

A fractal pattern of hierarchical genetic population structure in mixed stocks across fish segregated by dams revealed by genomic resources for curimba *Prochilodus lineatus*

Gabriel M. Yazbeck^{1,2,3}  | Gésica Aparecida Santana Nascimento¹  |
Leiliane Campos Carvalho²  | Raíssa Cristina Dias Graciano²  |
Rosiane de Paula Santos³  | Rafael Sachetto Oliveira⁴ 

¹Programa de Pós Graduação em Ecologia, Universidade Federal de São João Del Rei, São João, Brazil

²Programa de Pós Graduação em Biotecnologia, Universidade Federal de São João Del Rei, Divinópolis, Brazil

³Laboratório de Recursos Genéticos (LARGE), DEZOO, Universidade Federal de São João Del Rei, São João, Brazil

⁴Departamento de Ciência da Computação, Universidade Federal de São João Del Rei, São João, Brazil

Correspondence

Gabriel M. Yazbeck, Laboratório de Recursos Genéticos (LARGE), DEZOO, Universidade Federal de São João Del Rei, Praça Frei Orlando, 170, CEP 36.307-352, São João Del Rei, MG, Brazil.
Email: dna@ufs.br

Funding information

Companhia Energética de Minas Gerais, Grant/Award Number: GT345; Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Number: 303023/2023-6; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: Code 001

Abstract

Genomic resources, new microsatellite markers and a novel observation of a fractal pattern in the genetic population structure are presented for curimba, *Prochilodus lineatus*, a freshwater migratory model species of South America. Our main goals were to investigate the presence of mixed fish stocks and the effects of damming-induced fragmentation, as well as hatchery releases, over the genetic diversity of this species population. We focused on the upper segment of the highly fragmented Grande River, Minas Gerais, Brazil, within the megadiverse La Plata River basin. Using maximum likelihood estimation, different genetically admixed co-occurring fish stocks were identified. These served as input for a second analysis, the results of which were further examined as a hierarchical genetic structure investigation. Each of these three iterative steps consistently revealed two newly rearranged internal groups within the previously defined clusters (2³). We suggest that this nested pattern of mixed stocks in the main river channel reflects distinct, well-defined breeding sites that are differentially distributed across the basin's spatial scale levels (main river, first-order tributaries, second-order tributaries and so on). These potential sites would be geographically arranged based on a fractal disposition, mirroring the self-similar organization of the bottom-up stream network. This would imply homing behaviour in *P. lineatus*. Our work also provides high-throughput molecular data, including the first scaffold-level draft genome assembly for this species, alongside a valuable set of microsatellites for rapid, low-cost DNA marker development. The findings will aid researchers and environmental managers in addressing conservation challenges of Neotropical migratory fishes. The results support the hypothesis of isolating effects caused by dams and suggest restocking influences local genetic variability.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Journal of Fish Biology* published by John Wiley & Sons Ltd on behalf of Fisheries Society of the British Isles.

Importantly, the study highlights the role of tributaries of different orders in harbouring genetically diverse populations of migratory freshwater fishes. We also introduce the concept of *quasi*-demes, defined as nearly random-mating population units measurable at arbitrary scales across the basin's landscape, which become increasingly similar to true demes when smaller tributaries are considered.

KEYWORDS

genomics, microsatellites, migratory fish, Neotropics, population genetics

1 | INTRODUCTION

The curimba *Prochilodus lineatus* Valenciennes 1836 (Characiformes: Prochilodontidae), also known as sabálo, grumatá, streaked prochilod and other common names, is one of the most important freshwater migratory species for inland fisheries in South America. It corresponds to >50% of the fish biomass from the La Plata basin (Bowen, 2022a, 2022b), a megadiverse hydrological system (Cassemiro et al., 2023), reaching Argentina, Bolivia, Brazil, Paraguay and Uruguay. This major Neotropical basin is heavily impacted by anthropogenic activity (Reis et al., 2016). *P. lineatus* also naturally occurs in the Paraíba do Sul basin (Brazil) and has been recorded as an invasive species in China (Endruweit, 2014) and Vietnam (Kalous et al., 2012).

P. lineatus has a conspicuous specialized inverted funnel mouth, an adaptation from its detritivorous habits (Bowen, 2022b), and, due to its abundance and popularity, it is an emblematic example of the Neotropical potamodromous (i.e., freshwater migratory) life history. Reproductive migrations of these fishes are triggered by environmental cues, such as increasing daylight, the onset of the tropical rains and the lunar phase (e.g., Lopes et al., 2018). These upstream runs are regionally denominated *piracema*, a term used to refer both to the migratory process and to the fish species which partake in it. It typically occurs from around the early austral summer through to the end of the season (Carolsfeld et al., 2003). Thus, this process is potentially sensible to ongoing climate change (Alix et al., 2020).

Piracema fishes naturally spawn upon reaching upstream breeding grounds (e.g., Ribolli et al., 2020). Larvae and juveniles then passively drift downstream and reach nursery and feeding floodplains, where they grow and mature (Petry et al., 2003). Adults spawn several consecutive years during their life cycle (Carolsfeld et al., 2003). Although *piracema* fishes include species from other taxonomic orders, they are composed primarily of Characiformes and Siluriformes.

Species from the *Prochilodus* genus were among the first Neotropical migratory fishes to have its ex situ spawning mastered. This was motivated by the intensive exploitation of hydroelectricity generation and its impact over migratory species, aiming restocking operations, intended to counteract dwindling fisheries around regions affected by dams (Saraiva & Pompeu, 2016).

These restocking practices have been largely abandoned in some relevant places for hydroelectric production, such as the Brazilian state of Minas Gerais (MG), in the past decade. This was, in part, due to general relaxation of environmental protection-oriented regulations

(Zellhuber, 2016), although a rigorous, holistic (i.e., ecological, genetic and anthropological) assessment of the activity's potential benefits or pitfalls is still lacking. Given its status as the main targeted fish for stock enhancement in the past 50 years or so, *P. lineatus* is a natural candidate as a model species for gauging the impacts of dams and the effectiveness of possible mitigation measures regarding *piracema* fishes.

Fishes under the impacts of hydroelectric power production may benefit from objective genetic information raised through molecular markers such as microsatellites, which are amenable to simple genotyping routines for hatchery or wild fish monitoring operations (Wenne, 2023). The popularization of next-generation sequencing (NGS) techniques has made the characterization of new DNA markers increasingly accessible and straightforward (Yazbeck et al., 2024). Genomic resources currently available for *Prochilodus* species include full mitochondrial DNA sequences (Carmo et al., 2016; Chagas et al., 2016; Santos et al., 2021), NGS data for *P. lineatus* (Stornioli et al., 2021) and a reference genome for *Prochilodus magdalenae* (Yepes-Blandón et al., 2022).

Mixed fish stocks can follow a genetically *mixed* population model, characterized by each individual's genome originating from a single ancestral population source (genetic cluster), out of a *K* number of alternatives, where these demes do not share breeding grounds and events. Alternatively, mixed fish stocks might be portrayed by an *admixture* model. It considers different individual proportions of genomic heritage out of several ancestral genetic clusters where previous generations have potentially engaged in common reproductive events (Pritchard et al., 2000; Wang, 2022). Admixture, thus, provides a convenient framework for the problem of unravelling the genetic structure of *piracema* fishes under the impact of dams and decades of stocking operations largely conducted without planned considerations towards genetic issues.

Prior to the use of microsatellite and admixture methods, Sivasundar et al. (2001) detected no genetic population structure in *P. lineatus*, attributing this to its strong migratory ability. Rueda et al. (2013) were the first to apply both methods, revealing seasonal stock segregation in the Lower Uruguay River and suggesting mixed stocks for fisheries management. Later, Perini et al. (2021) described complex dynamics between the Pardo and Lower Grande rivers. Ferreira et al. (2017) investigated *P. lineatus* stocks from the Middle and Upper Paraná basin, whereas Rosa et al. (2022) studied those from the Mogi-Guaçu River, detecting temporal genetic structuring. Ferreira et al. (2023) also reported this along with mixed stocks in the

Parapanema River basin. These investigations point to the importance of free-flowing tributaries for spawning in this species, reporting high genetic variability and generally resolving two or three genetic admixture clusters. No genetic structure analysis exists for *P. lineatus* from the Upper Grande River, the uppermost region of the La Plata River basin's eastern flank.

The Grande River springs from an altitude of almost 2,000 m in Minas Gerais and runs through 1360 km until its confluence with the Paranaíba River, where both coalesce onto the Paraná River (a major component of the La Plata basin). It is an illustrative example of thoroughly impounded river, with 12 large dams, with the Camargos hydroelectric power plant being the last dam upbound implemented around 1960. It is closely followed downstream by another dam from the Itutinga power plant, operational since 1955, which delimits a singular, relatively small closed impoundment (without tributaries).

Immediately downstream the Camargos dam, a conservation-oriented hatchery was active for decades until 2019 in broodstock management and fish stocking operations of a largely unaccounted number of fry, fingerlings and adult fishes (possibly dozens of millions), mainly of *P. lineatus*. These stocking operations primarily occurred throughout tributaries from the Upper Grande River basin (particularly around municipalities upstream the Camargos dam). More relevant to this study, 9000 fry and fingerlings from this species were released in situ in 2016 at the enclosure between the Camargos and Itutinga dams, besides the likely occurrence of unintentional escapes from adult fish, often kept at net-pens in the reservoir. It provides a singularly convenient opportunity for evaluating the local genetic effects of restocking.

Further downstream lies the relatively new Funil power plant dam (constructed around the year 2000), the only dam accommodating a fish transposition mechanism, a lift (Suzuki et al., 2011). The degree in which these large dams genetically isolate *P. lineatus* in the Upper Grande River is not currently known nor are the genetic effects from fish stocking there.

Our core objectives in this work were (i) testing the hypothesis of the occurrence of mixed stocks across *P. lineatus* populations segregated by the past three upbound dams from the Upper Grande River, MG, Brazil; (ii) testing the hypothesis that dams are genetically isolating migratory fish in the region; and (iii) testing the hypothesis that fish stocking practices have altered genetic diversity in populations of fish trapped between dams. We aimed at doing so through the development of a genome-wide set of microsatellite *loci* for rapid marker development in *P. lineatus* while making available new genomic resources for this fish, a model for Neotropical freshwater migratory species in general.

2 | MATERIALS AND METHODS

2.1 | Ethics statement

All procedures presented in this study were conducted under the guidance and authorization of the Committee on Ethics in Animal Use

(CEUA-UFSJ) under licences 27/2011 and 6991130223, as well as CGEN licences A9D0E5 and AC5A784, and SISBIO licences 37222-2 and 85497-1. No experimental assays were performed on live specimens.

2.2 | Genomics, bioinformatics and microsatellite marker development

Genomic libraries of whole-genome shotgun fragments (500 bp) were produced for NGS out of a single adult female individual from the Itutinga Hatchery Station (21°19'20.7" S 44°36'56.9" W, Figure 1). The methodological steps and parameters for NGS and initial microsatellite characterization were the same as presented in Yazbeck et al. (2018) and carried out by an external service provider (SP) BGI (Tai Po, Hong Kong). The resulting 90 bases paired-end (PE) short-reads were organized as three sets of twin FASTQ files. We verified the quality of DNA short-reads using MultiQC (Ewels et al., 2016) and computed all of its 21-mers (k-mer substrings) in jellyfish (Marçais & Kingsford, 2011). These results were used to estimate the expected genome size of *P. lineatus* through a frequency histogram in R (version 4.4.1, R Core Team, 2024).

