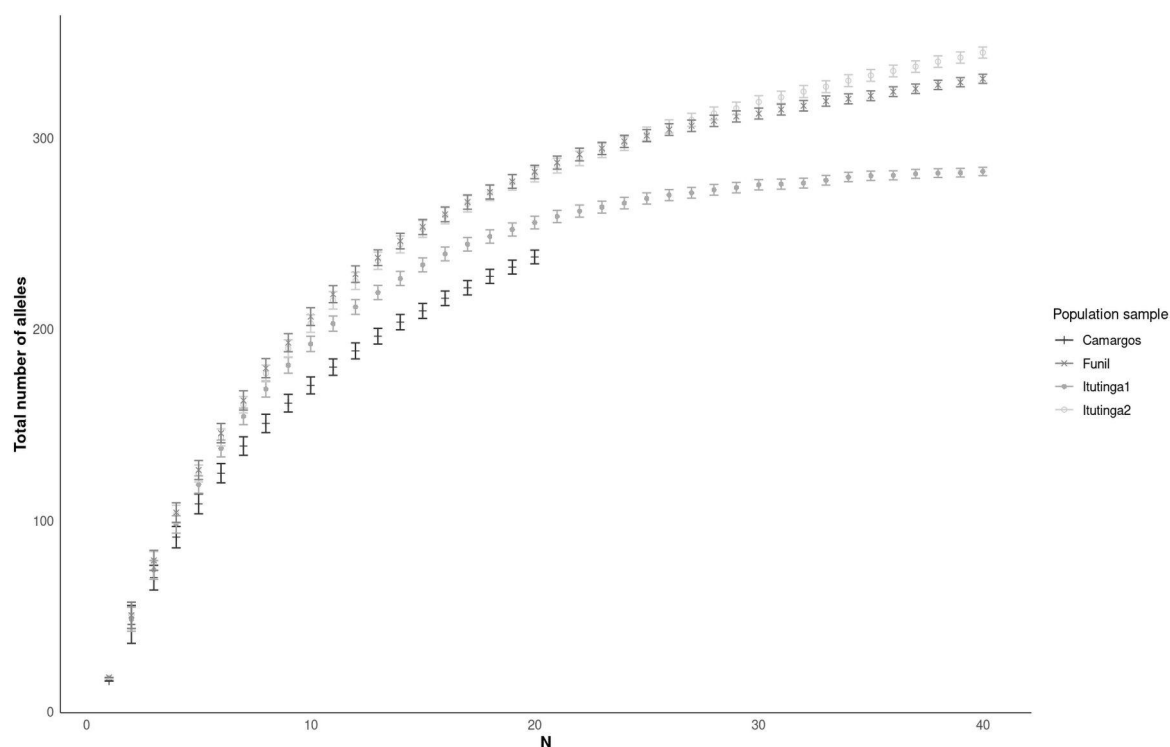


Figura 6: Estimativas de reamostragem *jackknife* e intervalos de confiança a 95% (linhas verticais) de 341 alelos, em 15 *loci* de *Prochilodus lineatus* das quatro amostras populacionais do alto rio Grande, MG, Brasil.

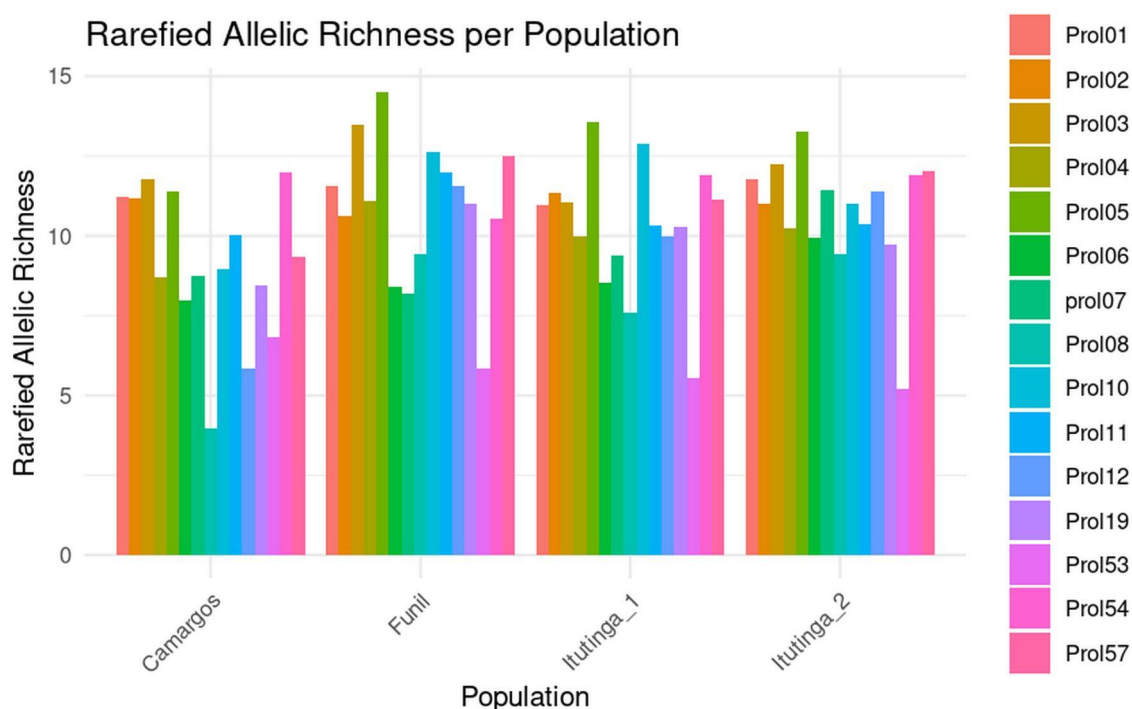


Figura 7: Estimativas de rarefação do número de alelos (riqueza alélica), em 15 *loci* em *Prochilodus lineatus* das quatro amostras populacionais do alto rio Grande, MG, Brasil.

Os resultados da AMOVA mostraram que apenas 1% da variação total foi explicada pelas quatro amostras populacionais, com a variação restante sendo explicada por diferenças entre indivíduos (99%). Todas as estimativas de fixação par-a-par para G''_{ST} e D de Jost foram diferentes de 0, incluindo entre amostras coletadas no mesmo local em momentos diferentes (Tabela 2). As amostras de Itutinga (2013 e 2023) apresentaram os menores valores, enquanto Camargos apresentou a maior variação e diferenciação par-a-par em geral, para ambos os respectivos índices.

Tabela 2: Sumário de índices de fixação G''_{ST} par-a-par (diagonal inferior) e da distância genética D de Jost (diagonal superior) para polimorfismos de microssatélites entre amostras populacionais de *Prochilodus lineatus* do alto rio Grande, MG. * $p < 0.05$; ** $p \leq 0.001$

	Camargos	Itutinga 1	Itutinga 2	Funil
Camargos	-	0,156**	0,222**	0,203**
Itutinga 1	0,171**	-	0,066*	0,087**
Itutinga 2	0,241**	0,072*	-	0,167**
Funil	0,221**	0,095**	0,180*	-

Considerando todos os 15 *loci* de microssatélites, entre as quatro amostras populacionais, o índice de fixação padronizado desenviesado G''_{ST} atingiu 0,16 ($\pm 0,03$ EP; 0 a 0,26 IC, $n=999$; $p=0,001$) e a estimativa geral de diferenciação genética do D de Jost foi de 0,15 ($\pm 0,03$; 0 a 0,24 IC, $n=999$; $p=0,001$).

Análise hierárquica da estrutura genética populacional

A estimativa de máxima verossimilhança (MLE) de admistura (Figura 8) inicialmente teve $K=2$ apontado como ótimo pela métrica D_{LK2} de Wang. No entanto, o F_{STIS} definiu $K=4$ como o número mais provável de grupos (Apêndice C – Figura suplementar 1). A análise de admistura de 1ª ordem permitiu a definição de dois *clusters* (com respectivos ID originais de peixes – Figura 8a). As estimativas de D_{LK2} para ambos os *clusters* produziram o valor de $K=2$ como ótimo na análise de 2ª ordem (Figura 8b), enquanto o F_{STIS} registrou valores de $K=5$ e $K=7$ (Apêndice C – Figura suplementar 2). Finalmente, esses quatro novos rearranjos resultantes foram estudados na análise de

3ª ordem (Figura 8c). O D_{LK2} nova e consistentemente apontou $K=2$ para cada *cluster* (Apêndice C – Figura suplementar 3).

As designações de indivíduos (ID) em *clusters* de todas ordens hierárquicas podem ser recuperadas no Apêndice A (Arquivo Suplementar 3). Abordei e descartei a possibilidade de panmixia ($K=1$), em cada nível de ordem hierárquica, inspecionando diretamente os valores de log-verossimilhança da análise de agrupamento, em relação aos resultados para o valor assumido para K (Apêndice A - Arquivo Suplementar 4 - Tabela suplementar externa 1).

Os diferentes arranjos de amostras mostraram uma diminuição consistente nos parâmetros de diversidade intrapopulacional para as ordens superiores, como heterozigosidade esperada e número de alelos (Apêndice A - Arquivo Suplementar 4 - Tabela suplementar externa 2), enquanto os indicadores interpoblacionais (G''_{ST} e D de Jost) aumentaram constantemente (Apêndice A - Arquivo Suplementar 4 - Tabelas suplementares externas 3-5). A análise EHW de *clusters* definidos por MLE de fato mostrou uma tendência em direção às expectativas de Hardy-Weinberg, nos arranjos de grupos de ordem superior. Em média, os *clusters* de 3ª ordem tiveram mais *loci* em EHW, com um dos *clusters* em particular tendo oito *loci* em equilíbrio (Apêndice A - Arquivo Suplementar 5).

A DAPC realizada *de novo* (i.e. sem informações de amostra populacional) apontou para dois *clusters* principais, de acordo com o BIC (Figura 9a). Esta DAPC inicial (Figura 9b) também revelou que dois alelos do *locus* Prol53 (contig C1608592 0.0 da montagem genômica) tiveram o maior peso na definição de contribuições das variáveis. Buscas BLAST no NCBI com as sequências preditas do produto de PCR de Prol53 e de seu contig de origem, ambas apontam para genes como *ABHD10*, relacionado ao metabolismo do fígado em humanos e *PGDB4*, uma proteína remotamente relacionada a elementos transponíveis *piggyBac*, implicada no neurodesenvolvimento e alteração dos movimentos normais dos olhos em humanos.

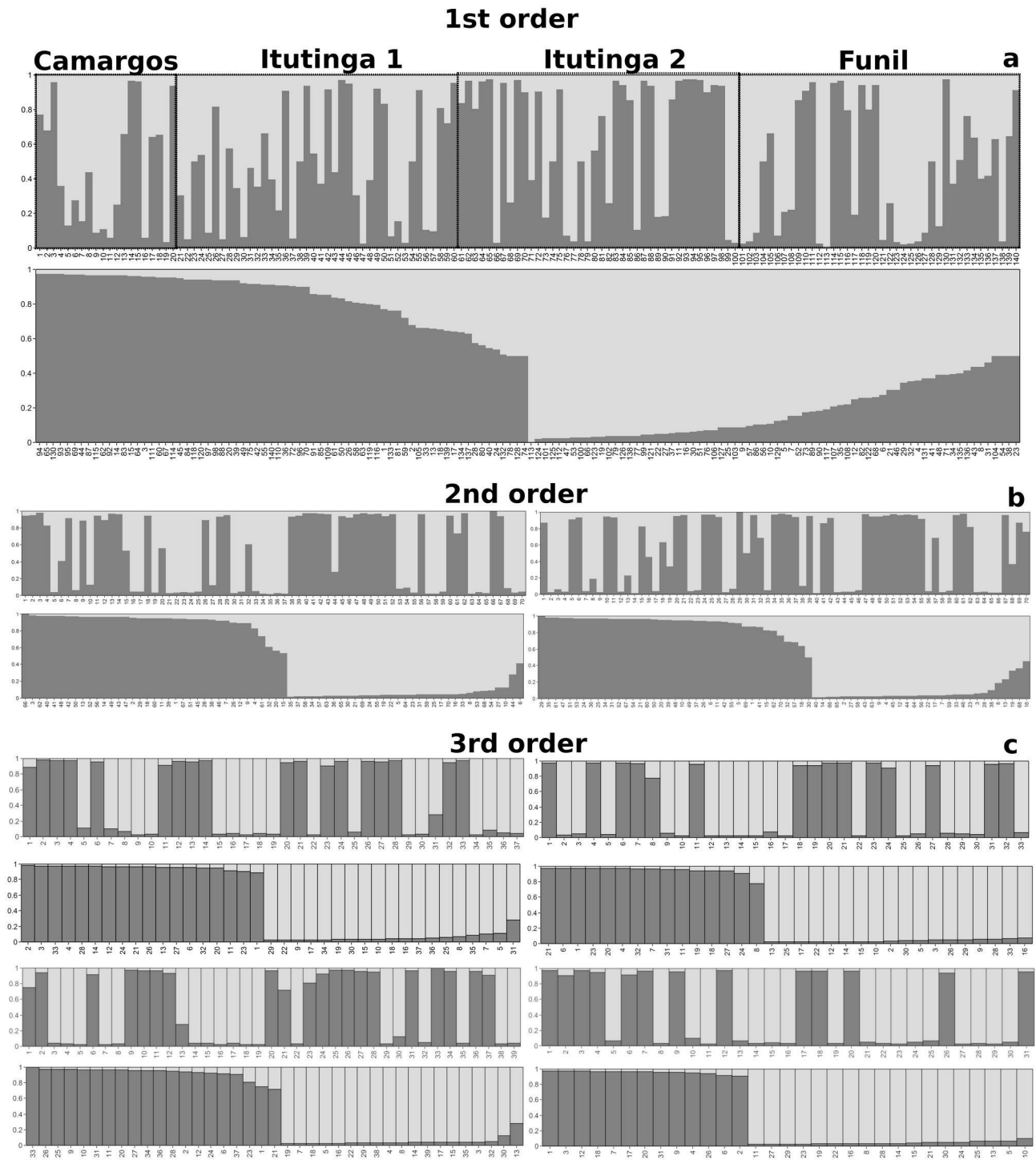


Figura 8: Coeficientes MLE de admistura individuais (barras) para $K=2$, em níveis inclusivos de estrutura genética hierárquica. Diferentes tons de cinza representam grupos alternativos de origem, com o tamanho proporcional das barras indicando as respectivas contribuições genômicas de demes ancestrais. Para cada painel, o diagrama acima exhibe as amostras na ordem de entrada dos dados, enquanto o diagrama abaixo mostra a amostra reorganizada nos grupos definidos. **a)** Análise de 1ª ordem, mostrando peixes agrupados de acordo com a amostra populacional e ID original. **b)** Análise de 2ª ordem de grupos definidos na 1ª ordem. **c)** Análise de 3ª ordem de grupos de 2ª ordem. Para 2ª e 3ª ordens, os números abaixo das barras são arbitrários e não equivalem ao ID original.

