UNIVERSIDADE ESTADUAL PAULISTA – UNESP CENTRO DE AQUICULTURA DA UNESP

DEFESA DE DOUTORADO

RECURSOS GENÔMICOS PARA *Megaleporinus macrocephalus*: SUBSÍDIOS PARA AQUICULTURA

Carolina Heloisa de Souza Borges

Jaboticabal, São Paulo

2022

UNIVERSIDADE ESTADUAL PAULISTA – UNESP CENTRO DE AQUICULTURA DA UNESP

DEFESA DE DOUTORADO

RECURSOS GENÔMICOS PARA *Megaleporinus macrocephalus*: SUBSÍDIOS PARA AQUICULTURA

Carolina Heloisa de Souza Borges

Orientador: Dr. Diogo Teruo Hashimoto

Coorientador: Dr. Ricardo Utsunomia

Tese apresentada ao Programa de Pós-graduação em Aquicultura do Centro de Aquicultura da UNESP - CAUNESP, como parte dos requisitos para obtenção do título de Doutor em Aquicultura

Jaboticabal, São Paulo

2022

Borges, Carolina Heloisa de Souza
D155a Recursos genômicos para *Megaleporinus macrocephalus*: subsídios para aquicultura / Carolina Heloisa de Souza Borges. – – Jaboticabal, 2022
x, 127 p. : il. ; 29 cm

Tese (doutorado) - Universidade Estadual Paulista, Centro de Aquicultura, 2022
Orientador: Diogo Teruo Hashimoto
Coorientador: Ricardo Utsunomia
Banca examinadora: Duílio Zerbinato, Frederico Henning, Pedro Galetti, Marcelo de Bello Cioffi
Bibliografia

1. Peixes tropicais. 2. Genoma. 3. Cromossomos sexuais. 4. Genes. I. Título. II. Jaboticabal-Centro de Aquicultura.

CDU 636.3.043

Ficha Catalográfica elaborada pela STATI - Biblioteca da UNESP
Campus de Jaboticabal/SP - Karina Gimenes Fernandes - CRB 8/7418

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Centro de Aqüicultura da Unesp - CAUNESP

CERTIFICADO DE APROVAÇÃO

TÍTULO: Recursos genômicos para *Megaleporinus macrocephalus*: subsídios para aquicultura

AUTORA: CAROLINA HELOISA DE SOUZA BORGES
ORIENTADOR: DIOGO TERUO HASHIMOTO

Aprovada como parte das exigências para obtenção do Título de Doutora em AQUICULTURA, pela Comissão Examinadora:

Prof. Dr. DIOGO TERUO HASHIMOTO (Participaçao Virtual)
Laboratorio de Genetica / Centro

Prof. Dr. FREDERICO HENNING (Participaçao Virtual)
Universidade Federal do Rio de Janeiro (UFRJ) / Departamento de Genética

Prof. Dr. Pedro Manoel Galetti Junior
Departamento de Genética e Evolução / Universidade Federal de São Carlos (UFSCAR)

Prof. Dr. MARCELO DE BELLO CIOFFI (Participaçao Virtual)
Departamento de Genética e Evolução / Universidade Federal de São Carlos (UFSCAR)

Prof. Dr. DUILIO MAZZONI ZERBINATO DE ANDRADE SILVA (Participaçao Virtual)
National Institute of Health, EUA

Jaboticabal, 29 de março de 2022

Centro de Aqüicultura da Unesp
Coordenação do Programa de Pós-Graduação
Via de Acesso Prof. Paulo Donato Castellane, s/n CEP 14884-900 Jaboticabal SP Brasil
Tel 55 16 32032110 ramal 214 fax 55 16 32032268
pgaqui@caunesp.unesp.br www.caunesp.unesp.br

Ao Billy e vó Dirce, dedico…

Felicidade foi-se embora e a saudade no meu peito ainda mora ….

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Ao meu orientador Dr. Diogo Teruo Hashimoto, em primeiro lugar, pela oportunidade do doutorado. Por ser um ótimo orientador e sempre estar muito presente. Por ter contribuído tanto com meu desenvolvimento profissional durante o doutorado.

Ao meu coorientador Dr. Ricardo Utsunomia, da Universidade Federal Rural do Rio de Janeiro (UFRRJ) por toda colaboração na realização deste trabalho.

Ao Dr. Ross Houston, Dr. Agustín Barría e Dra. Carolina Peñaloza, do Roslin Institute, da University of Edinburgh, por todo o suporte na Escócia/Brasil e pelo acompanhamento durante todo o desenvolvimento deste projeto.

À Dra. Marcela Uliano-Silva, do Sanger Institute, pelo exemplo profissional e por todo os conhecimentos compartilhados em *data analysis*.

Ao Dr. Yann Guiguen, do National Research Institute for Agriculture, Food and the Environment (INRAE), pelo aprendizado e por todo o suporte nas análises de *Pool-sex*.

Ao Dr. Ricardo Hattori, Dr. Arno Juliano Butzge, Dra. Yara e Túlio, do Intituto de Pesca de Campos do Jordão, por todo conhecimento compartilhado e ensinamentos na área de histologia e diferenciação sexual de peixes.

Ao Dr. Alessandro Varani, da Faculdade de Ciências Agrárias e Veterinárias (FCAV), por todo o aprendizado e especialmente por todo auxílio na anotação do genoma.