We used all the DNA short-reads, along a set of fasta format contigs containing microsatellites resolved by the SP upon the original undisclosed assembly, to execute a de novo genomic assembly. For this step, we used MEGAHIT (Li et al., 2015), using k-mer = 47, to generate new contigs and, then, submitting these to the SOAPdenovo-fusion module from SOAPdenovo 2 (Luo et al., 2012) to resolve scaffolds. This genome assembly was processed in the Sagarana high-performance computing cluster (CEPAD-ICB-UFMG, Belo Horizonte, Brazil). The National Center for Biotechnology Information's (NCBI – Sayers et al., 2022) contamination screening tool FCS (Astashyn et al., 2024) was used for filtering out mitochondrial assembly, sequences from microorganisms and library adaptors found in the final scaffold-level sequence. Assembly statistics were accessed using assemblyStats (<http://gif.biotech.iastate.edu>) and visualized with assembly-stats 17.02 (Challis, 2017).

We mapped the microsatellites characterized in the steps above in the new genomic assembly following the strategy delineated in Yazbeck et al. (2018). A sequence alignment map (SAM) was then produced using BBmap (Bushnell, 2014) by mapping short-reads unto the genomic assembly and finally converted to a binary format (BAM) using SAMtools (Li et al., 2009).

For this work, we developed new species-specific microsatellite markers to be applied in the population genetics analysis. We used the results of the genomic microsatellite *loci* search by the SP for choosing candidate *loci* for empirical tests. The microsatellites were sorted by decreasing motif size and number of repeats to maximize chances of finding polymorphic *loci* (Brandström & Ellegren, 2008). A total of 22 candidate primer pairs were chosen (12 from the top-sorted candidates and 10 at random) and then synthesized by an SP (Integration DNA Technologies, Coralville, USA). These were applied to wet bench assays in four individual DNA samples. Polymerase

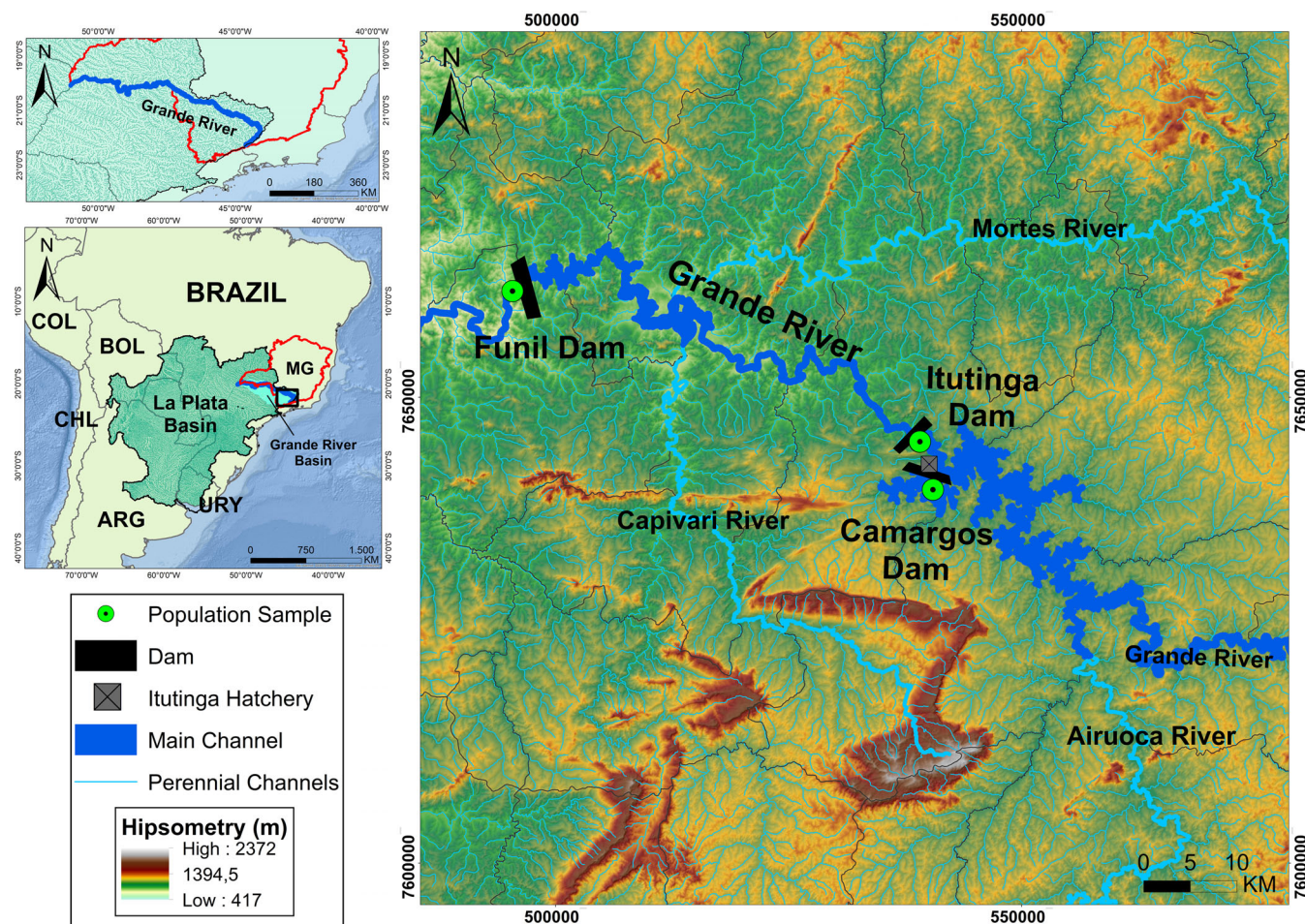


FIGURE 1 Geographical location of the studied population samples for curimba *Prochilodus lineatus* from the Upper Grande River basin, MG, Brazil.

chain reaction (PCR) was carried out in a final volume of 10 μ L, with initial conditions of 1 μ L of unquantified Chelex-extracted genomic DNA to seed the reaction, 10 pmol of each primer, $MgCl_2$ (2 mM), Tris-HCl (10 mM), KCl (50 mM), 0.2 mM of dNTP and 1 unit of *Taq* DNA polymerase enzyme (Phoneutria, Belo Horizonte, Brazil). The thermal PCR cycles were as follows: initial genomic DNA denaturation at 94°C for 4 min, followed by 30 iterated cycles of denaturation at 94°C for 30 s, primers annealing temperature (specific for each primer pair) for 30 s and amplicon elongation step at 72°C for 30 s, with a final elongation step at the same temperature for 5 min.

The resulting PCR products were visualized using 10% polyacrylamide gel electrophoresis at ~ 15 V/cm, with a 25-bp DNA fragment size ladder stained with ethidium bromide. Loci yielding polymorphic patterns were then assayed in a sample of 20 individuals to optimize PCR conditions. Gels were photographed under UVB light and scored using GelAnalyzer. Scored DNA fragments were then converted to nominal allele classes using an automated binning approach under supervision based on frequency histograms and formatted for downstream analysis. Band scoring was cross-validated by at least two authors.

2.3 | Genetic diversity

A total of 100 *P. lineatus* individual tissue samples (ethanol-preserved fin fragments) were used from the DNA bank of the *Laboratório de Recursos Genéticos* (LARGE-UFSJ). These samples were collected in 2013 from three locations (Figure 1) in the Upper the Grande River, MG: downstream the Funil dam, upstream the Itutinga dam and upstream the Camargos dam (see Suzuki et al., 2011 for a more detailed characterization from this general encompassed area). An additional sample of 40 individuals was obtained from the Itutinga dam reservoir in 2023 (Table 1).

After DNA marker development experiments, a total of 15 new empirically validated microsatellite loci were applied to 140 individuals of *P. lineatus*. Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were tested for each population sample using the methods implemented in GenePop 007 (Rousset, 2008), with default options for Markov-chain Monte-Carlo (MCMC) parameters. These were analysed under new significance threshold values for multiple tests following sequential Bonferroni corrections. Microchecker (van Oosterhout et al., 2004) was used for controlling scoring errors

TABLE 1 Evaluated *Prochilodus lineatus* individuals sampled from the last three dams from the Upper Grande River, MG, Brazil.

Sample	ID	Location	Year	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

Abbreviations: ID, individual label; N, sample size.

(stuttering, large allele dropout and null alleles). Genetic data handling, conversion and visualizations were mainly conducted with the *adegenet* package for R (Jombart, 2008). Polymorphic information content (PIC) was calculated according to Botstein et al. (1980).

Genetic diversity estimates for the population samples were obtained using GenAlex (Peakall & Smouse, 2012). We calculated expected and observed heterozygosities (H_E and H_O , respectively) and allelic richness, here expressed as the number of alleles (N_A), to assess diversity within population samples. Jackknife estimates of N_A were conducted using EstimateS (Colwell & Elsensohn, 2014), and confidence intervals (CIs) were calculated in R. Rarefaction analysis of N_A was performed to control for sample size bias by retaining 20 samples, using the *dplyr*, *tidyr* and *pegas* packages in R (Paradis, 2010). Visualizations were plotted using *ggplot2*.

2.4 | Population structure

Jost's differentiation statistic D , which focuses on allelic diversity among samples, rather than heterozygosity, was computed for its suitability for markers with multiple alleles. We also applied Hedrick's unbiased and standardized fixation index G''_{ST} . It is devised for circumventing small and differential sample sizes of genotypes from distinct demes or locations (Meirmans & Hedrick, 2011) and normalized by the maximum theoretical differentiation possible under observed homozygosity levels (Hedrick, 2005). Therefore, it is equally adequate and insightful for hypervariable DNA markers.

A two-level analysis of molecular variance (AMOVA) provided estimates of the variance component structure (ϕ_{ST}), within and among population samples, from pair-wise Euclidean distances among genotypes. All estimated statistics had its significance levels accessed through 999 random permutations. For D and G''_{ST} , standard errors (SE) were calculated by jackknifing over *loci*, and 95% CI were obtained by bootstrapping over *loci* (Peakall & Smouse, 2012).

The genotypic data were also processed using analytical tools implemented in PopCluster (Wang, 2022). A preliminary maximum likelihood estimation (MLE) analysis was performed upon a null-model for admixture (i.e., a strict mixture model, where an individual composite genotype is assigned exclusively to a single ancestral deme) for clustering individuals into K demes. It employs an annealing algorithm, which avoids estimates from being confined within a suboptimal peak,

enhancing its chances of finding the global maximum value from the data. These results were, then, used as a departure point for performing MLE of individual admixture coefficients. This allows for the refinement of the assignment of individuals to one of the K possible population clusters without supervision (i.e., information from the sample's origin). Unequal allele frequencies were assumed, and the weak scaling factor was used. The optimal K value explaining sample substructure ($K = 1$ –14, $N = 100$ replicate runs) was primarily based on the second-order rate of log-likelihood change, with the D_{LK2} estimator, together with computations of the F_{STIS} estimator (Wang, 2022). This initial procedure was labelled as a first-order analysis of population genetic structure.

Then, to explore possible hierarchically hidden strata of genetic structure (Janes et al., 2017; Vähä et al., 2007), the best replicate run results for the chosen K value were used for rearranging samples into new clusters. Each of these was, in turn, independently subjected to a second-order admixture analysis, ranging from $K = 1$ to $K = 7$. This was reiterated once more as a third-order analysis of population structure, assaying $K = 1$ –4. The resulting rearranged clusters across hierarchical orders were examined for genetic diversity and structure measurements as the original population samples.