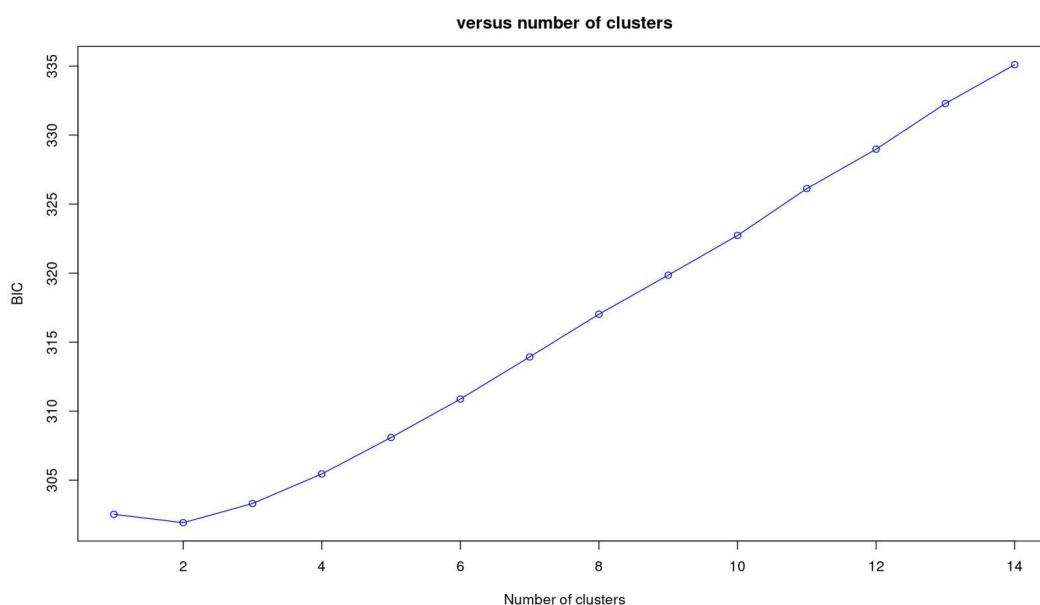
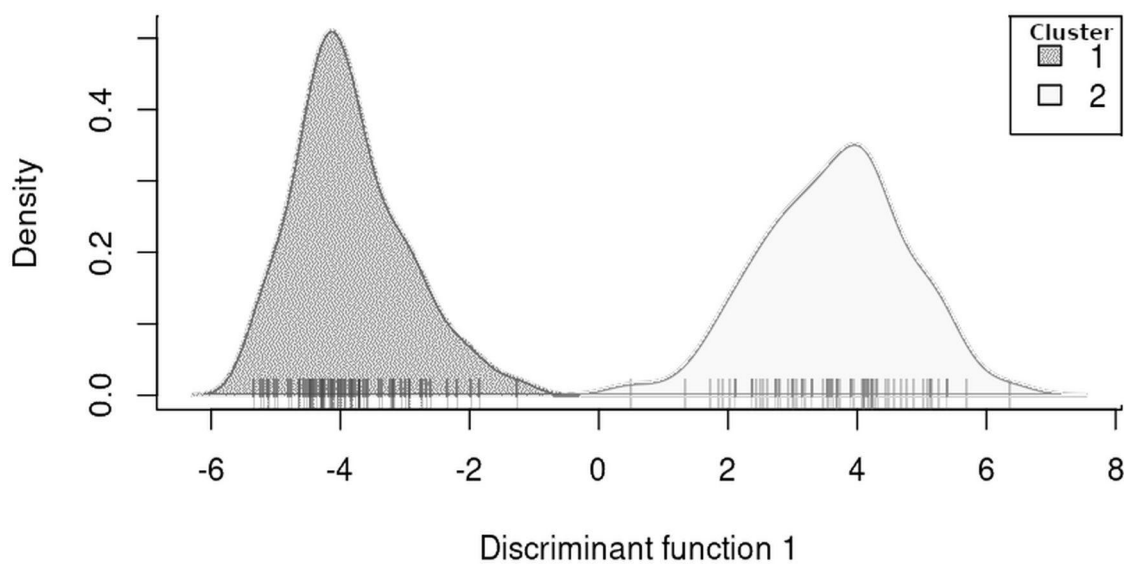
a**b**

Figura 9: Análise discriminante de componentes principais (DAPC) *de novo* para variabilidade da amostra total de *Prochilodus lineatus* do alto rio Grande, MG: **a)** Critério de informação Bayesiana (BIC) apontando o melhor número de *clusters* que capturam a variação total da amostra. **b)** Definição de dois *clusters* ratificados por uma única função discriminante.

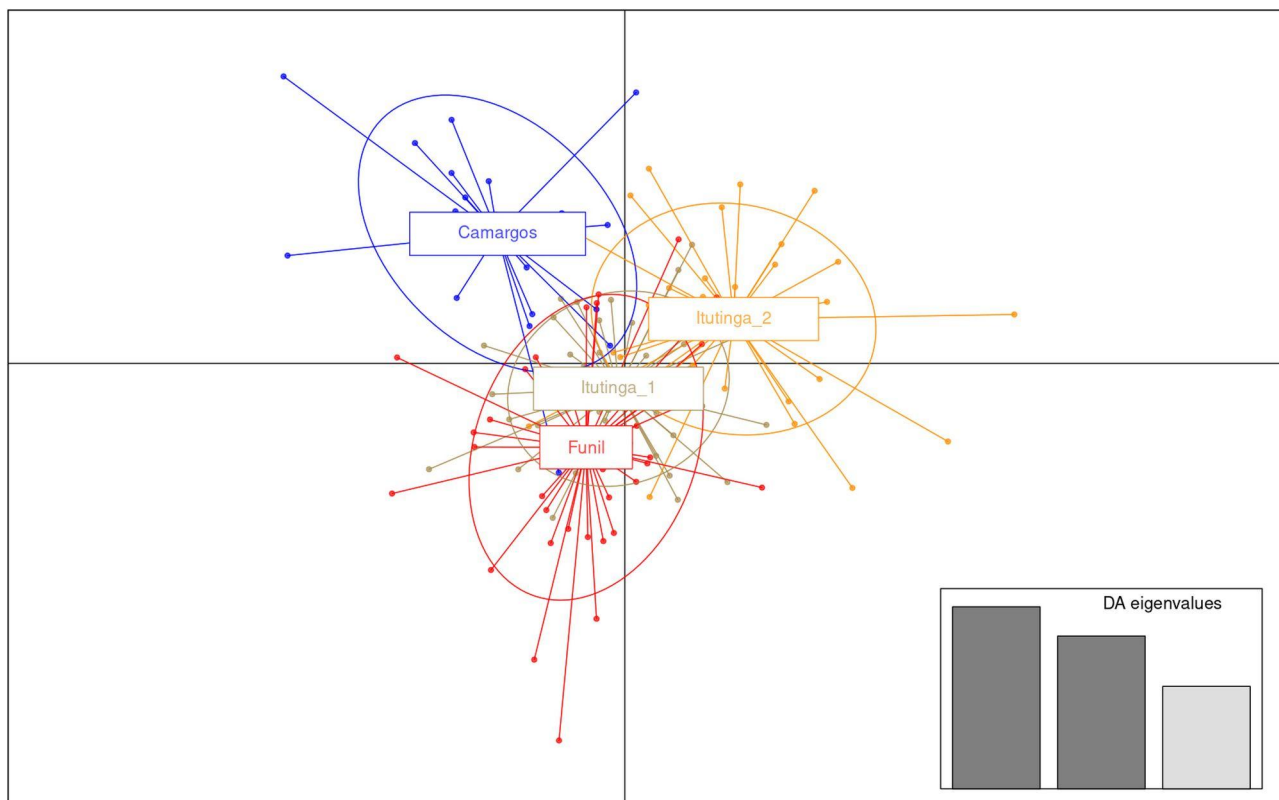


Figura 10: Análise discriminante de componentes principais (DAPC) para *Prochilodus lineatus* do alto rio Grande, MG, realizada com informações de origem da amostra populacional, com três funções discriminantes (os *eigenvalues* - autovalores - da análise discriminante em escala são mostrados no canto inferior direito).

O ensaio de validação cruzada apontou 30 como o número ideal de PC a ser retido (Apêndice C – Figura suplementar 4) e a ratificação do DAPC de acordo com este parâmetro levou a resultados muito convergentes (Apêndice C – Figura suplementar 5). A DAPC realizada com informações de amostras populacionais e este número de PC revelou um nível aparente de diferenciação genética entre as amostras populacionais (Figura 10).

A abordagem Bayesiana para o modelo de admistura (Figura 11) também resultou em valores alternativos para estimativas ótimas de K. O estimador $\Pr(X|K)$ de Pritchard define K=8, enquanto, ambos, os estimadores ΔK de Evanno e de parcimônia de Wang resolveram K=5. Os gráficos de admistura para MLE para os valores das melhores análises de K=5 e 8 são mostrados na Figura 12.

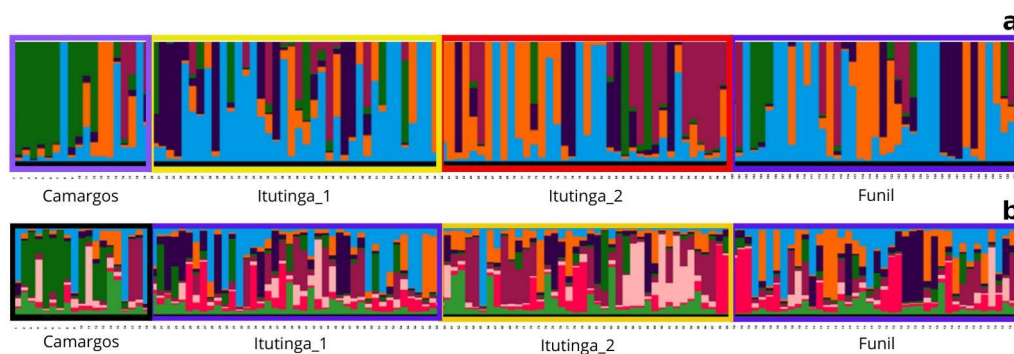


Figura 11: Resultados da análise Bayesiana de admistura de *Prochilodus lineatus* do alto rio Grande, MG. **a)** Cenário de K=5 *clusters* e **b)** Cenário de K=8. Cada cor representa um deme alternativo de origem em suas respectivas proporções nos genomas individuais (barras verticais).

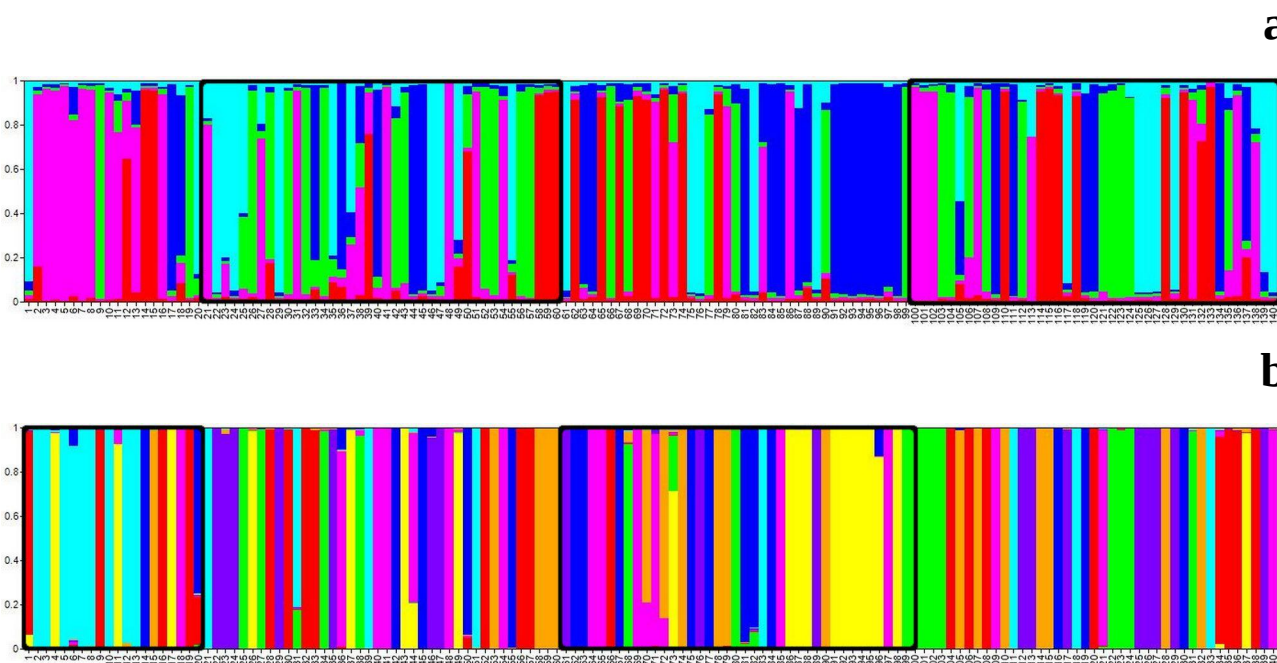


Figura 12: Resultados da análise de admistura por MLE para os cenários de K=5 (**a**) e 8 (**b**) em *Prochilodus lineatus* do alto rio Grande, MG. Cada cor representa um *cluster* alternativo de origem, com suas respectivas proporções de contribuição para genomas individuais (barras verticais). Os números indicam o ID original e os quadros, separam, respectivamente, as amostra populacionais de Camargos, Itutinga 1, Itutinga 2 e Funil.

Discussão

Genômica e bioinformática

A aplicação da genômica nos avanços da compreensão da história natural, gestão ambiental eficiente e práticas sustentáveis de aquicultura tem crescido rapidamente entre espécies de água doce. Genomas de referência de alta qualidade estão se tornando atualmente indispensáveis para

muitas diferentes disciplinas da Biologia (Jung *et al.*, 2020). Este trabalho contribui neste sentido para *Prochilodus lineatus*, um importante organismo-modelo potamôdromo Netropical e um dos principais recursos pesqueiros para potencialmente milhões de pessoas na América do Sul, ao apresentar publicamente seus primeiros recursos genômicos abrangentes de média cobertura.

Os dados de sequência de DNA de leituras curtas PE apresentados aqui superam os únicos dados FASTQ disponíveis anteriormente para *P. lineatus* (Stornioli *et al.*, 2021) por um fator de 12. No entanto, eles estão uma ordem de magnitude abaixo dos dados de leituras curtas relatados para uma espécie congênere colombiana, *P. magdalenae*, por Yepes-Blandón *et al.* (2023) que, entretanto, não divulgaram publicamente estas sequências. O conteúdo de CG e a média de qualidade das leituras foram os mesmos apresentados anteriormente para outros peixes piracema (*e.g.* Yazbeck *et al.* 2018; Graciano *et al.*, 2022), indicando um conjunto de dados com sequências de DNA de alta qualidade. Essas leituras curtas podem ser úteis em novas tentativas de montagem de genoma de referência no nível cromossômico ou telômero a telômero e possivelmente encontrarão aplicações em outros estudos (Jung *et al.* 2022; Sayers *et al.*, 2022a). Esses arquivos FASTQ para *P. lineatus* estão agora disponíveis publicamente para a comunidade científica no banco de dados SRA no NCBI (Sayers *et al.*, 2022b) e são o conjunto de dados mais extenso de leituras curtas divulgado publicamente para qualquer membro da família Prochilodontidae, até o momento.

Este trabalho também fornece a primeira montagem genômica de referência disponível para *P. lineatus*, resolvida no nível de *scaffolds*. Apesar de ser dramaticamente fragmentada, seu tamanho caracterizado é ligeiramente maior do que o relatado para *P. magdalenae* (Yepes-Blandón *et al.*, 2023), porém inferior à estimativa realizada aqui pelo conjunto de 21-meros (este método nem sempre é preciso - Pflug *et al.*, 2020). Esta montagem genômica atingiu apenas cerca de metade da contiguidade das montagens genômicas metodologicamente comparáveis em peixes de piracema (Yazbeck *et al.* 2018; Graciano *et al.*, 2022), embora para o genoma de *P. lineatus* tenham sido

empregadas cerca de três vezes mais leituras curtas. Comparada com a montagem de *P. magdalenae* a contiguidade do N₅₀ é ~62 vezes menor (esta empregou também dados de leituras longas).

Características genômicas intrínsecas desta espécie, como alto conteúdo de repetições, heterozigiosidade do espécime usado na montagem e outros fatores podem possivelmente dificultar a montagem *de novo* (Bradnam *et al.*, 2013). O genoma de *P. magdalenae*, por exemplo, foi anotado com cerca de 31% de elementos repetitivos (Yepes-Blandón *et al.*, 2023). Montagens *de novo* realizadas exclusivamente com leituras curtas (90 bases, aqui) não resolvem corretamente regiões com períodos repetitivos maiores que esse limite de comprimento (Wang e Wang, 2023). Isso fragmenta irremediavelmente a montagem. Esse não foi considerado um problema para os objetivos desta montagem, motivada primariamente pela busca de *loci* de microssatélites.

Essa primeira montagem do genoma de *P. lineatus* fornece dados de DNA sem precedentes nesta espécie e pode ser explorada de diferentes maneiras, como para a anotação inicial dos fragmentos maiores de sequências caracterizadas (*e.g.* Graciano *et al.*, 2021). Com o eventual uso de leituras longas de DNA (*e.g.* PacBio) e informações de captura conformacional de cromossomos (*e.g.* Hi-C) a montagem genômica e as leituras FASTQ produzidas aqui podem ser integradas, para a obtenção futura de um genoma de referência de alta qualidade, no nível de cromossomos para *P. lineatus*.

O arquivo de alinhamento de mapeamento de sequência (BAM) fornecido aqui, depositado no NCBI, pode ser usado diretamente na análise de profundidade do sequenciamento (ou seja, confiabilidade do conteúdo da sequência) de qualquer região de interesse específica resolvida na montagem genômica. Este também pode ser explorado para encontrar-se genes específicos ou caracterizar potenciais marcadores de polimorfismos de nucleotídeo único (SNP - Vaisband *et al.*, 2023).

O painel genômico de microssatélites produzido aqui, com mais de 37 mil *loci* potencias, agora pode ser facilmente acessado por diferentes grupos interessados (pesquisadores e gestores) para o rápido desenvolvimento de baixo custo de centenas de novos marcadores moleculares, por NGS ou uso tradicional. Uma desvantagem frequentemente apontada sobre o desenvolvimento de microssatélites, em comparação com SNP, é o custo *ab initio* (Putman e Carbone, 2016) e a disponibilização deste painel rompe essa inércia inicial.

Este painel tem, ainda, o potencial para uso no desenvolvimento tecnológico de ferramentas práticas para pisciculturas e gestão da pesca, como na produção de conjuntos multiplexados de marcadores (*e.g.* Carvalho *et al.*, 2019). Alguns desses recursos também podem, certamente, ser aplicados a outros peixes, particularmente espécies de Prochilodontidae e Characiformes em geral (*e.g.* Barbosa *et al.*, 2006; Carmo *et al.*, 2015).

Desenvolvimento e validação de marcadores moleculares

A taxa relativamente alta de sucesso em ensaios empíricos (cerca de 2/3 dos 22 *loci* testados) atesta a alta qualidade deste painel. *Loci* que não obtiveram amplificações podem ser nulos devido a artefatos bioinformáticos (*e.g.* quimeras na montagem), mas podem, alternativamente, ser devido a fatores biológicos, por exemplo, terem sido eventualmente caracterizados em um cromossomo supranumerário e tipados em indivíduos sem estes cromossomos. Os pares de *primers* que resultaram em padrões complexos podem ser possivelmente otimizados, caso eventualmente revisitados.

Os 15 novos marcadores de DNA apresentados originalmente em Nascimento (2024), reprisados e discutidos aqui, aumentam em ~84% o número de microssatélites disponíveis para *P. lineatus*. Observamos uma taxa variável de sucesso de amplificação/genotipagem entre esses novos marcadores. Alguns *loci* apresentaram uma taxa de falha de amplificação/genotipagem relativamente alta (*e.g.* Prol07). O microssatélite Prol53 teve uma taxa particularmente alta de amplificação e genotipagem e alto conteúdo de informação. Apesar do Microchecker sugerir a

possível presença de alelos nulos para a maioria dos *loci* em todas as populações, considero isso um artefato devido à subestruturação inerente da amostra, o que leva a efeitos de confusão críticos nesta análise (Waples, 2015).

Algumas abordagens (*e.g.* atribuições de admistura) são consideradas razoavelmente insensíveis a dados faltantes, particularmente quando não são devidos a uma causa sistemática (Falush *et al.*, 2007; Wang, 2022). Mesmo alelos nulos podem ter apenas um impacto marginal (Carlsson, 2008; Putman e Carbone, 2016, Budowle e Sajantila, 2024). Para algumas outras aplicações, a questão de dados ausentes é mais crítica, por exemplo, estimar o tamanho efetivo da população, N_e (Peel *et al.*, 2013).

Embora metodologias cada vez mais populares, como ddRADseq (Peterson *et al.*, 2012) ou GT-Seq (*Genotyping-by-the-Thousands* - Chang *et al.*, 2022), permitam a aplicação simultânea de até milhões de SNP, os *loci* de microssatélites ainda permanecerão uma classe de marcadores de DNA relevantes para uma série de aplicações rotineiras baseadas em informações genéticas como conservação, aquicultura, estudos de estratificação genética, análise de parentesco e etc (Hodel *et al.*, 2016). Alguns estudos apontam para uma resolução mais ou menos comparável de variabilidade e concordância de estrutura genética caracterizada entre centenas a milhões de SNP vs cerca de uma dúzia ou mais de microssatélites (*e.g.* Zimmerman *et al.*, 2020; Pérez-González *et al.*, 2023; Stacy *et al.*, 2025), apesar de apontarem aspectos diferenciais destas classes de marcadores favorecendo, geralmente, a superioridade de SNP (ver também Fischer *et al.*, 2017; Camacho-Sanchez *et al.*, 2020).

O número de alelos resolvidos em todos os 15 *loci* foi comparável à literatura anterior de caracterização ou aplicação de marcadores de microssatélites em populações naturais de *P. lineatus* (*e.g.* Yazbeck e Kalapothakis, 2007; Rueda *et al.*, 2011; Rueda *et al.*, 2013; e Perini *et al.*, 2021). Este novo conjunto de microssatélites, portanto, compreende marcadores de DNA muito

informativos, apropriados para algumas aplicações, como análises de parentesco ou diversidade, mesmo para amostras pequenas ou com alta taxa de dados faltantes (Budowle e Sajantila, 2024).

Esses microssatélites agora podem ser escurpulosamente avaliados por sequenciamento de Sanger ou ensaios NGS (Šarhanová *et al.*, 2018) para avaliar o seu grau médio de homoplasia (ou seja, alelos com o mesmo comprimento de fragmento, mas história ancestral independente) ou uso direto da informação integrada nas análises de estrutura genética populacional fina (*e.g.* Layton *et al.*, 2020).

Apesar de defender que o cientista deve sempre utilizar o melhor método possível (no caso de microssatélites, o uso de eletroforese capilar - EC - para precisão da estimativa de tamanho de fragmentos), argumento também pela solidez da abordagem criteriosa de PAGE. Isso se deve ao alto conteúdo de informação genética dos microssatélites (hipervariabilidade aliada a alta taxa de sinal/ruído), diante da limitação de recursos disponíveis e de sugestões da literatura de equivalência sobre estes métodos (PAGE vs EC) na avaliação de diversidade genética (*e.g.* Sánchez-Velásquez *et al.*, 2021).

Genética de populações

Um estudo de Wang *et al.* (2021) sugere que a aplicação de menos do que 11 ou 12 marcadores microssatélites pode levar a um viés importante na avaliação da estrutura genética populacional. Os resultados apresentados aqui constituem o maior esforço até agora, no que diz respeito ao número de *loci* microssatélites analisados, em *P. lineatus*. A avaliação da variabilidade genética revelou altos níveis de diversidade de microssatélites em peixes sob a influência de grandes barragens, da porção mais alta do rio Grande, MG (o extremo altitudinal da vertente oriental da bacia do Prata). Este é um resultado comum aos demais estudos anteriores de genética populacional em *P. lineatus*, com esforços de amostragem populacional comparáveis (*e.g.* Rueda *et al.*, 2013; Perini *et al.*, 2021). No geral, considereii nosso esforço de amostragem adequado para estimar representativamente as frequências alélicas dos *loci* (Hale *et al.*, 2012).

A análise da diversidade genética de amostras populacionais apontou para uma menor variabilidade genética para os cardumes à montante da barragem de Camargos, o último obstáculo artificial em direção à cabeceira do rio Grande e demais vales do alto da bacia, como o rio Aiuruoca, seu último grande tributário livre.

O alto rio Grande é uma área considerada prioritária para conservação de peixes migratórios no estado de Minas Gerais (Drummond, 2005) e está isolada do sistema à jusante pelos últimos 65 anos, desde a construção da barragem da usina hidrelétrica de Camargos. Não há passagem para peixes ali e larvas e juvenis provavelmente não a transpõem rio abaixo com facilidade (Suzuki *et al.*, 2011). Atividades de transposição de peixes, usando espécimes capturados à justante, e operações de peixamentos com estoques da piscicultura de Itutinga ocorreram na região nas últimas décadas, principalmente para o cumprimento de demandas de licenciamento ou no contexto de ações de educação ambiental.

Os resultados aqui apresentados, portanto, suportam a noção de que a barragem de Camargos está promovendo um importante isolamento de peixes do resto do sistema hidrográfico abaixo. Assim, conclui-se que as operações de peixamento e transposição manual podem ter sido ineficientes em favorecer a conectividade de fluxo gênico para o sistema à montante (falta de migrantes). Suas populações isoladas provavelmente estão sofrendo uma taxa de deriva genética aleatória alta, o que leva ao empobrecimento genético dentro dos demes e à diferenciação gradual entre eles com o passar do tempo. Essa noção é corroborada pelos resultados que mostram a maior quantidade de variação interpopulacional e diferenciação genética entre Camargos e todas as outras amostras populacionais. Este estudo, portanto, sustenta a hipótese de que as barragens estão isolando significativamente as populações de peixes. A DAPC realizada com informação de origem dos peixes também apoia essa interpretação, com base na inspeção visual da variação de microssatélites dentro e entre amostras populacionais.

A pequena variância entre as amostras populacionais registradas pela AMOVA é compatível com uma população única, quase panmítica (“*quasi-deme*”) ou, alternativamente, pode ocultar padrões mais complexos de variabilidade genética subjacentes à amostra total. No entanto, AMOVA e alguns índices de fixação (*e.g.* F_{ST} , G_{ST}) são considerados de interpretação pouco linear para marcadores hipervariáveis, como microssatélites (Jost, 2008), em termos de “diferenciação” genética. O índice de fixação G''_{ST} , que visa tornar a análise de fixação mais congruente com diferenciação, concorda razoavelmente bem com o estimador D de Jost, apontando para divergência relevante, em todas as comparações par-a-par. Isso apoia a hipótese de isolamento por barragens, sobre peixes impedidos de migrar livremente.

A amostra de 2023 de Itutinga é a mais diversa geneticamente (N_A e H_E), seguida por peixes capturados a jusante da barragem de Funil. Esta última, como levantado pela primeira vez em Yazbeck e Kalapothakis (2007), pode estar impedindo peixes migrantes, de estoques potencialmente diferentes, vindos de um trecho livre relativamente longo à jusante, de prosseguirem à montante, apesar do elevador de peixes. Este beco sem saída virtual pode estar levando a uma maior diversidade genética observada na amostra por acumulação, comparável ao efeito discutido por Perini *et al.* (2021). A montante desta barragem fica a foz de dois grandes tributários livres do rio Grande, que podem ser candidatos naturais de abrigos reprodutivos de diferentes estoques de *P. lineatus*, o rio Mortes e o rio Capivari (Suzuki *et al.*, 2011).

A população de *P. lineatus* no reservatório de Itutinga, por sua vez, sofreu variação genética e diferenciação de cerca de 7% em 10 anos. Apesar de ser fechado e fortemente isolado (sem tributários), parece ter acumulado variação genética local (riqueza alélica), desde 2013. Essa noção é contraintuitiva, uma vez que espera-se que populações pequenas e isoladas sofram perda de variabilidade genética (deriva).

Uma explicação plausível deste aparente paradoxo seria o efeito da piscicultura situada em uma de suas margens. A introdução repetida (peixamentos ou fuga) de peixes, oriundos de fontes

variadas de matrizes reprodutivas, e a tendência de ocorrência de gargalos genéticos, sobre eventos reprodutivos particulares (poucos casais), poderia estar produzindo coleções alternativas de alelos sendo introduzidos neste trecho fechado do rio Grande.

Embora este local, em 2023, tenha os maiores níveis de diversidade, também apresentou o maior coeficiente de endogamia, reforçando a noção de subestruturação induzida pela piscicultura. Por mais de cinco anos, no entanto, a estação de piscicultura opera principalmente para pesquisa científica e cessou a estocagem de peixes na região. Esses resultados apoiam a hipótese de que as pisciculturas podem influenciar a diversidade genética em populações de peixes aprisionados entre represas. Isso suscita a avaliação eventual da pertinência de retomada das operações de peixamento, orientadas por informações genéticas na bacia do alto rio Grande, MG.

Sobre a principal hipótese da pesquisa, da ocorrência de estoques mistos no alto rio Grande, MG, pode-se constatar que os resultados, no geral, são congruentes com um cenário de forte estratificação genética em todas as populações amostradas (*i.e.* há mais de um deme sendo amostrado). A interpretação se escora na observação de desvios consistentes em relação às proporções de Hardy-Weinberg em todas as amostras populacionais, sempre em direção ao déficit de heterozigotos, além de alguns casos de DL verificado. Eu considero esse resultado como uma expressão do efeito Wahlund, que implica na queda da heterozigosidade observada, em relação àquela esperada sob panmixia.

Análise hierárquica da estrutura genética populacional

O modelo de admistura é biologicamente informativo e aparentemente apropriado para se estudar peixes potamódromos Neotropicais. Esses peixes podem vivenciar múltiplas levas de piracema, ao longo de vários anos, o que pode resultar situações variáveis em relação ao momento e local da reprodução. Mais importante, indivíduos e estoques geneticamente distintos podem acabar potencialmente ocorrendo na mesma localidade e momento durante as subidas migratórias, ou

acabarem confinados entre represas adjacentes. Isso os isolaria de seus locais de reprodução originais, levando à exogamia subsequente entre diferentes demes (miscigenação).

A abordagem MLE para admistura estima parâmetros genéticos populacionais que maximizam sua função de verossimilhança (Wang, 2003; Wang, 2022) e tem se mostrado uma alternativa poderosa à hegemônica inferência Bayesiana. Esta é baseada nas probabilidades condicionais de genótipos, dado algum parâmetro genético populacional que é, por sua vez, altamente sensível às priores, muitas vezes subjetivas (van de Schoot *et al.*, 2021). Como os modelos de admistura são altamente hiperdimensionais (de 326 variáveis independentes, para $K=1$, até 6.384, para $K=14$, no caso estudado), a abordagem por MLE tem a vantagem de alta precisão, aliada a demandas computacionais mais modestas (Wang, 2022), o que me permitiu executar um número de réplicas mais expressivo. Houve alto grau de convergência entre réplicas (muitas têm variância zero) para diferentes K e concordância entre muitos ensaios alternativos (resultados não mostrados).