We also performed a parallel Bayesian inference using the STRUCTURE software suite (Pritchard et al., 2000) under the admixture model, conducted de novo (i.e., without population sample information). Separate Dirichlet hyperparameters (α) were estimated for each inferred cluster, allowing flexible prior distributions. Additionally, the analysis assumed correlated allele frequencies across clusters. A total of 5×10^5 steps were discarded before computing 1×10^6 MCMC iterations, with K varying from 1 to 8 and 20 replicates each. The best K value for the STRUCTURE analysis was primarily estimated using the methods implemented in Kfinder (Wang, 2019).

3 | RESULTS

3.1 | Genomics, bioinformatics and microsatellite marker development

Approximately 483.7 million short-reads (43.53 gigabases) were generated for *P. lineatus* and deposited at the Sequence Read Archive (SRA) database from NCBI (Bioproject accession code PRJNA1120875). Its CG content was 42%, and per base quality value

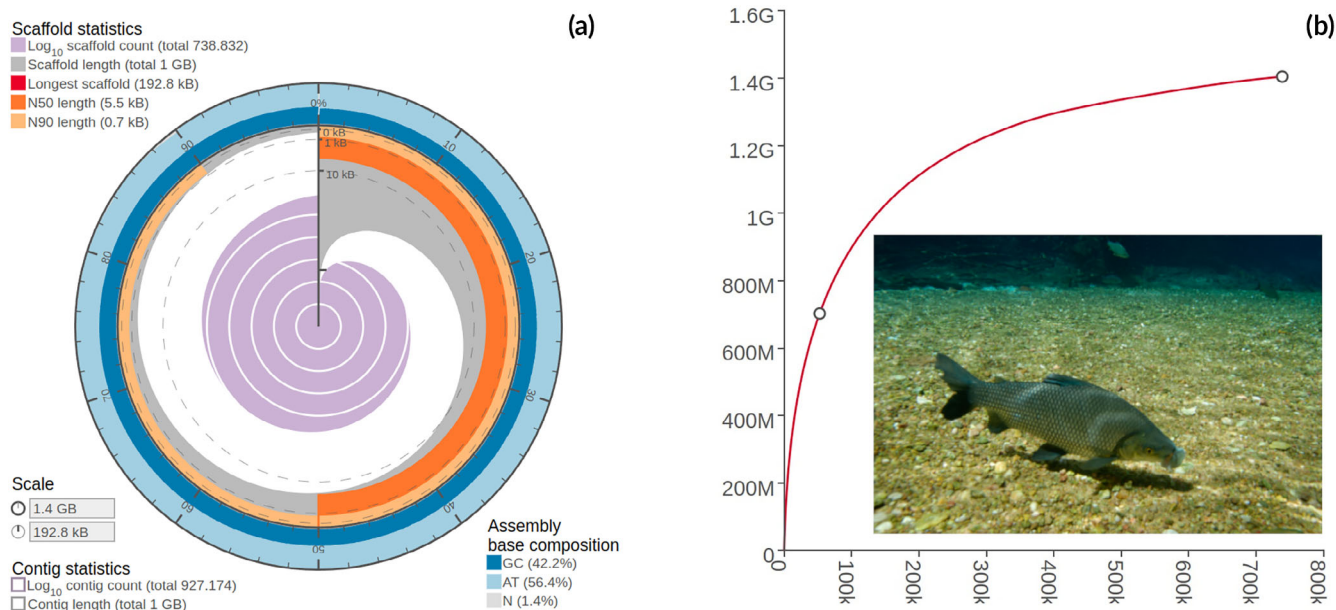


FIGURE 2 Genome assembly statistical features of *Prochilodus lineatus*. (a) The radius inside the circular plot marks the length of the longest scaffold resolved, and the grey segments represent the cumulative percentage of the assembly contained in scaffolds equal or longer than a particular length. (b) Cumulative contig/scaffold length of the genomic assembly as a function of the number of sequences resolved. The point in the middle of the curve marks the L₅₀ feature, the smallest number of sequences, which the sum surpasses half the total genome assembly length (point at the end of the curve – photography by André Seale). G = gigabases and k = thousands of sequences.

(Q) ranged from 31 to 39, with an average of 38 per read. Based on the k-mer counts, total genome size was estimated at 1.65 Gb.

The de novo genome assembly generated was 1.4 Gb, with ~31X coverage. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JBGFTF000000000. The version described in this paper is JBGFTF010000000. The total number of contigs/scaffolds was 738,832 sequences, with N₅₀ = 5503 bp and L₅₀ = 52,610. The average length of contigs was ~1901, with the longest being 192,837 bp (Figure 2). The BAM file associated with it can also be retrieved from NCBI (PRJNA1120875).

A total of 37,476 perfect di- to hexa-nucleotide microsatellite loci for *P. lineatus* were characterized in silico and compiled as supporting information (Supplementary Files 1 and 2). These were mapped to the genome assembly, but around 12% of putative loci could not be traced back to it. This set comprises approximately 50% di-, 13% tri-, 28% tetra-, 6% penta- and 3% hexa-nucleotides.

Of the 22 primer pairs tested, five (Prol14, Prol15, Prol18, Prol31 and Prol52) resulted in complex amplification patterns and two (Prol56 and Prol28) failed to amplify. Fifteen new microsatellite markers were successfully amplified, and the results for the pooled sample of 140 fish are presented in Table 2. Three markers required adjustments from the originally tested PCR conditions: Prol02 and Prol19 used ~6.7 pmol of each primer, with the latter using 35 thermal cycles, and Prol54 also used 35 cycles. The average number of alleles per locus was 22.7 (±1.3 SE). The PIC values were above 0.9 for all loci, except Prol53 (PIC = 0.683), considering the whole sample.

The amplification/genotyping failure rate varied from 2.1% (Prol53) to 32.8% (Prol07), with an average of 18.6% (±7.9 SE), including instances where one allele was scored and the other deemed undetermined for the same diploid genotype. Six DNA samples (four from Camargos, one from Itutinga 1 and one from Funil) failed to amplify between 5 and 10 loci. These were removed from the dataset, and the most downstream analyses were performed with and without them in parallel. Marginal or no differences were observed when comparing results and, thus, these samples were retained.

3.2 | Genetic diversity

Adjusted significance for 60 multiple HWE tests was set to $p < 0.008$ following a sequential Bonferroni correction. By this criterion, only six loci were deemed in HWE. The only locus found to conform to expected HWE proportions within each and every population sample, Prol07, was nevertheless not in equilibrium across all four pooled population samples ($p = 0.002$). All deviations were towards heterozygote deficits. Tests for LD within population samples had an adjusted significance level $\alpha = 0.0001$. No LD was inferred, except for six instances: Prol06 and Prol54 for Camargos; Prol01 and Prol19, Prol05 and Prol57 for Itutinga 1; Prol01 and Prol06, Prol04 and Prol07 for Itutinga 2; and Prol53 and Prol57 for Funil. The following loci pairs exhibited LD for the pooled sample of fish: Prol01 and Prol06; Prol04 and Prol07; Prol01 and Prol19; Prol06 and Prol54; Prol05 and Prol57; and Prol53 and Prol57. There was no evidence of scoring errors due

TABLE 2 Newly validated microsatellite markers for *Prochilodus lineatus*, along with the observed amplified DNA fragment length intervals for each locus and their respective defined allelic classes.

Locus	Primer sequence	Motif	T_m (°C)	Size range (bp)	N_A	N
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: TTCCCCATCTTTAGAAATGTCCTTC 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: AAATCCCATAGTTTGGAGTGACGC	ATCT	58.8	89–276	31	114
Prol06	F: GGCCACTATAAAAGGTAGGCACAA R: CATACGCTGTCCTAAATGTTTCAGC	AGAT	54.0	121–227	17	104
Prol07	F: TGGAGGCTGCAGGTATAAGTTTTT R: AAGGCTGGTTCCTAACCTCC	TTA	56.4	102–205	20	94
Prol08	F: AAGGGCAAGATCAGACAAACAAG R: GAATACTACGCTCATTTTCAAGAAAAA	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: GTTGACCTTTGTAGAGGGGTATG R: ACAGCCGTGGATATAGTCAGAAGC	TCA	56.4	117–280	23	107
Prol19	F: TAAATGTGGTTTCCTGTGCAATG R: TTGCAATTTTATCCATCTTTCCC	TAGA	60.0	87–200	18	106
Prol53	F: ATTGCTGTGTCACTTAAGCATGT R: TGCTACTGGCTGCTAAATTCCTT	AGAT	51.0	127–209	13	137
Prol54	F: CGAGACGTCTAGCTGACACCATAA R: TTGACTTCTTGCTGTCATGTCAA	AGAT	49.0	96–309	28	110
Prol57	F: AATGAAACACAGTCAGGCTCCTTC R: TTCCTTCAGGCAATGCACCTTTAT	ATA	49.0	139–317	27	122

Abbreviations: F, forward primer; N, sample size (successfully genotyped specimens); N_A , number of alleles; R, reverse primer; T_m , primers annealing temperature.

to stuttering or large-allele dropout; however, the presence of null alleles was suggested as a possibility due to the general excess of homozygotes for most allele size classes, for all loci, in all population samples, except for Prol07.

Allelic richness ranged from 10.8 (± 0.8 SE) alleles for Camargos to 16.9 (± 0.9 SE) alleles for Itutinga 2 (Table 3). This trend was confirmed by the jackknife estimates of allelic richness (Supplementary Figure 1), along with the results for the rarefaction analysis (Supplementary Figure 2). Camargos also had the lowest expected heterozygosity, while displaying the highest inbreeding coefficient value, along with the Itutinga 2 sample (Table 3). All population samples exhibited private alleles, with 7 alleles for Camargos and 37 alleles for Itutinga 2 (average 19.7 ± 12.7 SE).

3.3 | Population structure

The AMOVA results showed that only 1.6% of total variation was explained by the four population samples, with the remaining 98.4% being explained by variation among individuals ($\phi_{ST} = 0.016$, $p = 0.001$, degrees of freedom = 139). All pair-wise fixation estimates for G''_{ST} and Jost's D were significantly different from zero, including both samples collected in the same location at different times (Table 4). Itutinga samples (2013 and 2023) had the lowest values, whereas Camargos showed the highest overall pair-wise variation and differentiation for both the respective indices. Considering all 15 microsatellite loci, among the four population samples, the fixation index G''_{ST} reached 0.16 (± 0.03 SE; 0 to 0.26 CI, $n = 999$; $p = 0.001$),

TABLE 3 Genetic diversity results for each microsatellite maker across population samples in *Prochilodus lineatus*.

Sample	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
	Over 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
	Over 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
	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

TABLE 3 (Continued)

Sample	Locus	N	N _A	H _O	H _E	F	HWE (p)
Funil	Prol57	36	22	0.81	0.92	0.13	
	Over 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
	Over loci	32.6	16.67	0.50	0.89	0.44	0
Overall		140	341	0.45	0.91	0.50	0

Note: Probability for Hardy–Weinberg test had significance set at $\alpha = 0.008$. All probability values (p) are significant, except where noted (ns), and values deemed zero are actually extremely low non-zero values <0.00009 .

Abbreviations: F, inbreeding coefficient; HWE, exact test for Hardy–Weinberg equilibrium; N, sample size; N_A, number of alleles; H_O, observed heterozygosity; H_E, expected heterozygosity.

TABLE 4 Summary of pair-wise fixation index G''_{ST} (below diagonal) and Jost's genetic distance D (above diagonal) for microsatellite diversity among population samples of *Prochilodus lineatus* from the Upper Grande River, MG.