O estimador para K ótimo aqui adotado, D_{LK2} de Wang, é considerado geralmente mais confiável, em comparação com o estimador F_{STIS} (Wang, 2022), para uma gama de cenários. A análise MLE de admistura de 1ª ordem e a DAPC apontaram inicialmente para dois grupos, que não estão correlacionados com o esquema de amostragem. Ambos os *clusters* contêm membros de todas as populações amostradas. Isso sugere fortemente a ocorrência de estoques mistos de *P. lineatus* do leito principal do alto rio Grande. Descartei a hipótese de panmixia como a melhor possibilidade, examinando diretamente os valores de log-verossimilhança dos resultados das corridas $K=1$ contra o K escolhido. O D_{LK2} não capturaria $K=1$ (Wang, 2022), assim como o ΔK de Evanno para estimativa Bayesiana (Janes *et al.*, 2017), e o estimador F_{STIS} seria sempre zero nesse caso (Wang, 2022). Os resultados da análise Bayesiana de admistura também reforçam a noção de vários estoques de peixes diferentes em nossa amostra total, apoiando a hipótese da ocorrência de estoques mistos em todos pontos amostrados no leito do alto rio Grande, MG.

Diferentes valores de K revelam aspectos alternativos subjacentes à estrutura populacional e momentos distintos da história da população (Gilbert, 2016; Janes *et al.*, 2017; Gehri *et al.* 2021; Wang, 2022). Os dois principais grupos geneticamente diferenciados definidos na análise MLE de 1ª ordem podem ser o reflexo da existência de dois principais sítios de reprodução ancestrais, definidos por bifurcações demarcadas nas rotas migratórias (*e.g.* os rios Mortes, Capivari ou Aiuruoca), dividindo grupos no alto rio Grande. Ou, ainda, poderiam ser definidas por uma simples tendência assimétrica de navegação migratória lateral (preferência por acessos na margem direita vs esquerda).

A DAPC revelou alelos particulares de melhor desempenho para explicar a variação entre esses dois *clusters*, e a validação cruzada atestou sua robustez. No entanto, esse método pode ter dificuldades em cenários específicos, como com alta taxa de migração (como seria esperado em populações admisturadas) e baixo nível de diferenciação entre demes (como seria esperado de espécies com alta mobilidade), situações nas quais há uma tendência maior de subestimativa da presença de mais grupos internos (Miller *et al.*, 2020).

No conjunto, esses resultados confirmam a provável existência de pelo menos dois estoques mistos ao longo da área estudada, não explicados pela recente implementação de barragens. Evidências de estoques mistos em *P. lineatus* foram antecipadas, por exemplo, em Rueda *et al.* (2013); Avigliano *et al.* (2019) e indiretamente por Perini *et al.* (2021). A exploração Bayesiana do modelo de admistura sugeriu níveis mais profundos de subestruturação, apontando para até oito *clusters* possíveis na amostra total de peixes, um cenário compatível com os oito *clusters* definidos durante a “descamação” hierárquica de MLE da admistura.

Com o objetivo de aprofundar a exploração desses estoques geneticamente admistos de *P. lineatus* do alto rio Grande, os dois *clusters* inicialmente definidos por MLE foram desdobrados iterativamente, em mais duas rodadas de análises. Consistente e surpreendentemente, cada *cluster* examinado revelou dois novos subgrupos menores, aninhados dentro de *clusters* resolvidos na etapa

imediatamente anterior (2^3), como uma boneca matriosca. A julgar pelas amostras agrupadas em *clusters* definidos na Figura 8c, nas estimativas do F_{STIS} , ou na análise HWE dos *clusters* MLE, haveria potencialmente mais graus de admistura a serem desvendados, a partir de uma análise hierárquica mais aprofundada, embora a partir de tamanhos amostrais bastante críticos.

Os *clusters* definidos por MLE mostraram o comportamento esperado do modelo de admistura em direção às ordens superiores. A 3ª ordem teve mais homogeneidade genética dentro dos *clusters* e uma maximização da variação genética e distância entre os *clusters*. Também mostraram um número crescente de *loci* em EHW, atestando a significância da análise de admistura hierárquica e a utilidade desses novos marcadores microssatélites.

Estrutura genética populacional fractal

O padrão estratificado, autossimilar, inicialmente oculto e desvendado pela análise hierárquica da estrutura genética da amostra total de peixes parece revelar a existência de diferentes sítios de reprodução bem definidos, aninhados através de diferentes níveis de escala na bacia, dispostos de acordo com a geometria fractal (rio de 1ª ordem, tributários de 2ª e 3ª ordem e assim por diante). Isso me leva a concluir que *P. lineatus* provavelmente possui um alto grau de comportamento migratório filopátrico para sítios de procriação (*homing*). Esse é um tema comum em história natural de peixes e outros organismos migratórios e, apesar disso, ainda é muito pouco compreendido nos peixes dos Neotrópicos. Evidência de *homing* foi primeiramente levantada para *Prochilodus argenteus*, baseado em dados de telemetria (Godinho e Kynard, 2011) e discutida justamente para *P. lineatus* em Careaga e Carvajal-Vallejos (2019), a partir de dados morfométricos e lepidológicos (*i.e.* das escamas). Este trabalho reascende essa área de investigação para peixes migratórios na América do Sul.

Sendo particularmente abundante e competente na migração de longo alcance, vislumbro um passado anterior às barragens, onde *P. lineatus* provavelmente poderia alcançar virtualmente todos os tributários livres mais interiores das bacias, para reprodução. Considerando a amplitude da

colossal bacia do rio da Prata, que engloba uma diversidade de biomas e micro-habitats, isso provavelmente apontaria para o comportamento filopátrico como uma estratégia evolutivamente estável, devido à possível ocorrência de adaptações locais distintas, como tem sido demonstrada em outras espécies de peixes (*e.g.* Salles *et al.*, 2016; Mobley *et al.*, 2019; Vøllestad e Primmer, 2019).

Em um nível mais amplo, proponho aqui um modelo conceitual, uma hipótese, sobre a possível existência de camadas fractais da estrutura genética populacional em outros peixes migratórios, ou mesmo outros organismos migratórios, aquáticos ou não, diante de mistura temporária e sítios de reprodução segregados, sob regime predominante de fiolopatria (Figura 13).

Como fractais são conspícuos na paisagem (Mandelbrot, 1983), argumento por uma situação mais geral: espécies migratórias que partem de sítios de mistura temporária, como locais de alimentação e maturação, que apresentam, porém, afinidade filopátrica a sítios reprodutivos específicos em sua área total de ocorrência apresentariam esse padrão fractal da diversidade genética. Isso se daria supondo que estes organismos migrariam de forma razoavelmente direcionada por pistas migratórias (*e.g.* aves migratórias seguindo cursos d'água na bacia) ou por restrições de movimento (obstáculos migratórios) impostas por marcos dispostos em geometria fractal (autossimilaridade em diferentes escalas de magnificação).

Este modelo poderia abarcar diferentes variações em números, disposições de demes e etc. Sua principal predição é a observação deste padrão hierárquico fractal em diferentes escalas arbitrárias de consideração, através da área de distribuição de sítios reprodutivos, para sua estrutura genética de polimorfismos.

As únicas discussões diretas e explícitas sobre a relação dos fractais e a estrutura genética de populações, até onde sei, foram feita por Wingen *et al.* (2007), a respeito dos resultados de simulações de computador sobre a dispersão de longa distância de esporos de fungos, e por Lee (2020), no contexto da exploração da dimensão fractal em dados de variabilidade genética em humanos, dentro e entre genomas. Essa observação da natureza fractal hierárquica da estrutura

genética possivelmente encontrará aplicação e repercussão em outros campos diversos da pesquisa biológica.

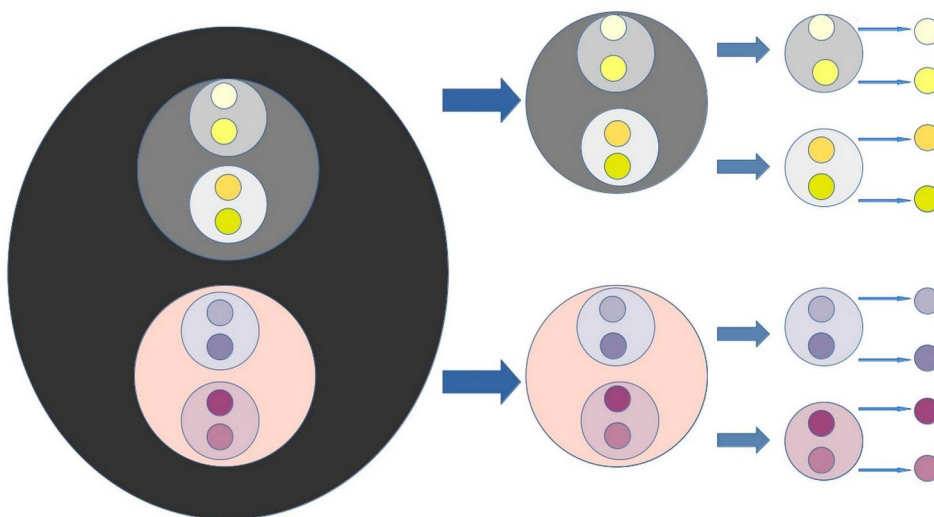


Figura 13: Ilustração esquemática do modelo conceitual geral de disposição fractal da estrutura da diversidade genética populacional em subpopulações misturadas (esquerda), com filopatria a sítios de reprodução segregados (à direita) em diferentes níveis de escala na paisagem. Círculos indicam agrupamentos genéticos de diferentes níveis inclusivos aninhados (escalas arbitrárias diferenciais de “demes” ou “quasi-demes”) as setas representam pistas ou restrições migratórias que seguem uma distribuição razoavelmente autossimilar e aninhada na paisagem.

Considerações finais

Na biologia da conservação, a compreensão dos graus de admistura é importante nas decisões de gestão (Wang, 2003; Wang, 2022). A confirmação de estoques mistos e da natureza hierarquicamente estratificada da variação genética de *P. lineatus* no leito principal de um rio intensamente represado provavelmente iluminará futuras pesquisas e sua gestão de recursos pesqueiros.

Caso seja corroborada essa descoberta teria implicação direta para a tomada de decisões de conservação e desenvolvimento em bacias Neotropicais. Ela destaca a relevância de tributários menores, ativamente procurados para a implementação de novas pequenas centrais hidrelétricas (Ferreira *et al.*, 2022), para a conservação da variação genética, em uma espécie de importância

socioambiental, pois destaca o papel potencial de tributários livres, de diferentes escalas, em toda a bacia, em abrigar estoques geneticamente distintos.

Em recente trabalho de incisivo impacto científico, McMillen e Levin (2024) argumentam pela relevância de um enfoque nas estruturas e processos aninhados, através de diferentes escalas, para um novo e mais amplo olhar sobre o fenômeno biológico. Na Era do *Big Data* (Schork, 2018), de LLM (Birhane *et al.*, 2023; Peng *et al.*, 2024), *deep learning* (e.g. Abramson *et al.*, 2024) e outros aspectos da IA, a investigação de questões genéticas, ecológicas e evolutivas cada vez mais profundas e complexas se tornarão mais e mais abordáveis, apontando para um novo cativante e promissor momento na Biologia.

Conclusão

- Este trabalho forneceu a primeira montagem genômica de referência para *Prochilodus lineatus*, junto com um pacote de recursos genéticos, incluindo um painel de microssatélites para o seu desenvolvimento rápido e de baixo custo;
- Foram confirmadas as hipóteses de que barragens influenciam a divergência genética entre localidades segregadas, além do efeito de pisciculturas na influência da diversidade genética de peixes aprisionados entre barragens;
- Foi corroborada a hipótese da ocorrência de cardumes mistos no leito principal do alto rio Grande, MG.
- Foi observada uma nova disposição fractal da diversidade genética, que implica em provável filopatria reprodutiva para *Prochilodus lineatus*;
- Foi apontado o possível papel de tributários de diferentes ordens em abrigar populações geneticamente diversas de peixes migratórios de água doce;
- Estes resultados possivelmente beneficiarão pesquisadores e gestores ambientais no enfrentamento de questões desafiadoras relacionadas à conservação de peixes migratórios Neotropicais e desenvolvimento; e finalmente

- Foi indutivamente proposto aqui um novo modelo geral para a disposição fractal da estrutura genética populacional, em organismos migratórios exibindo mistura temporária, porém com filopatria a sítios separados de procriação.

Referências bibliográficas

- ABRAMSON, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. **Nature**, [S. l.], vol. 630, no. 8016, p. 493–500, Jun. 2024.
- AGOSTINHO, A. A. *et al.* Fish diversity in the upper Paraná River basin: habitats, fisheries, management and conservation. **Aquatic Ecosystem Health & Management**, [S. l.], vol. 10, no. 2, p. 174–186, 8 Jun. 2007.
- ALEXANDER, D. H.; NOVEMBRE, J.; LANGE, K. Fast model-based estimation of ancestry in unrelated individuals. **Genome Research**, Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press publisher: Cold Spring Harbor Lab PMID: 19648217, vol. 19, no. 9, p. 1655–1664, 1 Sep. 2009.
- ALIX, M.; KJESBU, O. S.; ANDERSON, K. C. From gametogenesis to spawning: How climate-driven warming affects teleost reproductive biology. **Journal of Fish Biology**, [S. l.], vol. 97, no. 3, p. 607–632, Sep. 2020.
- ALLENDORF, F. W. Genetics and the conservation of natural populations: allozymes to genomes. **Molecular Ecology**, [S. l.], vol. 26, no. 2, p. 420–430, Jan. 2017.
- ALTSCHUL, S. F. *et al.* Basic local alignment search tool. **Journal of Molecular Biology**, [S. l.], vol. 215, no. 3, p. 403–410, 5 Oct. 1990.
- ANDREWS, S. FastQC A Quality Control tool for High Throughput Sequence Data. 2010. Available at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed on: 26 Jun. 2024.
- ANDRONACHE, I. *et al.* Dynamics of Forest Fragmentation and Connectivity Using Particle and Fractal Analysis. **Scientific Reports**, [S. l.], vol. 9, no. 1, p. 12228, 22 Aug. 2019.

- ANITAS, E. M. Fractal Analysis of DNA Sequences Using Frequency Chaos Game Representation and Small-Angle Scattering. **International Journal of Molecular Sciences**, [S. l.], vol. 23, no. 3, p. 1847, 6 Feb. 2022.
- ASTASHYN, A. *et al.* Rapid and sensitive detection of genome contamination at scale with FCS-GX. **Genome Biology**, [S. l.], vol. 25, no. 1, p. 60, 26 Feb. 2024.
- AVIGLIANO, E. *et al.* Mixed-stock and discriminant models use for assessing recruitment sources of estuarine fish populations in La Plata Basin (South America). **Journal of the Marine Biological Association of the United Kingdom**, [S. l.], vol. 99, no. 6, p. 1429–1433, Sep. 2019.
- BAGGIO, R. A. *et al.* Dams cause genetic homogenization in populations of fish that present homing behavior: Evidence from a demogenetic individual-based model. **Ecological Modelling**, [S. l.], vol. 384, p. 209–220, 24 Sep. 2018.
- BANCAUD, A. *et al.* A fractal model for nuclear organization: current evidence and biological implications. **Nucleic Acids Research**, [S. l.], vol. 40, no. 18, p. 8783–8792, 1 Oct. 2012.
- BARBOSA, A. C. D. R. *et al.* Thirteen polymorphic microsatellite *loci* in the Neotropical fish *Prochilodus argenteus* (Characiformes, Prochilodontidae). **Molecular Ecology Notes**, [S. l.], vol. 6, no. 3, p. 936–938, 2006.
- BARTHEM, R. B. *et al.* Goliath catfish spawning in the far western Amazon confirmed by the distribution of mature adults, drifting larvae and migrating juveniles. **Scientific Reports**, [S. l.], vol. 7, no. 1, p. 41784, 6 Feb. 2017.
- BASSINGTHWAIGHTE, J. B.; LIEBOVITCH, L. S.; WEST, B. J. **Fractal Physiology**. [S. l.]: Oxford University Press, 1994. Available at: <https://doi.org/10.1093/oso/9780195080131.001.0001>. Accessed on: 25 Mar. 2025.
- BATISTA, J. S.; ALVES-GOMES, J. A. Phylogeography of *Brachyplatystoma rousseauxii* (Siluriformes - Pimelodidae) in the Amazon Basin offers preliminary evidence for the first case of “homing” for an Amazonian migratory catfish. **Genetics and molecular research**, [S. l.], vol. 5, no. 4, p. 723–740, 1 Dec. 2006.
- BEAUMONT, M. A.; RANNALA, B. The Bayesian revolution in genetics. **Nature Reviews Genetics**, [S. l.], vol. 5, no. 4, p. 251–261, Apr. 2004.

- BERNOS, T. A.; JEFFRIES, K. M.; MANDRAK, N. E. Linking genomics and fish conservation decision making: a review. **Reviews in Fish Biology and Fisheries**, [S. l.], vol. 30, no. 4, p. 587–604, 1 Dec. 2020.
- BIRHANE, A. *et al.* Science in the age of large language models. **Nature Reviews Physics**, [S. l.], vol. 5, no. 5, p. 277–280, May 2023.
- BOWEN, S. H. Digestion and assimilation of benthic biofilm by the Sábalo, *Prochilodus lineatus*. **Journal of Fish Biology**, [S. l.], vol. 100, no. 1, p. 107–116, Jan. 2022a.
- BOWEN, S. H. The river flood pulse, benthic biofilm, and the nutrition of *Prochilodus lineatus*. **Environmental Biology of Fishes**, [S. l.], vol. 105, no. 2, p. 213–230, 1 Feb. 2022b.
- BRADNAM, K. R. *et al.* Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. **GigaScience**, [S. l.], vol. 2, no. 1, p. 10, 22 Jul. 2013.
- BRAGA-SILVA, A.; GALETTI, P. M. Evidence of isolation by time in freshwater migratory fish *Prochilodus costatus* (Characiformes, Prochilodontidae). **Hydrobiologia**, [S. l.], vol. 765, no. 1, p. 159–167, 1 Feb. 2016.
- BRANDSTRÖM, M.; ELLEGREN, H. Genome-wide analysis of microsatellite polymorphism in chicken circumventing the ascertainment bias. **Genome Research**, [S. l.], vol. 18, no. 6, p. 881–887, Jun. 2008.
- BUDOWLE, B.; SAJANTILA, A. Short tandem repeats - how microsatellites became the currency of forensic genetics. **Nature Reviews. Genetics**, [S. l.], vol. 25, no. 7, p. 450–451, Jul. 2024.
- BUSHNELL, B. **BBMap: A Fast, Accurate, Splice-Aware Aligner**, n. LBNL-7065E. Walnut Creek: Lawrence Berkeley National Laboratory, 19 Mar. 2014. Available at: <https://escholarship.org/uc/item/1h3515gn>. Accessed on: 19 Aug. 2024.
- CAMACHO-SANCHEZ, M. *et al.* Comparative assessment of range-wide patterns of genetic diversity and structure with SNPs and microsatellites: A case study with Iberian amphibians. **Ecology and Evolution**, [S. l.], vol. 10, no. 19, p. 10353–10363, 15 Sep. 2020.
- CAREAGA, M.; CARVAJAL-VALLEJOS, F. Los sábalo (Characiformes: Prochilodontidae: *Prochilodus*) de Bolivia. [S. l.: s. n.], 2019. p. 171–185.

- CARLSSON, J. Effects of microsatellite null alleles on assignment testing. **The Journal of Heredity**, [S. l.], vol. 99, no. 6, p. 616–623, 2008.
- CARMO, A. O. *et al.* Complete mitochondrial genome sequence of *Prochilodus lineatus* (Characiformes, Prochilodontidae). **Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis**, [S. l.], vol. 27, no. 3, p. 1946–1947, May 2016.
- CARMO, F. M. da S. *et al.* Optimization of heterologous microsatellites in Piracanjuba. **Pesquisa Agropecuária Brasileira**, [S. l.], vol. 50, p. 1236–1239, Dec. 2015.
- CAROLSFELD, J. *et al.* **Migratory Fishes of South America: Biology, Fisheries and Conservation Status**. [S. l.]: IDRC, 2003.
- CARVAJAL-VALLEJOS, F. M. *et al.* Genetic structure in the Amazonian catfish *Brachyplatystoma rousseauxii*: influence of life history strategies. **Genetica**, [S. l.], vol. 142, no. 4, p. 323–336, Aug. 2014.
- CARVALHO, L. C.; RIBEIRO, J. M.; YAZBECK, G. M. Microsatellite multiplex systems for *Brycon orbignyanus* (Characiformes: Bryconidae). **Brazilian Journal of Biology**, [S. l.], vol. 79, no. 4, p. 751–754, Nov. 2019.
- CASSEMIRO, F. A. S. *et al.* Landscape dynamics and diversification of the megadiverse South American freshwater fish fauna. **Proceedings of the National Academy of Sciences**, [S. l.], vol. 120, no. 2, p. e2211974120, 10 Jan. 2023.
- CASTRO, R.; VARI, R. Prochilodontidae (Fannel mouth characiforms). **Checklist of the Freshwater Fishes of South and Central America. Porto Alegre: EDIPUCRS, Brasil**, [S. l.], p. 65–70, 2003.
- CHAGAS, A. T. D. A. *et al.* Description and comparison of two economically important fish species mitogenomes: *Prochilodus argenteus* and *Prochilodus costatus* (Characiformes, Prochilodontidae). **Mitochondrial DNA Part A**, [S. l.], vol. 27, no. 4, p. 2852–2853, 3 Jul. 2016.
- CHALLIS, R. **rchallis/assembly-stats 17.02**. [S. l.]: Zenodo, 24 Feb. 2017. Available at: <https://zenodo.org/records/322347>. Accessed on: 13 Aug. 2024.

- CHANG, S. L. *et al.* Genotyping-in-Thousands by sequencing panel development and application for high-resolution monitoring of introgressive hybridization within sockeye salmon. **Scientific Reports**, [S. l.], vol. 12, no. 1, p. 3441, 2 Mar. 2022.
- COLLINS, S. M. *et al.* Population Structure of a Neotropical Migratory Fish: Contrasting Perspectives from Genetics and Otolith Microchemistry. **Transactions of the American Fisheries Society**, [S. l.], vol. 142, no. 5, p. 1192–1201, 1 Sep. 2013.
- COMTE, L.; OLDEN, J. D. Climatic vulnerability of the world's freshwater and marine fishes. **Nature Climate Change**, [S. l.], vol. 7, no. 10, p. 718–722, Oct. 2017.
- DAVIS, C. D. *et al.* Refining and defining riverscape genetics: How rivers influence population genetic structure. **Wiley Interdisciplinary Reviews: Water**, [S. l.], vol. 5, no. 2, p. e1269, 1 Mar. 2018.
- DRUMMOND, G. M. **Biodiversidade em Minas Gerais: um atlas para sua conservação**. Belo Horizonte: Fundação Biodiversitas, 2005.
- ELLEGREN, H. Microsatellites: simple sequences with complex evolution. **Nature Reviews. Genetics**, [S. l.], vol. 5, no. 6, p. 435–445, Jun. 2004.
- ENDRUWEIT, M. Four new records of fish species (Cypriniformes: Nemacheilidae, Balitoridae; Characiformes: Prochilodontidae) and corrections of two misidentified fish species (Tetraodontiformes: Tetraodontidae; Beloniformes: Belonidae) in Yunnan, China. **Dong Wu Xue Yan Jiu = Zoological Research**, [S. l.], vol. 35, no. 1, p. 51–58, Jan. 2014.
- ESPINDOLA, I. B.; DE LEITE, M. L. T. A.; RIBEIRO, W. C. South-American Transboundary Waters: The Management of the Guarani Aquifer System and the La Plata Basin Towards the Future. In: BREARS, R. C. (ed.). **The Palgrave Handbook of Climate Resilient Societies**. Cham: Springer International Publishing, 2021. p. 251–285. Available at: https://doi.org/10.1007/978-3-030-42462-6_51. Accessed on: 28 Apr. 2024.
- EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. **Molecular Ecology**, [S. l.], vol. 14, no. 8, p. 2611–2620, Jul. 2005.
- EWELS, P. *et al.* MultiQC: summarize analysis results for multiple tools and samples in a single report. **Bioinformatics**, [S. l.], vol. 32, no. 19, p. 3047–3048, 1 Oct. 2016.

- FALUSH, D.; STEPHENS, M.; PRITCHARD, J. K. Inference of population structure using multilocus genotype data: dominant markers and null alleles. **Molecular Ecology Notes**, [S. l.], vol. 7, no. 4, p. 574–578, 1 Jul. 2007.
- FALUSH, D.; STEPHENS, M.; PRITCHARD, J. K. Inference of population structure using multilocus genotype data: linked *loci* and correlated allele frequencies. **Genetics**, [S. l.], vol. 164, no. 4, p. 1567–1587, Aug. 2003.
- FENERICH-VERANI, N.; GODINHO, H. M.; NARAHARA, M. Y. The size composition of the eggs of curimatá, *Prochilodus scrofa* Steindachner 1881, induced to spawn with human chorionic gonadotropin (HCG). **Aquaculture**, [S. l.], vol. 42, no. 1, p. 37–41, 1 Nov. 1984.
- FERREIRA, Dhiego G. *et al.* Different stocks of *Prochilodus lineatus* along the Cinzas River basin: contributions from a free-flowing tributary in the Capivara Reservoir region, Paranapanema River basin (upper Paraná River). **Neotropical Ichthyology**, [S. l.], vol. 21, no. 4, p. e230080, 2023.
- FERREIRA, Dhiego Gomes *et al.* Genetic structure and diversity of migratory freshwater fish in a fragmented Neotropical river system. **Reviews in Fish Biology and Fisheries**, [S. l.], vol. 27, no. 1, p. 209–231, 1 Mar. 2017.
- FERREIRA, M. E. *et al.* Dams Pose a Critical Threat to Rivers in Brazil's Cerrado Hotspot. **Water**, [S. l.], vol. 14, no. 22, p. 3762, Jan. 2022.
- FISCHER, M. C. *et al.* Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. **BMC Genomics**, [S. l.], vol. 18, no. 1, p. 69, 11 Jan. 2017.
- FRABLE, B. W. *et al.* Biogeographic reconstruction of the migratory Neotropical fish family Prochilodontidae (Teleostei: Characiformes). **Zoologica Scripta**, [S. l.], vol. 51, no. 3, p. 348–364, May 2022.
- GARCÍA-CASTRO, K. L.; MÁRQUEZ, E. J. Temporal-scale assessment of population genetics of the freshwater fish *Prochilodus magdalenae* in an area impacted by construction of a dam. **Hydrobiologia**, [S. l.], vol. 851, no. 6, p. 1513–1531, 1 Mar. 2024.
- GEHRI, R. R.; GRUENTHAL, K.; LARSON, W. A. It's complicated: Heterogeneous patterns of genetic structure in five fish species from a fragmented river suggest multiple processes can

- drive differentiation. **Evolutionary Applications**, [S. l.], vol. 14, no. 8, p. 2079–2097, Aug. 2021.
- GILBERT, K. J. Identifying the number of population clusters with structure: problems and solutions. **Molecular Ecology Resources**, [S. l.], vol. 16, no. 3, p. 601–603, 2016.
- GODINHO, A. L.; KYNARD, B. Migration and Spawning of Radio-Tagged Zulega *Prochilodus argenteus* in a Dammed Brazilian River. **Transactions of the American Fisheries Society**, [S. l.], vol. 135, no. 3, p. 811–824, 1 May 2006.
- GODOY, M. P. **Contribuição à história natural e geral de Pirassununga**. Pirassununga: Prefeitura Municipal, 1970.
- GRACIANO, R. *et al.* Functional DNA annotation from a preliminary de novo genome assembly of *Brycon orbignyanus*, an endangered Neotropical migratory fish. **Latin American Data in Science**, [S. l.], vol. 1, no. 2, p. 42–48, 12 Nov. 2021.
- GRACIANO, R. C. D. *et al.* Genomic Resources for *Salminus brasiliensis*. **Frontiers in Genetics**, [S. l.], vol. 13, 28 Mar. 2022. Available at: <https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2022.855718/full>. Accessed on: 17 Jun. 2024.
- HALE, M. L.; BURG, T. M.; STEEVES, T. E. Sampling for Microsatellite-Based Population Genetic Studies: 25 to 30 Individuals per Population Is Enough to Accurately Estimate Allele Frequencies. **PLOS ONE**, [S. l.], vol. 7, no. 9, p. e45170, 12 Sep. 2012.
- HALLERMAN, E. M. (Org.). **Population genetics: principles and applications for fisheries scientists**. Bethesda, MD (USA): American Fisheries Society, 2003. Available at: <https://agris.fao.org/search/en/providers/122621/records/647396cfe01106880097fed4>. Accessed on: 8 Jul. 2024.
- HATANAKA, T.; HENRIQUE-SILVA, F.; GALETTI, P. M. Population Substructuring in a Migratory Freshwater Fish *Prochilodus argenteus* (Characiformes, Prochilodontidae) from the São Francisco River. **Genetica**, [S. l.], vol. 126, no. 1, p. 153–159, 1 Jan. 2006.
- HODEL, R. G. J. *et al.* The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st century. **Applications in Plant Sciences**, [S. l.], vol. 4, no. 6, p. apps.1600025, Jun. 2016.

- HOWE, N. S. *et al.* Genomic evidence for domestication selection in three hatchery populations of Chinook salmon, *Oncorhynchus tshawytscha*. **Evolutionary Applications**, [S. l.], vol. 17, no. 2, p. e13656, Feb. 2024.
- ILOSVAY, X. É. E.; MOLINOS, J. G.; OJEA, E. Stronger adaptive response among small-scale fishers experiencing greater climate change hazard exposure. **Communications Earth & Environment**, [S. l.], vol. 3, no. 1, p. 1–9, 20 Oct. 2022.
- JANES, J. K. *et al.* The K = 2 conundrum. **Molecular Ecology**, [S. l.], vol. 26, no. 14, p. 3594–3602, Jul. 2017.
- JOMBART, T. adegenet: a R package for the multivariate analysis of genetic markers. **Bioinformatics**, [S. l.], vol. 24, no. 11, p. 1403–1405, 1 Jun. 2008.
- JOMBART, T.; DEVILLARD, S.; BALLOUX, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. **BMC genetics**, [S. l.], vol. 11, p. 94, 15 Oct. 2010.
- JOST, L. G_{ST} and its relatives do not measure differentiation. **Molecular Ecology**, [S. l.], vol. 17, no. 18, p. 4015–4026, 2008.
- JUNG, H. *et al.* Twelve quick steps for genome assembly and annotation in the classroom. **PLOS Computational Biology**, [S. l.], vol. 16, no. 11, p. e1008325, 12 Nov. 2020.
- KALOUS, L. *et al.* The South American freshwater fish *Prochilodus lineatus* (Actinopterygii: Characiformes: Prochilodontidae): new species in Vietnamese aquaculture: *Prochilodus lineatus* in Vietnam. **Aquaculture Research**, [S. l.], vol. 43, no. 6, p. 955–958, May 2012.
- KOPELMAN, N. M. *et al.* Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. **Molecular Ecology Resources**, [S. l.], vol. 15, no. 5, p. 1179–1191, 2015.
- LANDÍNEZ-GARCÍA, R. M.; MARQUEZ, E. J. Microsatellite loci development and population genetics in Neotropical fish *Curimata mivartii* (Characiformes: Curimatidae). **PeerJ**, [S. l.], vol. 6, p. e5959, 13 Nov. 2018.