	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*	-

* $p < 0.05$.

** $p \leq 0.001$.

and the total genetic differentiation estimate of Jost's D was 0.15 (± 0.03 ; 0 to 0.24 CI, $n = 999$; $p = 0.001$).

MLE of admixture (Figure 3) pointed to $K = 2$ as the most likely number of clusters, according to Wang's D_{LK2} metric. However, F_{STIS} suggested $K = 4$ as the most likely number of groups (Supplementary Figure 3). The first-order admixture analysis allowed the definition of two clusters (with respective individual IDs, Figure 3a). Estimates of D_{LK2} for both these clusters yielded the value of $K = 2$ as optimal in the second-order analysis (Figure 3b), whereas F_{STIS} recorded values of $K = 5$ and $K = 7$ (Supplementary Figure 4). Finally, these four newly defined individual rearrangements were studied in the third-order analysis (Figure 3c), and D_{LK2} consistently pointed to $K = 2$ for

each and every subject cluster (Supplementary Figure 5; individual assignments for all hierarchical orders can be found in Supplementary File 3). We addressed and discarded the possibility of panmixia ($K = 1$) at this and each hierarchical-order level by directly inspecting log-likelihood values of the clustering analysis against the results for the assumed value for K (Supplementary File 4; Supplementary Table 1).

These different MLE sample arrangements showed a consistent decrease in intrapopulation diversity parameters for the higher orders, such as expected heterozygosity and number of alleles (Supplementary File 4; Supplementary Table 2), whereas interpopulation indicators (G''_{ST} and Jost's D) steadily increased (Supplementary File 4; Supplementary Tables 3–5). The HWE analysis of MLE-defined clusters indeed showed a tendency towards Hardy–Weinberg expectations in higher-order group arrangements. In average, third-order clusters had more loci in HWE, with one particular cluster showing eight loci in equilibrium (Supplementary File 5).

The Bayesian approach to admixture also resulted in alternative values for best K estimates. According to Pritchard's $p(X|K)$, the best is $K = 8$, with a log likelihood of -8990.74 , indicating a maximum posterior probability of $p(K|X) \approx 1$ compared to other K scenarios. Concurrently, both Evanno's ΔK and Wang's parsimony estimators resolved $K = 5$. Figure 4 presents individual admixture estimates ordered by population samples across the alternative K scenarios.

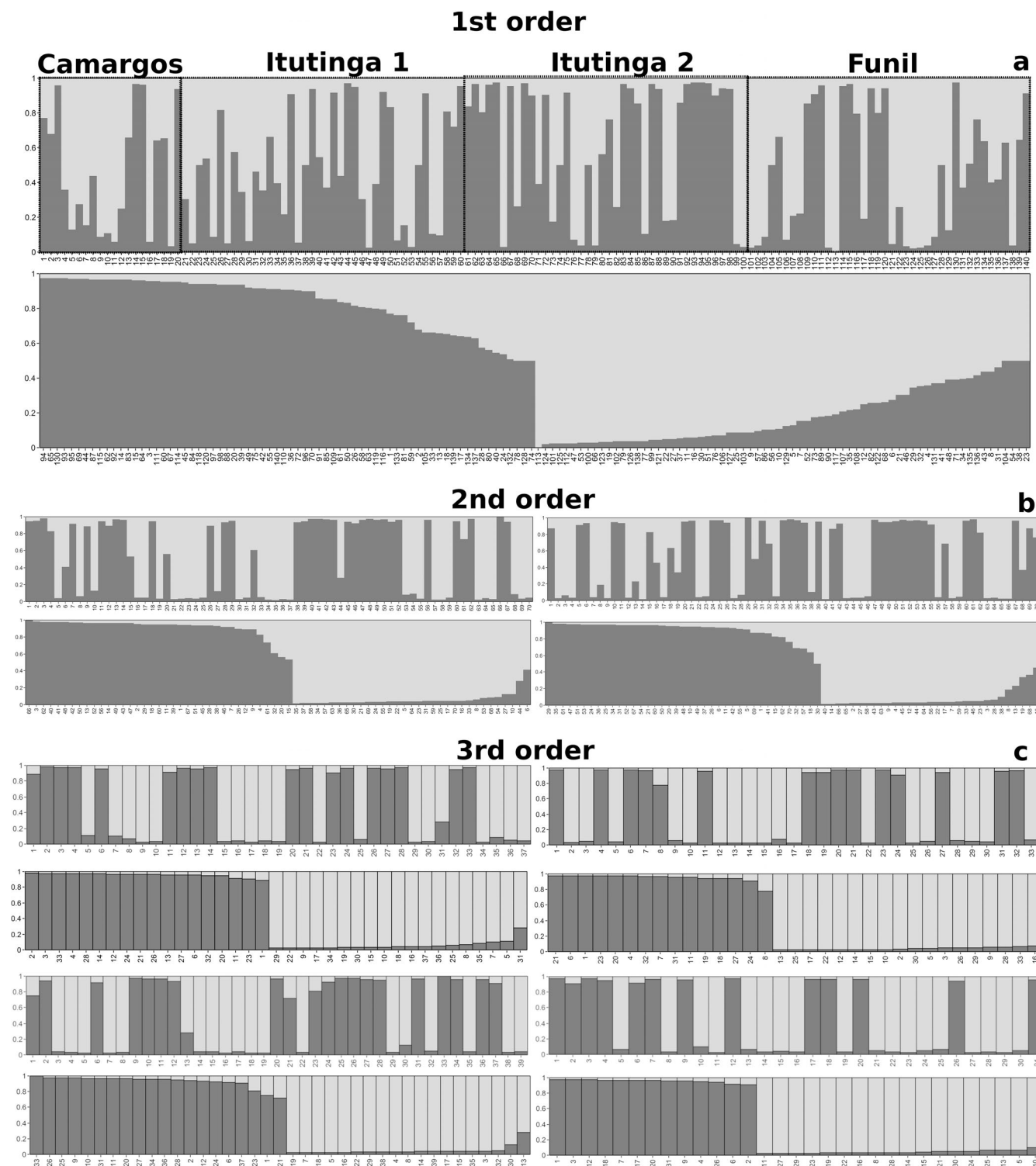


FIGURE 3 Individual admixture coefficients (bars) for $K = 2$ across inclusive levels of hierarchical genetic structure of *Prochilodus lineatus* populations from the Upper Grande River, MG, Brazil. Different shades of grey represent alternative clusters of origin, with the proportional size of bars indicating the respective genomic contributions from ancestral demes. For each panel, the diagram above shows the sample in the original input sequence, whereas the diagram below shows the sample organized as newly defined clusters. (a) First-order analysis, showing fish grouped based on the population sample and original ID. (b) Second-order analysis from clusters defined in the first order. (c) Third-order analysis of the second-order clusters. For second and third orders, the numbers below bars are arbitrary and not equivalent to the original ID.

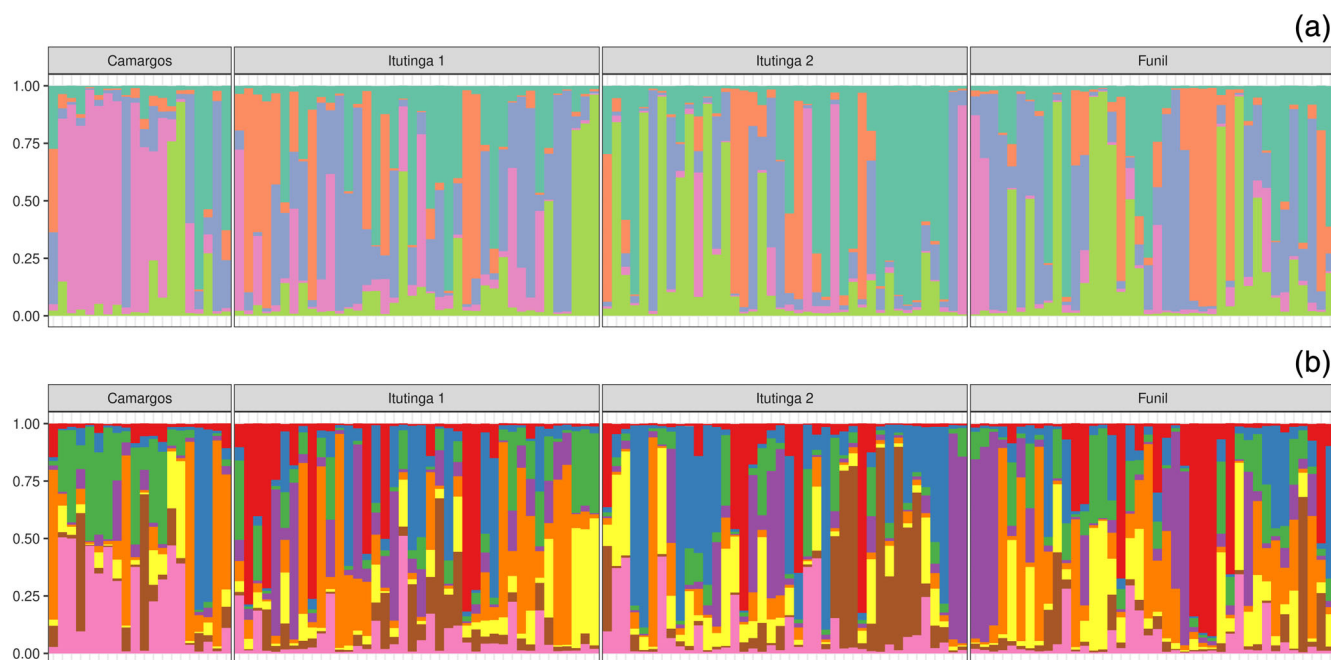


FIGURE 4 Individual admixture plots from the Bayesian clustering results for (a) $K = 5$ and (b) $K = 8$ of *Prochilodus lineatus* from the Upper Grande River, MG. The order of the samples are the incremental ID numbers, as presented in Table 1. Each colour represents a different cluster. Note the apparent differentiation at Camargos and Itutinga 2.

4 | DISCUSSION

4.1 | Genomics, bioinformatics and microsatellite marker development

Genome assemblies have become mandatory for many different biological disciplines (Jung et al., 2020). Our work takes a new step towards this goal by publicly presenting the first comprehensive genomic resources for this highly iconic species, *P. lineatus*. The volume of short-read DNA sequence data presented here surpasses the only previously available FASTQ data for *P. lineatus* (Stornioli et al., 2021) by a factor of 12. This number of sequences is, though, one order of magnitude below the short-read data reported for the trans-Andean congeneric species, *P. magdalenae* (Yepes-Blandón et al., 2022), which, however, is not currently publicly available. The CG content and quality rate of reads were the same as the ones previously presented for other *piracema* fish (e.g., Graciano et al., 2022; Yazbeck et al., 2018). It's also the most extensive dataset of short-reads publicly disclosed so far for any member of the Prochilodontidae family.

This work provides the first de novo genome assembly for *P. lineatus*. The reported assembly is slightly longer than the genome size described for *P. magdalenae* (Yepes-Blandón et al., 2022), despite its extensive fragmentation, and shorter than the size estimated by the k-mer method, which is not always precise (Pflug et al., 2020). The limited length of the short-reads used here contributed to curtailing our genome assembly, primarily motivated by the exclusive search of microsatellite loci. The full sequence of this *P. lineatus* specimen's

mitogenome was previously presented in Santos et al. (2021). This novel genome assembly might be explored for the initial annotation of longer-characterized sequences (e.g., Graciano et al., 2021).

The genome-wide microsatellite set produced here, with more than 37,000 putative loci, can now be free and easily accessed for the rapid, low-cost development of potentially hundreds of new empirically validated markers. It may also be explored for technological development of practical tools for hatcheries and management of impacted areas, such as multiplexed sets of markers (e.g., Carvalho et al., 2018). Some of these resources may also, arguably, be applied in different fishes, particularly in other prochilodontid and characid species (e.g., Barbosa et al., 2006; Carmo et al., 2015). The relatively high rate of success in empirical assays attests to the high quality of this genome-wide microsatellite set. These 15 new DNA markers presented here increase the number of available microsatellites for *P. lineatus* by ~84%.

We observed a variable rate of missing data among these new markers, with some loci (e.g., Prol07) presenting a relatively pronounced amount. The reason for this is likely multifactorial, knowing, the degree of DNA sample conservation, lack of amplification success or refraining from scoring uncertain banding patterns, such as 'shadows' (from extension slippage) and heteroduplexes (cross-annealing of single-stranded DNA from alleles with different repeat counts).

One noteworthy possibility for loci with variable rate of missing data among samples is that it could have been characterized from a supernumerary (or B) chromosome. *P. lineatus* ($2N = 54$) have been

described with a variable number of up to nine B chromosomes (Stornioli et al., 2021). Indeed, a quick BLAST search in the specimen's genome for the PliSat14 probe (NCBI accession MZ161107) associated with B chromosomes in *P. lineatus* (Stornioli et al., 2021) yields several hits (e.g., contigs C526428, C83258, C1638590 and C591000). Nevertheless, we have not yet associated these to any microsatellite marker used in this study.

Fortunately, approaches such as admixture assignments are found to be reasonably insensitive to missing data, especially when it is not due to a systematic cause (Falush et al., 2007; Wang, 2022). Even null alleles might have only a marginal impact (Budowle and Sajantila, 2024; Carlsson, 2008; Putman & Carbone, 2014). For some other applications, the issue of missing data is more critical (e.g., estimating effective population size, N_e – Peel et al., 2013).

The microsatellite Prol53 had a particularly high rate of successful genotype scoring. Despite exhibiting the lowest allelic richness and polymorphic information content, this *locus* had the most weight in explaining the two main clusters from the whole pooled data through a principal component analysis (results not shown).

The number of alleles resolved across the 15 *loci* was analogous to previous reports on the characterization and application of genome-specific microsatellite markers in this fish (e.g., Rueda et al., 2011; Yazbeck & Kalapothakis, 2007). The set of validated markers are deemed as highly informative as all showed PIC values above the arbitrary 0.5 threshold. Despite Microchecker suggesting the possible presence of null alleles for most *loci* in all populations, we consider this an artefact due to the inherent sample substructuring, which leads to critical confounding effects in this analysis (Waples, 2015).

4.2 | Genetic diversity

A study by Wang et al. (2021) suggests that applying fewer than 11 or 12 microsatellite markers may lead to important bias in population genetics structure assessments. The results presented here are the most extensive efforts so far in terms of the number of *loci* analysed in *P. lineatus*. It revealed high levels of microsatellite diversity in fish from the uppermost portion of the Grande River, MG, despite the influence of large dams. This result is consistent with previous studies on the population genetics of *P. lineatus* with comparable genomic sampling efforts (e.g., Perini et al., 2021; Rueda et al., 2013).

Despite the smaller sample size from Camargos (addressed through jackknife and rarefaction analysis for the allelic richness) and the variable rate of genotyping, the genetic diversity analysis within-population samples suggests a lower genetic richness in fish upstream of the Camargos dam. It is the last artificial obstacle towards the headwaters of the Grande River basin. This is an area considered as priority for conservation of migratory fishes in the state of Minas Gerais (Drummond, 2005), and it has been knowingly segregated from the downstream system for the past 65 years since the construction of the Camargos power plant dam. It does not bear a fish passage and larvae, and juveniles are not likely to successfully drift downstream

(Suzuki et al., 2011). Hatchery-based restocking of *P. lineatus* did occur in the region for the past decades (Saraiva & Pompeu, 2016). Yet, prior to more recent scientifically driven actions, fish stocking was mostly executed without purposefully defined goals, other than fulfilling regulatory agencies' licensing demands or in the context of environmental education actions. This portion of the river is still connected to a last major free-flowing upstream Grande River tributary, the Aiuruoca River, MG (Figure 1).

The Camargos dam likely isolates fish from the downstream hydrographic system, and stocking operations might have been limited in providing adequate gene flow connectivity for the system upstream. Its local populations are most likely undergoing stronger random genetic drift, which leads to genetic impoverishment within demes and differentiation among them. This notion is corroborated by the results showing the largest amount of interpopulation variation and differentiation in Camargos from all other samples. Thus, our results potentially support the hypothesis that dams are isolating fish populations in the Upper Grande River. Other competing possibilities for explaining this observed pattern of genetic variability should be noted as DNA marker artefacts (e.g., faulty primer annealing) or sampling bias (smallest sample size). These issue deserves deeper investigation in future research due to its relevant conservation implications.

The little difference among population samples registered by the AMOVA is compatible with a single, almost panmictic population, or could be explained by more complex, cryptic patterns of mixed genetic variability. Yet, it is deemed as less straightforward in the interpretation of hypervariable markers, such as microsatellites (Jost, 2008). The G''_{ST} fixation index, along with the allelic differentiation estimator Jost's D , pointed to significant deviations from panmixia, with all comparisons significantly different from zero. This supports the hypothesis of isolation by dams over fish unable to freely migrate.

The 2023 sample from the Itutinga reservoir (Itutinga 2) is the most genetically diverse followed by fish caught downstream of the Funil dam. The latter, as first raised in Yazbeck and Kalapothakis (2007), could be preventing migrating fish from downstream, potentially of different stocks, of proceeding further upward, despite its fish lift. This virtual *cul-de-sac* could be leading to a higher observed genetic diversity in the sample by accumulation, comparable to the trapped gene pool effect discussed by Perini et al. (2021). Upstream of the Funil dam lie the mouths of two major free-flowing tributaries of the Grande River, the Mortes River and the Capivari River (Figure 1), which could be natural candidates for harbouring different *P. lineatus* stocks (e.g., Suzuki et al., 2011).

The *P. lineatus* population at Itutinga's reservoir, in turn, has undergone genetic variation and differentiation of around 7% in 10 years. Despite being closed and strongly isolated, it seems to have locally accumulated genetic variation (allelic richness) since 2013. A plausible cause for this could be fish stocking of *P. lineatus* from the hatchery situated on its left bank. Hatchery-based stocking, or repeated escapes of broodstock from net-pens in the Camargos reservoir, with varying sources for spawning matrices, along with the marked tendency for genetic bottlenecks during particular

reproductive efforts (single or few spawned couples), could be yielding alternative sets of new alleles being introduced into this limited stretch of the Grande River. Although this location had the highest levels of diversity (N_A and H_E), it also had the highest inbreeding coefficient, reinforcing the notion of hatchery-driven substructuring. For more than 5 years now, though, the hatchery station has been operating mainly for scientific research and has ceased its fish stocking in the Grande River, MG, and surrounding region. These results support the hypothesis that hatcheries can influence the genetic diversity in populations of fish trapped between dams.

We conclude that stocking may be effective in preserving genetic diversity in isolated populations from the Upper Grande River basin. However, it should only be carried out after a thorough genetic characterization of existing stocks (see discussion below), with careful consideration of potential local adaptations. Additionally, hatchery breeding strategies must be meticulously planned, and the process should be closely monitored through efficiency evaluations and follow-up studies.

Regarding our main hypothesis on the occurrence of mixed stocks in the Upper Grande River, MG, the extensive and verified lack of adherence to Hardy-Weinberg proportions in all population samples, always showing a deficit of heterozygotes, along with some verified LD, suggest strong genetic stratification in all sampled populations (i.e., more than one deme being sampled) and, thus, are consistent with its prediction. The fact that it is so pervasive across *loci* allows us to initially disregard other putative confounding factors, such as eventual null alleles. It can be biologically interpreted as an expression of the Wahlund effect, leading to a pronounced downward departure from the expected heterozygosity under HWE.

4.3 | Population structure

The admixture model seems biologically meaningful and appropriate for studying potamodromous Neotropical fishes. These fish might experience multiple reproductive runs over several years, potentially resulting in variable breeding outcomes with respect to both the moment and place of spawning. More importantly, potential genetically distinct individuals and stocks might end up co-occurring in the same locality and time during migratory runs or trapped between adjacent dams. This would isolate them from their original breeding grounds, leading to subsequent exogamy among different demes.

The MLE approach for admixture, which estimates population genetics parameters that maximize a likelihood function (Wang, 2003; Wang, 2022), is a powerful and scalable alternative to the more hegemonic Bayesian inference. Bayesian analysis is based on the conditional probabilities of genotypes, given some population genetics parameter, which is, in turn, highly sensible to often subjective prior probabilities (van de Schoot et al., 2021). Because admixture models are highly hyperdimensional (in our particular case, from 326 independent variables, for $K = 1$, up to 6384, for $K = 14$), the MLE path has the advantage of high accuracy, coupled with lower computational demands (Wang, 2022), which allows for more extensive replication.

Here, it showed a high degree of convergence across replicates (many have zero variance) for different K .

We adopted Wang's D_{LK2} estimator for determining optimal K , as it has been shown to be generally more reliable across a variety of investigated scenarios compared to the F_{STIS} estimator (Wang, 2022). The first-order MLE analysis of admixture initially pointed for two groups, which did not correlate with sampling scheme. Both clusters contain members from all sampled populations. This strongly suggests the occurrence of mixed stocks of *P. lineatus* from the main channel of the Upper Grande River. Caution was taken to dismiss panmixia as the best possibility by directly examining the log-likelihood values of alternative models, as D_{LK2} would not possibly capture $K = 1$ (Wang, 2022), just like Evanno's ΔK for Bayesian estimation (Janes et al., 2017), and given that the F_{STIS} estimator would always be zero (Wang, 2022). The results from the Bayesian analysis of admixture also reinforce the notion of several, different fish stocks in our total sample, not correlated with sampling locations or moments, supporting the hypothesis of the occurrence of historical mixed stocks throughout the Upper Grande River basin.

Different values of K , thus, can reveal multiple underlying aspects of population structure and distinct stages in the population's history (Gehri et al., 2021; Gilbert, 2016; Janes et al., 2017; Wang, 2022). As discussed by Perini et al. (2021) and Ferreira et al. (2023), mixed stocks might reflect the use of different affluent streams as breeding sites. The two major genetically differentiated groups defined in the first-order MLE analysis, then, could be the reflex of the existence of two main ancient breeding grounds (e.g., the Mortes, Capivari or the Aiuruoca rivers), defining forks along migratory routes, splitting groups in the Upper Grande River. This result indicates the probable existence of at least two distinct stocks mixed along the studied area, not explained by the recent implementation of dams. Evidence for mixed stocks in *P. lineatus* was also previously forwarded by Rueda et al. (2013) and Avigliano et al. (2019).