- LANDÍNEZ-GARCÍA, R. M.; NARVÁEZ, J. C.; MÁRQUEZ, E. J. Population genetics of the freshwater fish *Prochilodus magdalenae* (Characiformes: Prochilodontidae), using species-specific microsatellite *loci*. **PeerJ**, [S. l.], vol. 8, p. e10327, 11 Nov. 2020.
- LAWSON, D. J.; VAN DORP, L.; FALUSH, D. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. **Nature Communications**, [S. l.], vol. 9, no. 1, p. 3258, 14 Aug. 2018.
- LAYTON, K. K. S. *et al.* Resolving fine-scale population structure and fishery exploitation using sequenced microsatellites in a northern fish. **Evolutionary Applications**, [S. l.], vol. 13, no. 5, p. 1055–1068, 2020.
- LEE, C.-Y. The fractal dimension as a measure for characterizing genetic variation of the human genome. **Computational Biology and Chemistry**, [S. l.], vol. 87, p. 107278, 1 Aug. 2020.
- LEGGETT, S. E. *et al.* Motility-limited aggregation of mammary epithelial cells into fractal-like clusters. **Proceedings of the National Academy of Sciences**, [S. l.], vol. 116, no. 35, p. 17298–17306, 27 Aug. 2019.
- LEI, Y. *et al.* Genome-wide characterization of microsatellite DNA in fishes: survey and analysis of their abundance and frequency in genome-specific regions. **BMC Genomics**, [S. l.], vol. 22, no. 1, p. 421, 7 Jun. 2021.
- LI, D. *et al.* MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. **Bioinformatics (Oxford, England)**, [S. l.], vol. 31, no. 10, p. 1674–1676, 15 May 2015.
- LI, H. *et al.* The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, [S. l.], vol. 25, no. 16, p. 2078–2079, 15 Aug. 2009.
- LI, Y.-L.; LIU, J.-X. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. **Molecular Ecology Resources**, [S. l.], vol. 18, no. 1, p. 176–177, 2018.
- LIU, C.-C. *et al.* Exploring Population Structure with Admixture Models and Principal Component Analysis. In: DUTHEIL, J. Y. (ed.). **Statistical Population Genomics**. New York, NY: Springer US, 2020. p. 67–86. Available at: https://doi.org/10.1007/978-1-0716-0199-0_4. Accessed on: 29 Jun. 2024.

- LUO, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. **GigaScience**, [S. l.], vol. 1, p. 18, 27 Dec. 2012.
- MANDELBROT, Benoît B. **Les objets fractals: forme, hasard et dimension**. Paris: Flammarion, 1975.
- MANDELBROT, Benoit B. **The Fractal Geometry of Nature**. [S. l.]: Henry Holt and Company, 1983.
- MARÇAIS, G.; KINGSFORD, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. **Bioinformatics**, [S. l.], vol. 27, no. 6, p. 764–770, 15 Mar. 2011.
- MCCOMISH, B. J. *et al.* Ancient and Modern Genomes Reveal Microsatellites Maintain a Dynamic Equilibrium Through Deep Time. **Genome Biology and Evolution**, [S. l.], vol. 16, no. 3, p. evae017, 1 Mar. 2024.
- MCMILLEN, P.; LEVIN, M. Collective intelligence: A unifying concept for integrating biology across scales and substrates. **Communications Biology**, [S. l.], vol. 7, no. 1, p. 1–17, 28 Mar. 2024.
- MELO, B. F. *et al.* Little Divergence Among Mitochondrial Lineages of *Prochilodus* (Teleostei, Characiformes). **Frontiers in Genetics**, [S. l.], vol. 9, p. 107, 4 Apr. 2018.
- MILLER, J. M.; CULLINGHAM, C. I.; PEERY, R. M. The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. **Heredity**, [S. l.], vol. 125, no. 5, p. 269–280, Nov. 2020.
- MOBLEY, K. B. *et al.* Home ground advantage: Local Atlantic salmon have higher reproductive fitness than dispersers in the wild. **Science Advances**, [S. l.], vol. 5, no. 2, p. eaav1112, 27 Feb. 2019.
- NASCIMENTO, G. A. S. **Diversidade genética de *Prochilodus lineatus* em regiões sob influência de barragens de hidrelétricas no sistema do alto rio Grande – MG**. 2024. 125 f. Dissertação – Universidade Federal de São João Del Rei, São João Del Rei, 2024.
- NGUYEN, T. H. T. *et al.* Modelling Tools to Analyze and Assess the Ecological Impact of Hydropower Dams. **Water**, [S. l.], vol. 10, no. 3, p. 259, Mar. 2018.

- OLIVEIRA, C.; FORESTI, F.; HILSDORF, A. W. S. Genetics of neotropical fish: from chromosomes to populations. **Fish Physiology and Biochemistry**, [S. l.], vol. 35, no. 1, p. 81–100, Mar. 2009.
- OLIVEIRA-FARIAS, K. K. C. *et al.* Genetic structure and management of the Neotropical migratory fish *Megaleporinus obtusidens* on a highly impacted river basin. **Hydrobiologia**, [S. l.], vol. 849, no. 7, p. 1645–1658, 1 Apr. 2022.
- OSATHANUNKUL, M.; SUWANNAPOOM, C. Sustainable fisheries management through reliable restocking and stock enhancement evaluation with environmental DNA. **Scientific Reports**, [S. l.], vol. 13, no. 1, p. 11297, 12 Jul. 2023.
- PAULS, E.; BERTOLLO, L. A. C. Evidence for a System of Supernumerary Chromosomes in *Prochilodus scrofa* Steindachner, 1881 (Pisces, Prochilodontidae). **Caryologia**, [S. l.], vol. 36, no. 4, p. 307–314, 1 Jan. 1983.
- PEAKALL, R.; SMOUSE, P. E. GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. **Bioinformatics**, [S. l.], vol. 28, no. 19, p. 2537–2539, 1 Oct. 2012.
- PEEL, D. *et al.* Accounting for missing data in the estimation of contemporary genetic effective population size ($N(e)$). **Molecular Ecology Resources**, [S. l.], vol. 13, no. 2, p. 243–253, Mar. 2013.
- PENG, D. *et al.* Large-language models facilitate discovery of the molecular signatures regulating sleep and activity. **Nature Communications**, [S. l.], vol. 15, no. 1, p. 3685, 1 May 2024.
- PEREIRA, L. H. G.; FORESTI, F.; OLIVEIRA, C. Genetic structure of the migratory catfish *Pseudoplatystoma corruscans* (Siluriformes: Pimelodidae) suggests homing behaviour. **Ecology of Freshwater Fish**, [S. l.], vol. 18, no. 2, p. 215–225, 2009.
- PÉREZ-GONZÁLEZ, J. *et al.* Comparative Analysis of Microsatellite and SNP Markers for Genetic Management of Red Deer. **Animals : an Open Access Journal from MDPI**, [S. l.], vol. 13, no. 21, p. 3374, 31 Oct. 2023.
- PERINI, V. R. *et al.* Metapopulation dynamics of the migratory fish *Prochilodus lineatus* (Characiformes: Prochilodontidae) in a lotic remnant of the Grande River, Southeastern Brazil. **Neotropical Ichthyology**, [S. l.], vol. 19, no. 4, p. e200046, 2021.

- PETERSON, B. K. *et al.* Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. **PLOS ONE**, [S. l.], vol. 7, no. 5, p. e37135, 31 May 2012.
- PETRY, A. C.; AGOSTINHO, A. A.; GOMES, L. C. Fish assemblages of tropical floodplain lagoons: exploring the role of connectivity in a dry year. **Neotropical Ichthyology**, [S. l.], vol. 1, p. 111–119, Dec. 2003.
- PFLUG, J. M. *et al.* Measuring Genome Sizes Using Read-Depth, k-mers, and Flow Cytometry: Methodological Comparisons in Beetles (Coleoptera). **G3 (Bethesda, Md.)**, [S. l.], vol. 10, no. 9, p. 3047–3060, 2 Sep. 2020.
- PIMENTEL, J. da S. M. *et al.* Genetic evaluation of migratory fish: Implications for conservation and stocking programs. **Ecology and Evolution**, [S. l.], vol. 10, no. 19, p. 10314–10324, 2020.
- PIORSKI, N. M. *et al.* Contribution of conservation genetics in assessing neotropical freshwater fish biodiversity. **Brazilian Journal of Biology**, [S. l.], vol. 68, p. 1039–1050, Nov. 2008.
- PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, [S. l.], vol. 155, no. 2, p. 945–959, Jun. 2000.
- PUTMAN, A. I.; CARBONE, I. Challenges in analysis and interpretation of microsatellite data for population genetic studies. **Ecology and Evolution**, [S. l.], vol. 4, no. 22, p. 4399–4428, 2014.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing, 14 2024. Available at: <https://www.R-project.org>.
- REIS, R. E. *et al.* Fish biodiversity and conservation in South America. **Journal of Fish Biology**, [S. l.], vol. 89, no. 1, p. 12–47, Jul. 2016.
- RIBOLLI, J. *et al.* Genetic evidence supports polygamous mating system in a wild population of *Prochilodus lineatus* (Characiformes: Prochilodontidae), a Neotropical shoal spawner fish. **Neotropical Ichthyology**, [S. l.], vol. 18, p. e190123, 26 Jun. 2020.
- RIBOLLI, J. *et al.* Isolation-by-time population structure in potamodromous Dourado *Salminus brasiliensis* in southern Brazil. **Conservation Genetics**, [S. l.], vol. 18, no. 1, p. 67–76, 1 Feb. 2017.

- RICHARDSON, L. F. The problem of contiguity : An appendix to statistics of deadly quarrels. **General Systems Yearbook**, [S. l.], vol. 6, p. 139–187, 1961.
- ROGNES, T. Faster Smith-Waterman database searches with inter-sequence SIMD parallelisation. **BMC Bioinformatics**, [S. l.], vol. 12, no. 1, p. 221, 1 Jun. 2011.
- ROSA, I. F. D. *et al.* Temporal genetic structure of a stock of *Prochilodus lineatus* (Characiformes: Prochilodontidae) in the Mogi-Guaçu River ecosystem, São Paulo, Brazil. **Neotropical Ichthyology**, [S. l.], vol. 20, no. 2, p. e210156, 2022.
- ROUSSET, F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. **Molecular Ecology Resources**, [S. l.], vol. 8, no. 1, p. 103–106, 2008.
- RUEDA, Eva C. *et al.* Isolation and characterization of polymorphic microsatellite *loci* in the migratory freshwater fish *Prochilodus lineatus* (Characiformes: Prochilodontidae). **Conservation Genetics Resources**, [S. l.], vol. 3, no. 4, p. 681–684, 1 Oct. 2011.
- RUEDA, Eva Carolina *et al.* Seasonal variation in genetic population structure of sábalo (*Prochilodus lineatus*) in the Lower Uruguay River. **Genetica**, [S. l.], vol. 141, no. 7–9, p. 401–407, Sep. 2013.
- RUZICH, J. *et al.* Isolation by a hydroelectric dam induces minimal impacts on genetic diversity and population structure in six fish species. **Conservation Genetics**, [S. l.], vol. 20, no. 6, p. 1421–1436, 1 Dec. 2019.
- SALLES, O. C. *et al.* First genealogy for a wild marine fish population reveals multigenerational philopatry. **Proceedings of the National Academy of Sciences**, [S. l.], vol. 113, no. 46, p. 13245–13250, 15 Nov. 2016.
- SÁNCHEZ-VELÁSQUEZ, J. J. *et al.* Performance comparison of gel and capillary electrophoresis-based microsatellite genotyping strategies in a population research and kinship testing framework. **BMC Research Notes**, [S. l.], vol. 14, no. 1, p. 444, 7 Dec. 2021a.
- SANTOS, R. P. *et al.* Diversification of *Prochilodus* in the eastern Brazilian Shield: Evidence from complete mitochondrial genomes (Teleostei, Prochilodontidae). **Journal of Zoological Systematics and Evolutionary Research**, [S. l.], vol. 59, no. 5, p. 1053–1063, Jul. 2021.

- SARAIVA, S. O.; POMPEU, P. S. Fish hatching and its effects on the morphology of *Prochilodus lineatus* (Actinopterygii: Prochilodontidae). **Brazilian Journal of Biology**, [S. l.], vol. 76, no. 1, p. 209–217, 22 Jan. 2016.
- ŠARHANOVÁ, P. *et al.* SSR-seq: Genotyping of microsatellites using next-generation sequencing reveals higher level of polymorphism as compared to traditional fragment size scoring. **Ecology and Evolution**, [S. l.], vol. 8, no. 22, p. 10817–10833, Nov. 2018.
- SAYERS, E. W. *et al.* Database resources of the national center for biotechnology information. **Nucleic Acids Research**, [S. l.], vol. 50, no. D1, p. D20–D26, 7 Jan. 2022.
- SAYERS, E. W.; O’SULLIVAN, C.; KARSCH-MIZRACHI, I. Using GenBank and SRA. **Methods in Molecular Biology (Clifton, N.J.)**, [S. l.], vol. 2443, p. 1–25, 2022.
- SCHLÖTTERER, C. The evolution of molecular markers--just a matter of fashion? **Nature Reviews. Genetics**, [S. l.], vol. 5, no. 1, p. 63–69, Jan. 2004.
- SCHORK, N. J. The big data revolution and human genetics. **Human Molecular Genetics**, [S. l.], vol. 27, no. R1, p. R1, 1 May 2018.
- SENDKER, F. L. *et al.* Emergence of fractal geometries in the evolution of a metabolic enzyme. **Nature**, [S. l.], vol. 628, no. 8009, p. 894–900, Apr. 2024.
- SHEDD, K. R. *et al.* Reduced relative fitness in hatchery-origin Pink Salmon in two streams in Prince William Sound, Alaska. **Evolutionary Applications**, [S. l.], vol. 15, no. 3, p. 429–446, Mar. 2022.
- SHENDURE, J. *et al.* DNA sequencing at 40: past, present and future. **Nature**, [S. l.], vol. 550, no. 7676, p. 345–353, 19 Oct. 2017.
- SIVASUNDAR, A.; BERMINGHAM, E.; ORTÍ, G. Population structure and biogeography of migratory freshwater fishes (*Prochilodus*: Characiformes) in major South American rivers. **Molecular Ecology**, [S. l.], vol. 10, no. 2, p. 407–417, Feb. 2001.
- SMOUSE, P. E.; WAPLES, R. S.; TWOREK, J. A. A Genetic Mixture Analysis for use with Incomplete Source Population Data. **Canadian Journal of Fisheries and Aquatic Sciences**, [S. l.], vol. 47, no. 3, p. 620–634, Mar. 1990.

- SRIVASTAVA, S. *et al.* Patterns of microsatellite distribution across eukaryotic genomes. **BMC Genomics**, [S. l.], vol. 20, p. 153, 22 Feb. 2019.
- STACY, E. M. *et al.* Comparing microsatellites and single nucleotide polymorphisms to evaluate genetic structure and diversity in wolverines (*Gulo gulo*) across Alaska and western Canada. **Journal of Mammalogy**, [S. l.], , p. gya151, 15 Jan. 2025.
- STORNIOLI, J. H. F. *et al.* The B Chromosomes of *Prochilodus lineatus* (Teleostei, Characiformes) Are Highly Enriched in Satellite DNAs. **Cells**, [S. l.], vol. 10, no. 6, p. 1527, 17 Jun. 2021.
- SUZUKI, F. M.; PIRES, L. V.; POMPEU, P. S. Passage of fish larvae and eggs through the Funil, Itutinga and Camargos Reservoirs on the upper Rio Grande (Minas Gerais, Brazil). **Neotropical Ichthyology**, [S. l.], vol. 9, p. 617–622, 2011.
- TAGLIACOLLO, V. A. *et al.* Time-calibrated phylogeny of neotropical freshwater fishes. **Frontiers in Bioinformatics**, [S. l.], vol. 4, 6 Dec. 2024. Available at: <https://www.frontiersin.org/journals/bioinformatics/articles/10.3389/fbinf.2024.1433995/full>. Accessed on: 18 Mar. 2025.
- TELLES, M. P. de C. *et al.* A geographical genetics framework for inferring homing reproductive behavior in fishes. **Genetica**, [S. l.], vol. 139, no. 2, p. 243–253, 1 Feb. 2011.
- TEMNYKH, S. *et al.* Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res* 11: 1441-1452. **Genome research**, [S. l.], vol. 11, p. 1441–52, 1 Aug. 2001.
- UNTERGASSER, A. *et al.* Primer3—new capabilities and interfaces. **Nucleic Acids Research**, [S. l.], vol. 40, no. 15, p. e115, Aug. 2012.
- VÄRHÄ, J.-P. *et al.* Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. **Molecular Ecology**, [S. l.], vol. 16, no. 13, p. 2638–2654, Jul. 2007.
- VAISBAND, M. *et al.* Validation of genetic variants from NGS data using deep convolutional neural networks. **BMC Bioinformatics**, [S. l.], vol. 24, no. 1, p. 158, 20 Apr. 2023.

- VALENZUELA-QUIÑONEZ, F. How fisheries management can benefit from genomics? **Briefings in Functional Genomics**, [S. l.], vol. 15, no. 5, p. 352–357, 1 Sep. 2016.
- VAN DE SCHOOT, R. *et al.* Bayesian statistics and modelling. **Nature Reviews Methods Primers**, [S. l.], vol. 1, no. 1, p. 1–26, 14 Jan. 2021.
- VAN OOSTERHOUT, C. *et al.* micro-checker: software for identifying and correcting genotyping errors in microsatellite data. **Molecular Ecology Notes**, [S. l.], vol. 4, no. 3, p. 535–538, 2004.
- VERITY, R.; NICHOLS, R. A. What is genetic differentiation, and how should we measure it— G_{ST} , D , neither or both? **Molecular Ecology**, [S. l.], vol. 23, no. 17, p. 4216–4225, 2014.
- VØLLESTAD, L. A.; PRIMMER, C. R. Understanding local adaptation in a freshwater salmonid fish: evolution of a research programme. **ICES Journal of Marine Science**, [S. l.], vol. 76, no. 6, p. 1404–1414, 1 Dec. 2019.
- WANG, H. *et al.* Impact of different numbers of microsatellite markers on population genetic results using SLAF-seq data for *Rhododendron* species. **Scientific Reports**, [S. l.], vol. 11, no. 1, p. 8597, 21 Apr. 2021.
- WANG, J. A parsimony estimator of the number of populations from a STRUCTURE-like analysis. **Molecular Ecology Resources**, [S. l.], vol. 19, no. 4, p. 970–981, Jul. 2019.
- WANG, J. Fast and accurate population admixture inference from genotype data from a few microsatellites to millions of SNPs. **Heredity**, [S. l.], vol. 129, no. 2, p. 79–92, Aug. 2022.
- WANG, J. Maximum-Likelihood Estimation of Admixture Proportions From Genetic Data. **Genetics**, [S. l.], vol. 164, no. 2, p. 747–765, 1 Jun. 2003.
- WANG, J. PopCluster: A Population Genetics Model-Based Toolset for Simulating, Inferring and Visualising Individual Admixture and Population Structure. **Molecular Ecology Resources**, [S. l.], vol. n/a, no. n/a, p. e14058, 2024.
- WANG, J. The computer program structure for assigning individuals to populations: easy to use but easier to misuse. **Molecular Ecology Resources**, [S. l.], vol. 17, no. 5, p. 981–990, 1 Sep. 2017.
- WANG, P.; WANG, F. A proposed metric set for evaluation of genome assembly quality. **Trends in Genetics**, [S. l.], vol. 39, no. 3, p. 175–186, 1 Mar. 2023.

- WAPLES, R. S. Testing for Hardy–Weinberg Proportions: Have We Lost the Plot? **Journal of Heredity**, [S. l.], vol. 106, no. 1, p. 1–19, 1 Jan. 2015.
- WENNE, R. Microsatellites as Molecular Markers with Applications in Exploitation and Conservation of Aquatic Animal Populations. **Genes**, [S. l.], vol. 14, no. 4, p. 808, 27 Mar. 2023.
- WIKIMEDIA COMMONS. [s. d.]. Available at: <https://commons.wikimedia.org>. Accessed on: 18 Mar. 2025.
- WILKES, M. A. *et al.* Not just a migration problem: Metapopulations, habitat shifts, and gene flow are also important for fishway science and management. **River Research and Applications**, [S. l.], vol. 35, no. 10, p. 1688–1696, Dec. 2019.
- WINGEN, L. U.; BROWN, J. K. M.; SHAW, M. W. The population genetic structure of clonal organisms generated by exponentially bounded and fat-tailed dispersal. **Genetics**, [S. l.], vol. 177, no. 1, p. 435–448, Sep. 2007.
- XU, M. *et al.* Global freshwater fish invasion linked to the presence of closely related species. **Nature Communications**, [S. l.], vol. 15, no. 1, p. 1411, 15 Feb. 2024.
- XU, T.; MOORE, I. D.; GALLANT, J. C. Fractals, fractal dimensions and landscapes — a review. **Geomorphology**, [S. l.], vol. 8, no. 4, p. 245–262, 1 Dec. 1993.
- YAZBECK, G. M.; KALAPOTHAKIS, E. Isolation and characterization of microsatellite DNA in the piracema fish *Prochilodus lineatus* (Characiformes). **Genetics and molecular research: GMR**, [S. l.], vol. 6, no. 4, p. 1026–1034, 5 Oct. 2007.
- YAZBECK, G. M. *et al.* A broad genomic panel of microsatellite *loci* from *Brycon orbignyanus* (Characiformes: Bryconidae) an endangered migratory Neotropical fish. **Scientific Reports**, [S. l.], vol. 8, no. 1, p. 8511, 31 May 2018.
- YEPES-BLANDÓN, J. A. *et al.* Draft genome assembly for the colombian freshwater bocachico fish, *Prochilodus magdalenae*. **Frontiers in Genetics**, [S. l.], vol. 13, p. 989788, 2023.
- ZELLHUBER, A. Environmental Policy in Brazil. Tensions Between Conservation and the Ideology of Growth. In: DE LA FONTAINE, D.; STEHNKEN, T. (eds.). **The Political System of Brazil**. Berlin, Heidelberg: Springer, 2016. p. 329–350. Available at: https://doi.org/10.1007/978-3-642-40023-0_18. Accessed on: 30 May 2024.

ZIMMERMAN, S. J.; ALDRIDGE, C. L.; OYLER-MCCANCE, S. J. An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. **BMC Genomics**, [S. l.], vol. 21, no. 1, p. 382, 1 Jun. 2020.

3Blue1Brown - FRACTALS ARE TYPICALLY NOT SELF-SIMILAR. [S. l.: s. n.], 27 Jan. 2017.
Available at: <https://www.youtube.com/watch?v=gB9n2gHsHN4>. Accessed on: 18 Mar. 2025.

Apêndices

Apêndice A – Arquivos suplementares externos

Arquivo suplementar 1: Sequências de DNA de microssatélites e *primers* apresentados no formato de texto simples (txt) para *Prochilodus lineatus*. A ordem das colunas é “Locus”; “Original fasta label”; “Motif”; “Repeats”; “5'-flank”; “3'-flank”; “F-primer”; “F-Tm”; “R-Primer”; “R-Tm”; “PCR product”; “Product length”; “Location”; “Scaffold or contig label in the newly presented assembly, when mapped”. Este arquivo está disponível no FigShare, doi: 10.6084/m9.figshare.26848654.

<https://doi.org/10.6084/m9.figshare.26848654>.

Arquivo suplementar 2: Lista de sequências de DNA no formato fasta para contigs da montagem original de onde os *loci* microssatélites para *Prochilodus lineatus* foram originalmente caracterizados. Este arquivo está disponível no FigShare, doi: 10.6084/m9.figshare.26848015

<https://doi.org/10.6084/m9.figshare.26848015>.

Arquivo suplementar 3: Amostras populacionais de *Prochilodus lineatus*, identificações de peixes e seus rearranjos de acordo com *clusters* definidos por MLE de 1ª, 2ª e 3ª ordens. Este arquivo está disponível no FigShare, doi: 10.6084/m9.figshare.26871031.

<https://doi.org/10.6084/m9.figshare.26871031>.

Arquivo suplementar 4: Tabelas suplementares externas, incluindo valores de log-verossimilhança da análise de agrupamento para panmixia e estimativas de diversidade genética para *Prochilodus lineatus* para *clusters* de 1ª, 2ª e 3ª ordens de análise hierárquica. Este arquivo está disponível em FigShare, doi: 10.6084/m9.figshare.27020857.

<https://doi.org/10.6084/m9.figshare.27020857>.

Arquivo suplementar 5: Testes exatos de Hardy-Weinberg para *Prochilodus lineatus* em *clusters* definidos por MLE de 1ª, 2ª e 3ª ordens da análise hierárquica. Este arquivo está disponível em FigShare, doi: 10.6084/m9.figshare.27020869.

<https://doi.org/10.6084/m9.figshare.27020869>.

Arquivo suplementar extra : Manuscrito relacionado à esta tese, depositado em repositório *pre-print* (versão de manuscrito submetido ao *Journal of Fish Biology* em 28 de setembro de 2024):

<https://www.biorxiv.org/content/10.1101/2025.03.14.643328>

Apêndice B - Tabelas Suplementares

Tabela suplementar 1: Marcadores microsatélites validados para *Prochilodus lineatus*, com de intervalos de comprimento de fragmentos de DNA amplificados observados para cada *locus* e suas respectivas classes alélicas definidas.

<i>Locus</i>	<i>Primers</i>	<i>Tema</i>	<i>T_m</i> (°C)	Intervalo de fragmentos (bp)	<i>N_A</i>	<i>N</i>
Prol01	F: CAAAAGAAAAACGAAAATAAAAGAAACA R: CCCTCTGGGATCAATATAGTATGTATGT	AGAT	54.0	109-261	24	117
Prol02	F: TTAAACTCATTCAACCGGCCTAC R: GTTGGAGGTTACTTGATAGGGAGGTA	TATC	54.0	113-389	24	119
Prol03	F: TTCCCATCTTTAGAAATGTCCTTC R: CTCCTGGGGATAAACCATCAATCT	AGAT	63.6	134-375	27	122
Prol04	F: TCTGAAACTATTTTTATAGGTAGGGGGA R: TCCTCTCACCTACAACTTCCTCAA	TAA	56.4	113-264	25	125
Prol05	F: TCTATCCATCCAATCTATCCATCCA R: AAATTCCATAGTTTGGAGTGACGC	ATCT	58.8	89-276	31	114
Prol06	F: GGCCACTATAAAAGGTAGGCACAA R: CATACGCTGCTCTAAATGTTTCAGC	AGAT	54.0	121-227	17	104
Prol07	F: TGGAGGCTGCAGGTATAAGTTTTT R: AAGGCTGGTCCCTAACCTCC	TTA	56.4	102-205	20	94
Prol08	F: AAGGGCAAGATCAGACAAACAAG R: GAATACTACGCTCATTTCAGAAGAAAAA	TAT	63.6	146-288	16	122
Prol10	F: CACAGACAGACACATATGGGGA R: CAGGTATGTTCTATTGAAATGCTAACAC	AT	61.2	145-316	25	100
Prol11	F: AGAGCTTCAATAGGCTGAAGTCGT R: TGAATTTCCCTCTGGGATCAATAA	TAGA	56.4	113-222	23	110
Prol12	F: GTTGCACCTTTGTAGAGGGGTATG R: ACAGCCGTGGATATAGTCAGAAGC	TCA	56.4	117-280	23	107
Prol19	F: TAAAATGTGGTTTCCTGTGCAATG R: TTTGCATTTTATCCATCTTTTCCC	TAGA	60.0	87-200	18	106
Prol53	F: ATTGCTGTGTCACTTAAGCATGT R: TGCTACTGGCTGCTAAATTTCTT	AGAT	51.0	127-209	13	137
Prol54	F: CGAGACGTCTAGCTGACACCATAA R: TTGACTTCTTGCTGTCATGTCAAA	AGAT	49.0	96-309	28	110
Prol57	F: AATGAAACACAGTCAGGCTCCTTC R: TTCCTTCAGGCAATGCACTTTTAT	ATA	49.0	139-317	27	122

T_m=Temperatura de anelamento; *N*=Tamanho amostral (genotipagens realizadas); *F*=*Primer* direto; e *R*=*Primer* reverso; *N_A*=Riqueza alélica (número de alelos).

Tabela suplementar 2: Diversidade genética de microssatélites em amostras populacionais de *Prochilodus lineatus* do alto rio Grande, MG. Testes exatos de Hardy-Weinberg tiveram significância definida como $\alpha=0,008$. Todos os valores de probabilidade (p) são significativos, exceto onde indicado (^{ns}), e os valores considerados zero são, na verdade, valores extremamente baixos diferentes de zero $<0,00009$.

Amostras	Locus	N	N _A	H _O	H _E	F	HWE (p)
Camargos	Prol01	17	14	0.65	0.91	0.29	0.0023
	Prol02	15	14	0.60	0.89	0.33	0
	Prol03	13	13	0.31	0.91	0.66	0
	Prol04	17	10	0.41	0.87	0.53	0
	Prol05	14	13	0.36	0.9	0.6	0
	Prol06	14	9	0.43	0.81	0.47	0
	Prol07	15	10	0.73	0.85	0.14	0.0191 ^{ns}
	Prol08	13	4	0.08	0.65	0.88	0
	Prol10	16	10	0.69	0.85	0.19	0.0049
	Prol11	17	12	0.35	0.88	0.6	0
	Prol12	13	6	0	0.77	1	0
	Prol19	12	9	0.50	0.84	0.41	0.0004
	Prol53	18	9	0.33	0.77	0.57	0
	Prol54	17	15	0.41	0.91	0.55	0
	Prol57	17	11	0.41	0.87	0.53	0
	Loci	15.2	10.6	0.42	0.85	0.52	0
Itutinga 1	Prol01	33	16	0.27	0.91	0.7	0
	Prol02	33	18	0.55	0.92	0.41	0
	Prol03	31	16	0.52	0.91	0.43	0
	Prol04	36	13	0.31	0.9	0.66	0
	Prol05	26	22	0.62	0.94	0.34	0
	Prol06	34	11	0.21	0.87	0.76	0
	Prol07	20	12	0.95	0.88	-0.08	0.0157 ^{ns}
	Prol08	38	10	0.45	0.85	0.48	0
	Prol10	33	20	0.73	0.94	0.22	0
	Prol11	30	16	0.57	0.89	0.37	0
	Prol12	31	13	0.16	0.9	0.82	0
	Prol19	30	13	0.23	0.91	0.74	0
	Prol53	39	9	0.31	0.64	0.52	0
	Prol54	34	20	0.38	0.92	0.59	0
	Prol57	38	16	0.47	0.91	0.48	0
	Loci	32.4	15	0.45	0.89	0.50	0
Itutinga 2	Prol01	34	19	0.29	0.92	0.68	0
	Prol02	36	18	0.56	0.91	0.39	0
	Prol03	39	19	0.36	0.93	0.61	0
	Prol04	33	14	0.27	0.91	0.7	0

Funil	Prol05	38	24	0.39	0.94	0.58	0
	Prol06	27	14	0.11	0.89	0.87	0
	Prol07	27	18	0.81	0.91	0.1	0.0498 ^{ns}
	Prol08	39	14	0.62	0.87	0.3	0
	Prol10	27	15	1	0.91	-0.1	0
	Prol11	33	17	0.27	0.88	0.69	0
	Prol12	32	17	0.34	0.91	0.62	0
	Prol19	31	13	0.13	0.89	0.86	0
	Prol53	40	10	0.18	0.65	0.73	0
	Prol54	34	19	0.29	0.92	0.68	0
	Prol57	36	22	0.81	0.92	0.13	0.0210 ^{ns}
	Loci	33.7	16.9	0.43	0.89	0.52	0
	Prol01	33	16	0.15	0.92	0.84	0
	Prol02	35	17	0.49	0.91	0.46	0
	Prol03	39	21	0.51	0.95	0.46	0
	Prol04	39	17	0.51	0.92	0.44	0
	Prol05	36	27	0.56	0.95	0.42	0
	Prol06	29	12	0.21	0.79	0.74	0
	Prol07	32	11	0.91	0.85	-0.07	0.4959 ^{ns}
	Prol08	32	13	0.66	0.86	0.24	0
	Prol10	24	18	1	0.93	-0.08	0.7323 ^{ns}
	Prol11	30	18	0.4	0.92	0.57	0
	Prol12	31	19	0.29	0.9	0.68	0
	Prol19	33	15	0.15	0.91	0.83	0
	Prol53	40	11	0.23	0.68	0.67	0
	Prol54	25	14	0.72	0.91	0.2	0.0052
	Prol57	31	21	0.71	0.93	0.23	0
	Loci	32.6	16.67	0.50	0.89	0.44	0
	Total	140	341	0.45	0.91	0.50	0

N=Tamanho; N_A=Riqueza alélica (número de alelos); H_O=Heterozigosidade observada; H_E=Heterozigosidade Esperada; F=Coeficiente de endogamia; EHW=Teste exato para equilíbrio de Hardy-Weinberg.