Aiming to deepen the exploration of these seemingly co-occurring genetically mixed stocks of *P. lineatus* from the main channel of the Upper Grande River, the two clusters initially defined by MLE were iteratively unfolded for two more analyses rounds. Consistently, and surprisingly, every examined cluster revealed two newly characterized smaller subgroups, nested within clusters resolved in the immediate previous step (2^3). Judging by the sorted samples of newly defined clusters from Figure 3c, the F_{STIS} estimates from Supplementary Figure 5 or the HWE analysis from MLE clusters, there would be potentially more admixture degrees to be untangled from further hierarchical analysis, although we would reach rather critical sample sizes. The MLE-defined clusters showed the expected admixture model behaviour towards higher orders. The third-order had more genetic homogeneity within clusters and a maximization of genetic variation and distance among clusters. It also showed increasing numbers of *loci* in HWE, attesting to the meaningfulness of the hierarchical admixture analysis and the usefulness of these new microsatellite markers.

The Bayesian exploration of the admixture model suggested deeper levels of substructuring, pointing to as many as eight possible

clusters in the whole sample of fish, a scenario compatible with the eight demes defined during the 'peeling' of MLE layers. The $K = 5$ scenario was favoured based on methods that balance model fit and parsimony. A fair visual agreement of cluster boundaries can be verified for $K = 8$ when placing the samples accordingly to the clusters defined at $K = 5$ (results not shown). This pattern supports the interpretation of nested clusters within clusters consistent with a hierarchical arrangement of the genetic structure of the mixed stocks.

The observed stratified, self-similar, hidden pattern of the population genetic structure across the larger sample of fish may reveal the existence of potential different, well-defined, ancestral breeding grounds, hierarchically nested within differential scale levels, throughout the basin, organized based on fractal geometry (e.g., main river, first-order, second-order tributaries and so on). One parsimonious explanatory hypothesis is that these reiterated subdivisions of clusters into two, across scales, reflect a preferential directional navigation tendency, either towards streams on opposite sides or, alternatively, along cardinal orientations, when individuals encounter forks and tributaries (e.g., left-leaning versus right-leaning migrators or east versus west, and so forth). This asymmetrical bias would operate at multiple levels within the hydrographic system, partly defining the apparent self-similar pattern.

This view also helps reconcile the former notion of a single panmictic population exhibiting high levels of genetic variability, alongside low but significant levels of among-population genetic structure generally reported in this species. One hallmark of a fractal is the capacity to be measured across arbitrary scale levels, with values increasing as higher levels of magnification (i.e., smaller scales) are considered. The observed increase in F_{ST} values at superior hierarchical orders supports this interpretation.

Thus, we propose that populations of *P. lineatus* across the hydrographic landscape may be regarded as *quasi*-demes, exhibiting almost random mating across arbitrary scale levels throughout its distribution, potentially becoming progressively more similar to true demes as smaller tributaries are examined.

If this association is confirmed, we can conclude that it may constitute the first indirect genetic evidence of a strong degree of homing behaviour in a Characiformes *piracema* fish species. Philopatric spawning, or homing behaviour, a thoroughly investigated issue in salmonid and other migratory species, is still poorly understood in the Neotropics. It has been documented for Neotropical catfishes (e.g., Batista & Alves-Gomes, 2006; Pereira et al., 2009; Hauser et al., 2020). But most importantly, it was suggested for the congeneric *P. argenteus*, based on radio-tagging results from the São Francisco River basin (Godinho & Kynard, 2006) and discussed for *P. lineatus* itself by Careaga and Carvajal-Vallejos (2019), from morphometric data. Our work revives this area of inquiry for migratory fish in South America.

Being particularly abundant and a conspicuously competent, far-ranging migratory species, we argue that in a pre-damming past, *P. lineatus* could probably reach the innermost free-flowing tributaries for spawning, upon the colossal La Plata basin, which cuts through a diversified range of biomes and microhabitats. We hypothesize that

this would likely point towards philopatric behaviour as an evolutionary stable strategy due to the possible occurrence of distinct local adaptations, as it has been shown in other fish species (e.g., Mobley et al., 2019; Salles et al., 2016; Vøllestad & Primmer, 2019).

In conservation biology, the understanding of degrees of admixture is important in management decisions (Wang, 2003; Wang, 2022). The confirmation of mixed stocks and the hierarchical genetically stratified nature of *P. lineatus* from the main channel of an important large and intensely dammed river will likely enlighten its future research and fisheries management. If further corroborated, these findings, particularly its ensuing propositions, would have direct implications to environmental and development decision-making. It highlights the possible central relevance of smaller key tributaries (actively sought for the implementation of new small hydroelectric power plants, e.g., Ferreira et al., 2022) for the conservation of genetic variation in a socio-environmentally relevant species, as it highlights the importance of free-flowing tributaries of different orders throughout the basin.

On a broader level, we raise the question whether these apparent fractal layers from the genetic structure could imply in a more widespread phenomena for *piracema*, other migratory fishes or even other non-aquatic migratory organisms, showing a temporary mixture, but segregated breeding grounds. We believe this is worthy of further investigation, as fractal structures are pervasive in nature, not only in biology (e.g., Bassingthwaite, 1992; Leggett et al., 2019; Meyer et al., 2020; Sendker et al., 2024), but it is also conspicuously present in major geographical landscape attributes like mountain ranges, forest borders, rivers and coastlines (Andronache et al., 2019; Xu et al., 1993). The only direct discussions on the relationship between fractals and genetic population structure were laid by Wingen et al. (2007), based on computer simulation results of long-distance dispersal of clonal fungal spores, and by Lee (2020), who explored the fractal dimension of human genetic variability data both within and among genomes.

Our results will very likely benefit researchers and environmental managers in tackling challenging issues related to the conservation of Neotropical migratory fishes, as they support the hypothesis of the isolating effects of dams and local influence of fish stocking over genetic variability. The former seems to have been partially alleviated by the availability of tributaries around affected areas, as previously acknowledged in other *P. lineatus* genetics studies. More importantly, the results presented here point to the potential role of tributaries of different orders in harbouring genetically diverse (possibly unique) populations of migratory freshwater fishes.

AUTHOR CONTRIBUTIONS

Gabriel M. Yazbeck conceived, conducted and supervised the study, raised funding, performed bioinformatics and population genetics analyses, wrote and edited the manuscript; Gélica Aparecida Santana Nascimento performed marker development and population genetics analysis and wrote the manuscript; Leiliane Campos Carvalho performed marker development; Raíssa Cristina Dias Graciano performed bioinformatics analysis; Rosiane de Paula Santos performed

bioinformatics analysis; Rafael Sachetto Oliveira supervised and performed bioinformatics pipelines. All contributors read and agreed upon the final manuscript.

ACKNOWLEDGEMENTS

We thank the fishermen couple Donizete Monteiro de Paula and Sônia Bárbara de Oliveira Paula, biologists and photographers Lucia and André Seale, biologists João de Magalhães Lopes, Miriam Aparecida de Castro, Raquel Coelho Loures and everyone at Programa Peixe Vivo, José Mauro Ribeiro for bioinformatics assistance, Mônica Soares da Silva for assistance with marker development, Fausto Moreira da Silva Carmo for assistance with specimen collection and DNA extraction. We are also grateful to Pedro Manoel Galetti Jr., Paulo De Marco Jr., Fabrício Rodrigues dos Santos and Paulo dos Santos Pompeu for their valuable comments. This work was financed by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq grant 303023/2023-6), Companhia Energética de Minas Gerais and Agência Nacional de Energia Elétrica (CEMIG-ANEEL grant GT345), by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES finance Code 001) and by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG grant APQ-04569-10). The Article Processing Charge for the publication of this research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (ROR identifier: 00x0ma614).

FUNDING INFORMATION

This study was funded by Conselho Nacional de Pesquisa e Desenvolvimento Tecnológico (CNPq), Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Companhia Energética de Minas Gerais (CEMIG), Agência Nacional de Energia Elétrica (ANEEL) and Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG).

ORCID

Gabriel M. Yazbeck  <https://orcid.org/0000-0003-1729-348X>

Gésica Aparecida Santana Nascimento  <https://orcid.org/0009-0002-2779-0834>

Leiliane Campos Carvalho  <https://orcid.org/0009-0004-1177-4631>

Raíssa Cristina Dias Graciano  <https://orcid.org/0000-0002-7481-1135>

Rosiane de Paula Santos  <https://orcid.org/0000-0003-3233-6863>

Rafael Sachetto Oliveira  <https://orcid.org/0000-0003-0800-5984>

REFERENCES

- Alix, M., Kjesbu, O. S., & Anderson, K. C. (2020). From gametogenesis to spawning: How climate-driven warming affects teleost reproductive biology. *Journal of Fish Biology*, 97(3), 607–632. <https://doi.org/10.1111/jfb.14439>
- Andronache, I., Marin, M., Fischer, R., Ahammer, H., Radulovic, M., Ciobotaru, A.-M., Jelinek, H. F., Di Ieva, A., Pintili, R.-D., Drăghici, C.-C., Herman, G. V., Nicula, A.-S., Simion, A.-G., Loghin, I.-V., Diaconu, D.-C., & Peptenatu, D. (2019). Dynamics of Forest fragmentation and connectivity using particle and fractal analysis. *Scientific Reports*, 9(1), 12228. <https://doi.org/10.1038/s41598-019-48277-z>
- Astashyn, A., Tvedte, E. S., Sweeney, D., Sapojnikov, V., Bouk, N., Joukov, V., Mozes, E., Strobe, P. K., Sylla, P. M., Wagner, L., Bidwell, S. L., Brown, L. C., Clark, K., Davis, E. W., Smith-White, B., Hlavina, W., Pruitt, K. D., Schneider, V. A., & Murphy, T. D. (2024). Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biology*, 25(1), 60. <https://doi.org/10.1186/s13059-024-03198-7>
- Avigliano, E., Pisonero, J., Bordel, N., Dománico, A., & Volpedo, A. V. (2019). 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*, 99(6), 1429–1433. <https://doi.org/10.1017/S002531541900016X>
- Barbosa, A. C. D. R., Corrêa, T. C., Galzerani, F., Galetti, P. M., Jr., & Hatanaka, T. (2006). Thirteen polymorphic microsatellite loci in the Neotropical fish *Prochilodus argenteus* (Characiformes, Prochilodontidae). *Molecular Ecology Notes*, 6(3), 936–938. <https://doi.org/10.1111/j.1471-8286.2006.01406.x>
- Bassingthwaight, J. B. (1992). Fractal vascular growth patterns. *Acta Stereologica*, 11(Suppl 1), 305–319.
- Batista, J. S., & Alves-Gomes, J. A. (2006). Phylogeography of *Brachyplatystoma rousseauxii* (Siluriformes—Pimelodidae) in the Amazon Basin offers preliminary evidence for the first case of “homings” for an Amazonian migratory catfish. *Genetics and Molecular Research*, 5(4), 723–740.
- Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32(3), 314–331.
- Bowen, S. H. (2022a). Digestion and assimilation of benthic biofilm by the Sábalo, *Prochilodus lineatus*. *Journal of Fish Biology*, 100(1), 107–116. <https://doi.org/10.1111/jfb.14924>
- Bowen, S. H. (2022b). The river flood pulse, benthic biofilm, and the nutrition of *Prochilodus lineatus*. *Environmental Biology of Fishes*, 105(2), 213–230. <https://doi.org/10.1007/s10641-022-01211-1>
- Brandström, M., & Ellegren, H. (2008). Genome-wide analysis of microsatellite polymorphism in chicken circumventing the ascertainment bias. *Genome Research*, 18(6), 881–887. <https://doi.org/10.1101/gr.075242.107>
- Budowle, B., & Sajantila A. (2024). Short Tandem Repeats - How Microsatellites Became the Currency of Forensic Genetics. *Nature Reviews Genetics*, 25(7), 450–51. <https://doi.org/10.1038/s41576-024-00721-1>
- Bushnell, B. (2014). *BBMap: A Fast, Accurate, Splice-Aware Aligner* (No. LBNL-7065E; Lawrence Berkeley National Laboratory. LBNL Report, p. 3). Lawrence Berkeley National Laboratory. <https://escholarship.org/uc/item/1h3515gn>.
- Careaga, M., & Carvajal-Vallejos, F. (2019). Los sábalo (Characiformes: Prochilodontidae: Prochilodus) de Bolivia (pp. 171–185).
- Carlsson, J. (2008). Effects of microsatellite null alleles on assignment testing. *The Journal of Heredity*, 99(6), 616–623. <https://doi.org/10.1093/jhered/esn048>
- Carmo, A. O., Brandão Dias, P. F. P., Martins, A. P. V., Bedore, A. G., & Kalapothakis, E. (2016). Complete mitochondrial genome sequence of *Prochilodus lineatus* (Characiformes, Prochilodontidae). *Mitochondrial DNA Part A, DNA Mapping, Sequencing, and Analysis*, 27(3), 1946–1947. <https://doi.org/10.3109/19401736.2014.971300>
- Carmo, F. M. d. S., Polo, É. M., da Silva, M. A., & Yazbeck, G. d. M. (2015). Optimization of heterologous microsatellites in Piracanjuba. *Pesquisa Agropecuária Brasileira*, 50, 1236–1239. <https://doi.org/10.1590/S0100-204X2015001200015>
- Carolsfeld, J., Harvey, B., Carmen, R., & Anton, B. (2003). *Migratory fishes of South America: Biology, fisheries and conservation status*. IDRC.
- Carvalho, L. C., Ribeiro, J. M., & Yazbeck, G. M. (2018). Microsatellite multiplex systems for *Brycon orbignyanus* (Characiformes: Bryconidae). *Brazilian Journal of Biology*, 79, 751–754. <https://doi.org/10.1590/1519-6984.192528>
- Casemiro, F. A. S., Albert, J. S., Antonelli, A., Menegotto, A., Wüest, R. O., Cerezer, F., Coelho, M. T. P., Reis, R. E., Tan, M., Tagliacollo, V.,