Apêndice C – Figuras suplementares

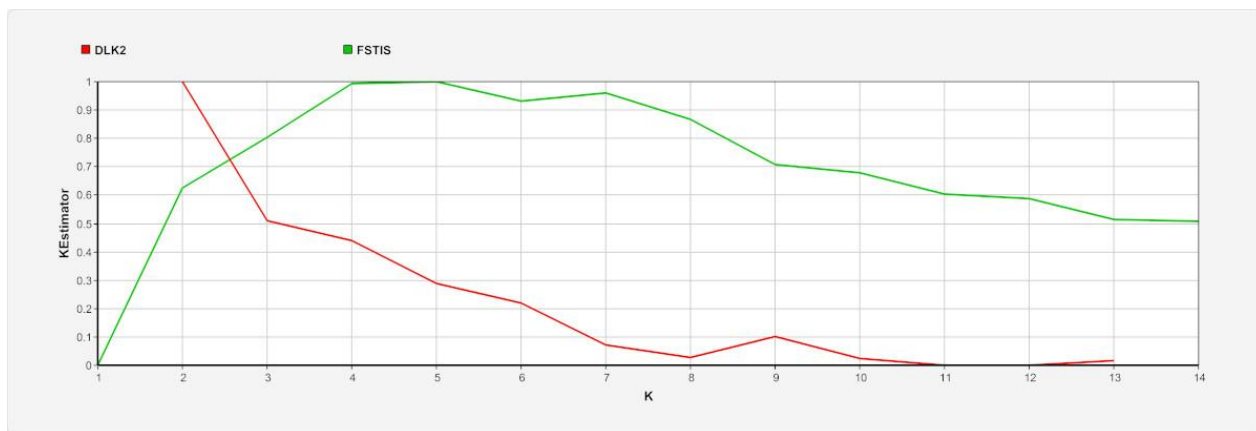


Figura suplementar 1: Estimativa ótima de K por dois métodos, D_{LK2} e F_{STIS} para MLE da admistura de 1st ordem em *Prochilodus lineatus* do alto rio Grande, MG.

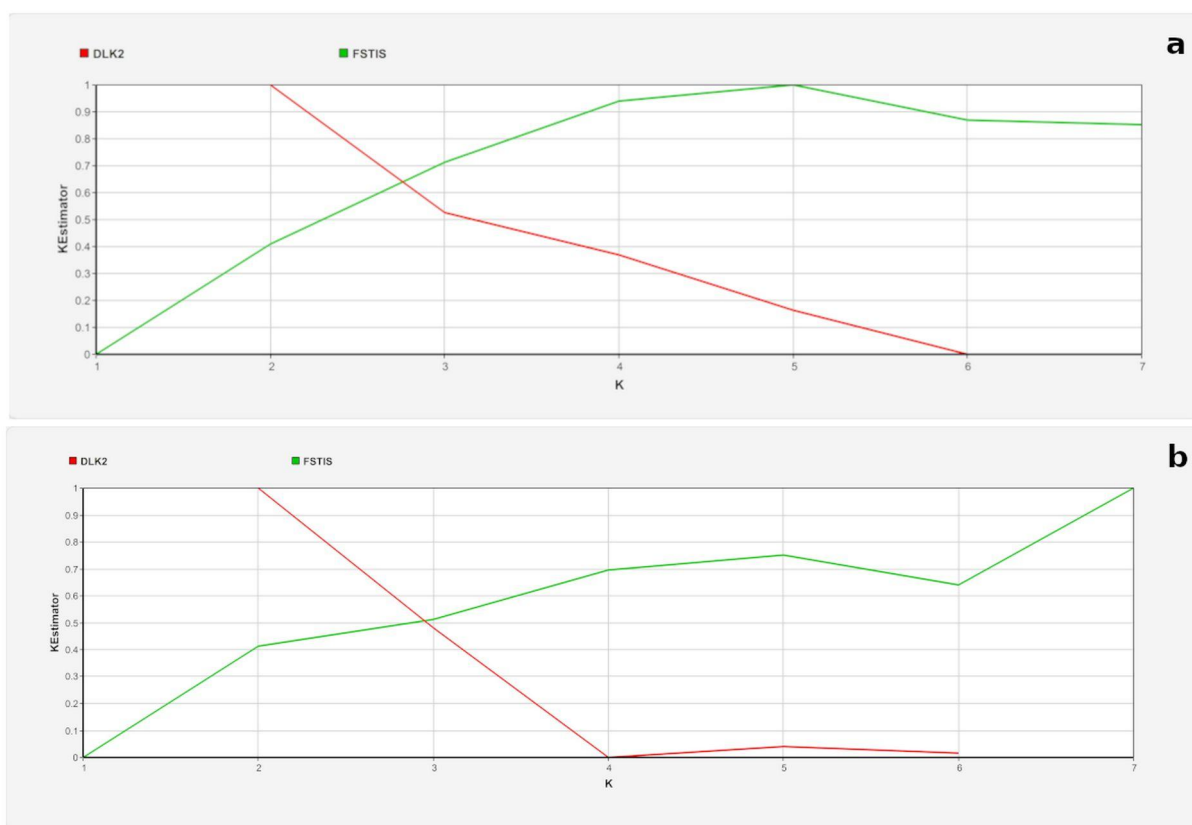


Figura suplementar 2: Estimativa ótima de K por dois métodos, D_{LK2} e F_{STIS} para MLE da admistura de 2ª ordem em *Prochilodus lineatus* do alto rio Grande, MG. **a)** Cluster 1 da 1ª ordem. **b)** Cluster 2 da 1ª ordem.

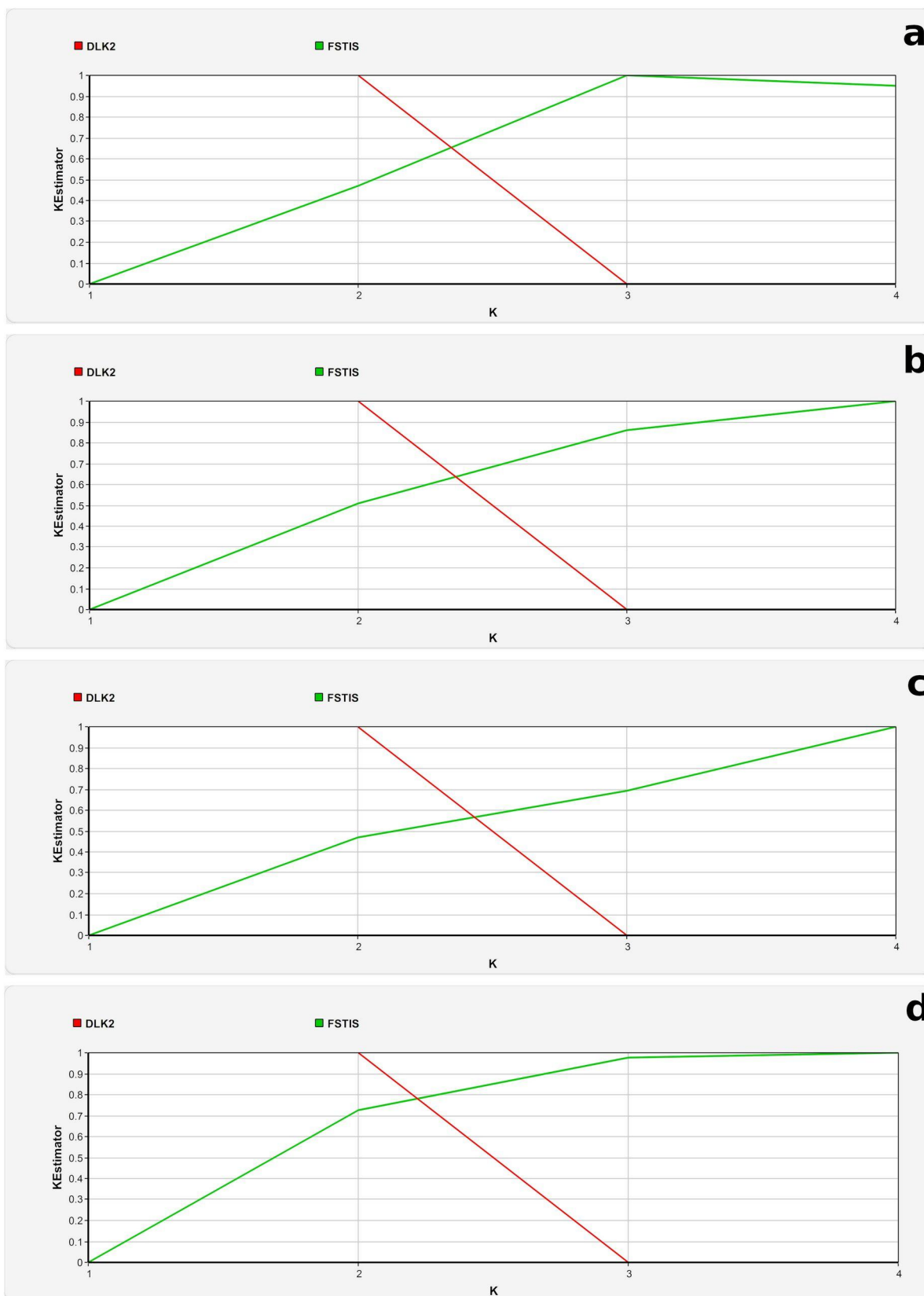


Figura suplementar 3: Estimativa ótima de K por dois métodos, D_{LK2} e F_{STIS} para MLE da admistura de 3^a ordem em *Prochilodus lineatus* do alto rio Grande, MG. **a)** Cluster 1 da 2^a ordem. **b)** Cluster 2 da 2^a ordem. **c)** Cluster 3 from the 2^a ordem. **d)** Cluster 4 da 2^a ordem.

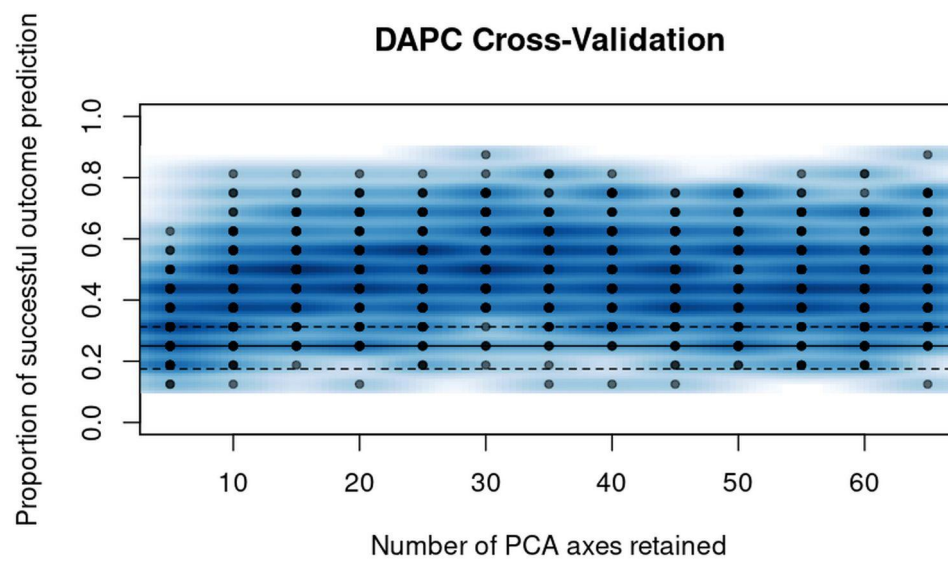


Figura suplementar 4: Validação cruzada por *machine learning* da DAPC com 100 réplicas.

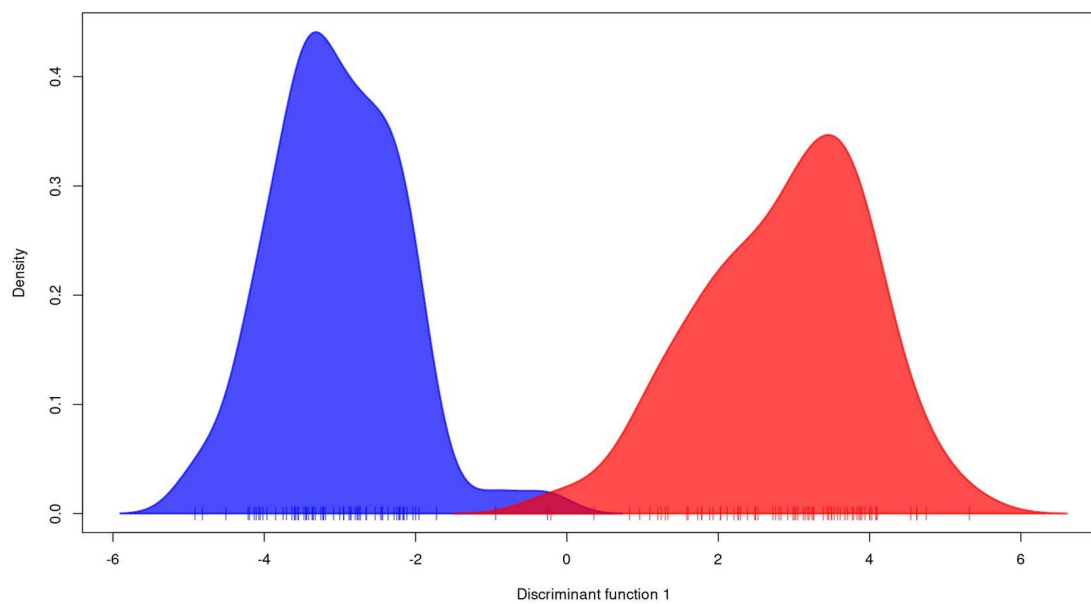


Figura suplementar 5: Reprise da DAPC *de novo*, agora com número ótimo de PC da validação cruzada.