- Bailly, D., da Silva, V. F. B., Frota, A., da Graça, W. J., Ré, R., Ramos, T., Oliveira, A. G., Dias, M. S., Colwell, R. K., ... Graham, C. H. (2023). Landscape dynamics and diversification of the megadiverse south American freshwater fish fauna. *Proceedings of the National Academy of Sciences of the United States of America*, 120(2), e2211974120. <https://doi.org/10.1073/pnas.2211974120>
- Chagas, A. T. D. A., Carmo, A. O., Costa, M. A., Resende, L. C., Brandão Dias, P. F. P., Martins, A. P. V., & Kalapothakis, E. (2016). Description and comparison of two economically important fish species mitogenomes: *Prochilodus argenteus* and *Prochilodus costatus* (Characiformes, Prochilodontidae). *Mitochondrial DNA Part A DNA Mapping, Sequencing, and Analysis*, 27(4), 2852–2853. <https://doi.org/10.3109/19401736.2015.1053125>
- Challis, R. (2017). *Rjchallis/assembly-stats 17.02* (Version 17.02) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.322347>
- Colwell, R. K., & Elsensohn, J. E. (2014). EstimateS turns 20: Statistical estimation of species richness and shared species from samples, with non-parametric extrapolation. *Ecography*, 37(6), 609–613. <https://doi.org/10.1111/ecog.00814>
- Drummond, G. M. (2005). *Biodiversidade em Minas Gerais: Um atlas para sua conservação*. Fundação Biodiversitas.
- Endruweit, M. (2014). Four new records of fish species (Cypriniformes: Nemacheilidae, Balitoridae; Characiformes: Prochilodontidae) and corrections of two misidentified fish species (Tetraodontiformes: Tetraodontidae; Beloniformes: Belontiidae) in Yunnan, China. *Dong Wu Xue Yan Jiu = Zoological Research*, 35(1), 51–58. <https://doi.org/10.11813/j.issn.0254-5853.2014.1.051>
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
- Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology Notes*, 7(4), 574–578. <https://doi.org/10.1111/j.1471-8286.2007.01758.x>
- Ferreira, D. G., Galindo, B. A., Limeira, D. M., Garla, A., Marques, A. J. C., Moreira, A. A., Rodrigues, M. V. H., Clemente, R. C., Shimpó, J. Y., Martins, G. A., Almeida, F. S., Apolinário-Silva, C., Zanatta, A. S., Sofia, S. H., & Souza-Shibatta, L. (2023). 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*, 21(4), e230080. <https://doi.org/10.1590/1982-0224-2023-0080>
- Ferreira, D. G., Souza-Shibatta, L., Shibatta, O. A., Sofia, S. H., Carlsson, J., Dias, J. H. P., Makrakis, S., & Makrakis, M. C. (2017). Genetic structure and diversity of migratory freshwater fish in a fragmented Neotropical river system. *Reviews in Fish Biology and Fisheries*, 27(1), 209–231. <https://doi.org/10.1007/s11160-016-9441-2>
- Ferreira, M. E., Nogueira, S. H. d. M., Latrubesse, E. M., Macedo, M. N., Callisto, M., Bezerra Neto, J. F., & Fernandes, G. W. (2022). Dams pose a critical threat to Rivers in Brazil's Cerrado hotspot. *Water (Basel)*, 14(22), 22. <https://doi.org/10.3390/w14223762>
- Gehri, R. R., Gruenthal, K., & Larson, W. A. (2021). It's complicated: Heterogeneous patterns of genetic structure in five fish species from a fragmented river suggest multiple processes can drive differentiation. *Evolutionary Applications*, 14(8), 2079–2097. <https://doi.org/10.1111/eva.13268>
- Gilbert, K. J. (2016). Identifying the number of population clusters with structure: Problems and solutions. *Molecular Ecology Resources*, 16(3), 601–603. <https://doi.org/10.1111/1755-0998.12521>
- Godinho, A. L., & Kynard, B. (2006). Migration and spawning of radio-tagged *Prochilodus argenteus* in a dammed Brazilian river. *Transactions of the American Fisheries Society*, 135(3), 811–824. <https://doi.org/10.1577/T04-176.1>
- Graciano, R. C. D., Oliveira, R. S., Santos, I. M., & Yazbeck, G. M. (2022). Genomic resources for *Salminus brasiliensis*. *Frontiers in Genetics*, 13, 855718. <https://doi.org/10.3389/fgene.2022.855718>
- Graciano, R., Oliveira, R. S., dos Santos, I. M., & Yazbeck, G. d. M. (2021). Functional DNA annotation from a preliminary de novo genome assembly of *Brycon orbignyanus*, an endangered Neotropical migratory fish. *Latin American Data in Science*, 1(2), 2. <https://doi.org/10.53805/lads.v1i2.12>
- Hauser, M., Duponchelle, F., Hermann, T. W., Limburg, K. E., Castello, L., Stewart, D. J., Torrente-Vilara, G., García-Vásquez, A., García-Davila, C., Pouilly, M., Pecheyran, C., Ponzevera, E., Renno, J.-F., Moret, A. S., & Doria, C. R. C. (2020). Unmasking continental natal homing in goliath catfish from the upper Amazon. *Freshwater Biology*, 65(2), 325–336. <https://doi.org/10.1111/fwb.13427>
- Hedrick, P. W. (2005). A standardized genetic differentiation measure. *Evolution*, 59(8), 1633–1638. <https://doi.org/10.1111/j.0014-3820.2005.tb01814.x>
- Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The $K = 2$ conundrum. *Molecular Ecology*, 26(14), 3594–3602. <https://doi.org/10.1111/mec.14187>
- Jombart, T. (2008). ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jost, L. (2008). G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, 17(18), 4015–4026. <https://doi.org/10.1111/j.1365-294X.2008.03887.x>
- Jung, H., Ventura, T., Chung, J. S., Kim, W.-J., Nam, B.-H., Kong, H. J., Kim, Y.-O., Jeon, M.-S., & Eyun, S. (2020). Twelve quick steps for genome assembly and annotation in the classroom. *PLoS Computational Biology*, 16(11), e1008325. <https://doi.org/10.1371/journal.pcbi.1008325>
- Kalous, L., Bui, A. T., Petrtýl, M., Bohlen, J., & Chaloupková, P. (2012). The south American freshwater fish *Prochilodus lineatus* (Actinopterygii: Characiformes: Prochilodontidae): New species in Vietnamese aquaculture: *Prochilodus lineatus* in Vietnam. *Aquaculture Research*, 43(6), 955–958. <https://doi.org/10.1111/j.1365-2109.2011.02895.x>
- Lee, C.-Y. (2020). The fractal dimension as a measure for characterizing genetic variation of the human genome. *Computational Biology and Chemistry*, 87(August), 107278. <https://doi.org/10.1016/j.compbiolchem.2020.107278>
- Leggett, S. E., Neronha, Z. J., Bhaskar, D., Sim, J. Y., Perdikari, T. M., & Wong, I. Y. (2019). Motility-limited aggregation of mammary epithelial cells into fractal-like clusters. *Proceedings of the National Academy of Sciences*, 116(35), 17298–17306. <https://doi.org/10.1073/pnas.1905958116>
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics (Oxford, England)*, 31(10), 1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lopes, J. d. M., Alves, C. B. M., Peressin, A., & Pompeu, P. S. (2018). Influence of rainfall, hydrological fluctuations, and lunar phase on spawning migration timing of the Neotropical fish *Prochilodus costatus*. *Hydrobiologia*, 818(1), 145–161. <https://doi.org/10.1007/s10750-018-3601-4>
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., ... Wang, J. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Giga-Science*, 1, 18. <https://doi.org/10.1186/2047-217X-1-18>
- Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6), 764–770. <https://doi.org/10.1093/bioinformatics/btr011>

- Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources*, 11(1), 5–18. <https://doi.org/10.1111/j.1755-0998.2010.02927.x>
- Meyer, H. V., Dawes, T. J. W., Serrani, M., Bai, W., Tokarczuk, P., Cai, J., de Marvao, A., Henry, A., Lumbers, R. T., Gierten, J., Thumberger, T., Wittbrodt, J., Ware, J. S., Rueckert, D., Matthews, P. M., Prasad, S. K., Costantino, M. L., Cook, S. A., Birney, E., & O'Regan, D. P. (2020). Genetic and functional insights into the fractal structure of the heart. *Nature*, 584(7822), 589–594. <https://doi.org/10.1038/s41586-020-2635-8>
- Mobley, K. B., Granroth-Wilding, H., Ellmen, M., Vähä, J.-P., Aykanat, T., Johnston, S. E., Orell, P., Erkinaro, J., & Primmer, C. R. (2019). Home ground advantage: Local Atlantic salmon have higher reproductive fitness than dispersers in the wild. *Science Advances*, 5(2), eaav1112. <https://doi.org/10.1126/sciadv.aav1112>
- Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26(3), 419–420. <https://doi.org/10.1093/bioinformatics/btp696>
- Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in excel. Population genetic software for teaching and research—An update. *Bioinformatics*, 28(19), 2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Peel, D., Waples, R. S., Macbeth, G. M., Do, C., & Ovenden, J. R. (2013). Accounting for missing data in the estimation of contemporary genetic effective population size (N_e). *Molecular Ecology Resources*, 13(2), 243–253. <https://doi.org/10.1111/1755-0998.12049>
- Pereira, L. H. G., Foresti, F., & Oliveira, C. (2009). Genetic structure of the migratory catfish *Pseudoplatystoma corruscans* (Siluriformes: Pimelodidae) suggests homing behaviour. *Ecology of Freshwater Fish*, 18(2), 215–225. <https://doi.org/10.1111/j.1600-0633.2008.00338.x>
- Perini, V. R., Paschoalini, A. L., Bazzoli, N., Rizzo, E., & Carvalho, D. C. (2021). Metapopulation dynamics of the migratory fish *Prochilodus lineatus* (Characiformes: Prochilodontidae) in a lotic remnant of the Grande River, southeastern Brazil. *Neotropical Ichthyology*, 19(4), e200046. <https://doi.org/10.1590/1982-0224-2020-0046>
- Petry, A. C., Agostinho, A. A., & Gomes, L. C. (2003). Fish assemblages of tropical floodplain lagoons: Exploring the role of connectivity in a dry year. *Neotropical Ichthyology*, 1, 111–119. <https://doi.org/10.1590/S1679-62252003000200005>
- Pflug, J. M., Holmes, V. R., Burrus, C., Johnston, J. S., & Maddison, D. R. (2020). Measuring genome sizes using read-depth, k-mers, and flow cytometry: Methodological comparisons in beetles (Coleoptera). *G3: Genes, Genomes, Genetics*, 10(9), 3047–3060. <https://doi.org/10.1534/g3.120.401028>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution*, 4(22), 4399–4428. <https://doi.org/10.1002/ece3.1305>
- R Core Team. (2024). R: A Language and Environment for Statistical Computing (Version 4.4.1) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org>
- Reis, R. E., Albert, J. S., Di Dario, F., Mincarone, M. M., Petry, P., & Rocha, L. A. (2016). Fish biodiversity and conservation in South America. *Journal of Fish Biology*, 89(1), 12–47. <https://doi.org/10.1111/jfb.13016>
- Ribolli, J., Miño, C. I., Scaranto, B. M. S., Reynalte-Tataje, D. A., & Zaniboni Filho, E. (2020). Genetic evidence supports polygamous mating system in a wild population of *Prochilodus Lineatus* (Characiformes: Prochilodontidae), a Neotropical shoal Spawner fish. *Neotropical Ichthyology*, 18(2), e190123. <https://doi.org/10.1590/1982-0224-2019-0123>
- Rosa, I. F. D., Oliveira, D. J. D., Cruz, V. P. D., Ashikaga, F. Y., Costa, G. O., Doretto, L. B., Senhorini, J. A., Rocha, R. D. C., Foresti, F. P., Oliveira, C., & Foresti, F. (2022). Temporal genetic structure of a stock of *Prochilodus lineatus* (Characiformes: Prochilodontidae) in the Mogi-Guaçu River ecosystem, São Paulo, Brazil. *Neotropical Ichthyology*, 20(2), e210156. <https://doi.org/10.1590/1982-0224-2021-0156>
- Rousset, F. (2008). genepop'007: A complete re-implementation of the genepop software for windows and Linux. *Molecular Ecology Resources*, 8(1), 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Rueda, E. C., Carriquiriborde, P., Monzón, A. M., Somoza, G. M., & Ortí, G. (2013). Seasonal variation in genetic population structure of sábalo (FRue) in the lower Uruguay River. *Genetica*, 141(7–9), 401–407. <https://doi.org/10.1007/s10709-013-9739-0>
- Rueda, E. C., Sommer, J., Scarabotti, P., Markariani, R., & Ortí, G. (2011). Isolation and characterization of polymorphic microsatellite loci in the migratory freshwater fish *Prochilodus lineatus* (Characiformes: Prochilodontidae). *Conservation Genetics Resources*, 3(4), 681–684. <https://doi.org/10.1007/s12686-011-9432-5>
- Salles, O. C., Pujol, B., Maynard, J. A., Almany, G. R., Berumen, M. L., Jones, G. P., Saenz-Agudelo, P., Srinivasan, M., Thorrold, S. R., & Planes, S. (2016). First genealogy for a wild marine fish population reveals multigenerational philopatry. *Proceedings of the National Academy of Sciences*, 113(46), 13245–13250. <https://doi.org/10.1073/pnas.1611797113>
- Santos, R. P., Melo, B. F., Yazbeck, G. M., Oliveira, R. S., Hilário, H. O., Prosdoci, F., & Carvalho, D. C. (2021). Diversification of *Prochilodus* in the eastern Brazilian shield: Evidence from complete mitochondrial genomes (Teleostei, Prochilodontidae). *Journal of Zoological Systematics and Evolutionary Research*, 59(5), 1053–1063. <https://doi.org/10.1111/jzs.12475>
- Saraiva, S. O., & Pompeu, P. S. (2016). Fish hatchery and its effects on the morphology of *Prochilodus lineatus* (Actinopterygii: Prochilodontidae). *Brazilian Journal of Biology*, 76(1), 209–217. <https://doi.org/10.1590/1519-6984.18514>
- Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., ... Sherry, S. T. (2022). Database resources of the National Center for biotechnology information. *Nucleic Acids Research*, 50(D1), D20–D26. <https://doi.org/10.1093/nar/gkab1112>
- Sendker, F. L., Lo, Y. K., Heimerl, T., Bohn, S., Persson, L. J., Mais, C.-N., Sadowska, W., Paczia, N., Nußbaum, E., del Carmen Sánchez Olmos, M., Forchhammer, K., Schindler, D., Erb, T. J., Benesch, J. L. P., Marklund, E. G., Bange, G., Schuller, J. M., & Hochberg, G. K. A. (2024). Emergence of fractal geometries in the evolution of a metabolic enzyme. *Nature*, 628(8009), 894–900. <https://doi.org/10.1038/s41586-024-07287-2>
- Sivasundar, A., Bermingham, E., & Ortí, G. (2001). Population structure and biogeography of migratory freshwater fishes (*Prochilodus*: Characiformes) in major south American rivers. *Molecular Ecology*, 10(2), 407–417. <https://doi.org/10.1046/j.1365-294x.2001.01194.x>
- Stornioli, J. H. F., Goes, C. A. G., Calegari, R. M., dos Santos, R. Z., Giglio, L. M., Foresti, F., Oliveira, C., Penitente, M., Porto-Foresti, F., & Utsunomia, R. (2021). The B chromosomes of *Prochilodus lineatus* (Teleostei, Characiformes) are highly enriched in satellite DNAs. *Cells*, 10(6), 1527. <https://doi.org/10.3390/cells10061527>
- Suzuki, F. M., Pires, L. V., & Pompeu, P. S. (2011). Passage of fish larvae and eggs through the Funil, Itutinga and Camargos reservoirs on the upper Rio Grande (Minas Gerais, Brazil). *Neotropical Ichthyology*, 9, 617–622. <https://doi.org/10.1590/S1679-62252011000300014>
- Vähä, J.-P., Erkinaro, J., Niemelä, E., & Primmer, C. R. (2007). Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology*, 16(13), 2638–2654. <https://doi.org/10.1111/j.1365-294X.2007.03329.x>
- van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märten, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., & Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews*

- Methods Primers, 1(1), 1–26. <https://doi.org/10.1038/s43586-020-00001-2>
- van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2004). Micro-checker: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4(3), 535–538. <https://doi.org/10.1111/j.1471-8286.2004.00684.x>
- Vøllestad, L. A., & Primmer, C. R. (2019). Understanding local adaptation in a freshwater salmonid fish: Evolution of a research programme. *ICES Journal of Marine Science*, 76(6), 1404–1414. <https://doi.org/10.1093/icesjms/fsz037>
- Wang, H., Yang, B., Wang, H., & Xiao, H. (2021). Impact of different numbers of microsatellite markers on population genetic results using SLAF-seq data for *rhododendron* species. *Scientific Reports*, 11(1), 8597. <https://doi.org/10.1038/s41598-021-87945-x>
- Wang, J. (2003). Maximum-likelihood estimation of admixture proportions from genetic data. *Genetics*, 164(2), 747–765. <https://doi.org/10.1093/genetics/164.2.747>
- Wang, J. (2019). A parsimony estimator of the number of populations from a STRUCTURE-like analysis. *Molecular Ecology Resources*, 19(4), 970–981. <https://doi.org/10.1111/1755-0998.13000>
- Wang, J. (2022). Fast and accurate population admixture inference from genotype data from a few microsatellites to millions of SNPs. *Heredity*, 129(2), 79–92. <https://doi.org/10.1038/s41437-022-00535-z>
- Waples, R. S. (2015). Testing for hardy–Weinberg proportions: Have we lost the plot? *Journal of Heredity*, 106(1), 1–19. <https://doi.org/10.1093/jhered/esu062>
- Wenne, R. (2023). Microsatellites as molecular markers with applications in exploitation and conservation of aquatic animal populations. *Genes*, 14(4), 808. <https://doi.org/10.3390/genes14040808>
- Wingen, L. U., Brown, J. K. M., & Shaw, M. W. (2007). The population genetic structure of clonal organisms generated by exponentially bounded and fat-tailed dispersal. *Genetics*, 177(1), 435–448. <https://doi.org/10.1534/genetics.107.077206>
- Xu, T., Moore, I. D., & Gallant, J. C. (1993). Fractals, fractal dimensions and landscapes—A review. *Geomorphology*, 8(4), 245–262. [https://doi.org/10.1016/0169-555X\(93\)90022-T](https://doi.org/10.1016/0169-555X(93)90022-T)
- Yazbeck, G. M., & Kalapothakis, E. (2007). Isolation and characterization of microsatellite DNA in the piracema fish *Prochilodus lineatus* (Characiformes). *Genetics and Molecular Research*, 6(4), 1026–1034.
- Yazbeck, G. M., Graciano, R. C. D., Santos, R. P., & Oliveira, R. S. (2024). Let them eat cake: When the small aims at being LARGE or the empowering effects of bioinformatics in NGS wonderland. *Journal of Information and Data Management*, 15(1), 1. <https://doi.org/10.5753/jidm.2024.2626>
- Yazbeck, G. M., Oliveira, R. S., Ribeiro, J. M., Graciano, R. D., Santos, R. P., Carmo, F. M. S., & Lavenier, D. (2018). A broad genomic panel of microsatellite loci from *Brycon orbignyanus* (Characiformes: Bryconidae) an endangered migratory Neotropical fish. *Scientific Reports*, 8(1), 8511. <https://doi.org/10.1038/s41598-018-26623-x>
- Yepes-Blandón, J. A., Bian, C., Benítez-Galeano, M. J., Aristizabal-Regino, J. L., Estrada-Posada, A. L., Mir, D., Vásquez-Machado, G., Atencio-García, V. J., Shi, Q., & Rodríguez-Osorio, N. (2022). Draft genome assembly for the colombian freshwater bocachico fish, *Prochilodus magdalenae*. *Frontiers in Genetics*, 13, 989788. <https://doi.org/10.3389/fgene.2022.989788>
- Zellhuber, A. (2016). Environmental policy in Brazil. Tensions between conservation and the ideology of growth. In D. de la Fontaine & T. Stehnen (Eds.), *The political system of Brazil* (pp. 329–350). Springer. https://doi.org/10.1007/978-3-642-40023-0_18

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Yazbeck, G. M., Nascimento, G. A. S., Carvalho, L. C., Graciano, R. C. D., de Paula Santos, R., & Oliveira, R. S. (2026). A fractal pattern of hierarchical genetic population structure in mixed stocks across fish segregated by dams revealed by genomic resources for curimba *Prochilodus lineatus*. *Journal of Fish Biology*, 108(4), 1029–1046. <https://doi.org/10.1111/jfb.70278>