Ao Dr. José Manuel Yáñez, da Universidad de Chile, por todo suporte durante minha estadia no Chile e por me abrir as portas à genômica.

À dona Sueli, do projeto Peixes, por toda boa vontade com a gente e auxílio nas coletas.

Aos integrantes e ex-integrantes do Laboratório de Genética em Aquicultura e Conservação (LaGeAC): Vito, John, Shysley, Rubens, Valéria, Celma, Marcelo, Natália, Milena e Raquel, por toda colaboração na realização deste trabalho e principalmente pelas conversas, cafézinhos, rolês e cervejas compartilhadas.

Ao Valdecir e ao Perereca, técnicos do Centro de Aquicultura (CAUNESP), por toda a ajuda sempre e em tudo.

A todos os funcionários do Caunesp, em especial às tias: Lucia, Eleuza e Elaine, e ao David, da Pós-Graduação.

À minha vó Dirce, meu anjo-da-guarda, e meu maior exemplo.

Ao Billy, meu fiel-escudeiro, por toda sua lealdade, carinho e amor incondicional.

À minha mãe Patrícia e meu pai João, meus exemplos, por sempre me ensinarem a importância dos estudos, pelo amor incondicional e apoio imensurável.

À minha irmãzinha Valentina, simplesmente por existir e por deixar meus dias muito mais felizes.

Ao meu colega e namorado, Inácio Mateus Assane, por todos os bons momentos, conversas, parceria e principalmente por ser tão estranho quanto eu.

À toda minha família, especialmente meus primos-irmãos: Kaity, Marcelo e Giuli, à minha madrinha Durva, minha tia Vera e meus tios Ginho e Tedy.

Às minhas grandes comadres, Thais e Maria Carol, que mesmo há milhares de quilômetros de distância estão sempre presentes.

À todo pessoal de Jabuka, por todos os momentos especiais compartilhados durante este período.

À todos que contribuíram de alguma forma com este trabalho.

À Faculdade de Ciências Agrárias e Veterinárias (FCAV) e ao Centro de Aquicultura (CAUNESP) pelo privilégio de poder trabalhar neste lugar maravilhoso.

Sumário

Lista de Tabelas

Lista de Figuras

Resumo Geral

O piauçu, *Megaleporinus macrocephalus*, pertencente à familia Anostomidae, apresenta alto valor econômico para a aquicultura brasileira. Devido a presença de um sistema ZZ/ZW de cromossomos sexuais, ele representa a única espécie relevante para a aquicultura brasileira que pode ser manipulada cromossomicamente. Adicionalmente, por apresentarem um notável dimorfismo sexual no crescimento, com as fêmeas adultas atingindo tamanhos bem maiores que os machos, um genoma de referência de alta qualidade seria um recurso muito útil para correta identificação dos mecanismos moleculares envolvidos na determinação sexual. Assim, o presente trabalho teve como objetivo a construção de um genoma de referência valioso para estudos de genômica funcional no piauçu, além de fornecer *insights* sobre a evolução dos cromossomos sexuais e mecanismos de determinação do sexo. No presente estudo, usamos leituras longas PacBio, leituras curtas e dados de Hi-C para montar um genoma de alta qualidade, pseudo haploide e em nível cromossômico. O genoma contém 27 cromossomos, incluindo o cromossomo sexual, o que corresponde ao número de cromossomos haploide da espécie. O cromossomo sexual é composto por 24 Mb da região pseudo-autossômica de Z e W e regiões específicas de Z e W (18 Mb e 3 Mb, respectivamente). O novo genoma forneceu informações sobre características intrigantes da espécie e foi usado como referência em várias análises relacionadas à evolução do cromossomo sexual e determinação do sexo de *M. macrocephalus* e em uma espécie próxima, *Leporinus friderici*. Usamos informações de segregação de famílias para construir um mapa genético que destacou as diferenças entre os mapas físico (genoma) e o genético, e evidenciou a supressão da recombinação no cromossomo sexual. Sequências completas do genoma de machos e fêmeas revelaram diferenças sexo-específicas que permitiram a delimitação de diferentes estratos no cromossomo sexual e reflexões sobre seu processo evolutivo. Dados de expressão de machos e fêmeas no período de

determinação do sexo permitiu a identificação dos genes CCDC114 e IGFBP-6 como candidatos à determinação do sexo no piauçu.

Palavras-chave: cromossomo Z, cromossomo W, genoma, Characiformes, Anostomidae

General Abstract

In the present study, we used PacBio long reads, short reads, and Hi-C data to assemble a high-quality chromosome-level pseudo-haploid genome for *Megaleporinus macrocephalus* containing 27 chromosomes, including the sex chromosome, which corresponds to the karyotype data of the species. The sex chromosome is composed of 24 Mb of the pseudo-autosomal region of Z and W, and specific regions of Z and W (18 Mb and 3 Mb, respectively). The new genome provided insights into intriguing characteristics of the species and was used as a reference in various analyses related to the sex chromosome evolution and sex determination of *M. macrocephalus* and a closely related species, *Leporinus friderici*. We used family segregation information to construct a genetic map that highlighted differences between physical and genetic maps, and evidentiated recombination suppression in the sex chromosome. Whole-genome sequences of males and females revealed sex-biased differences that permitted the delimitation of different *strata* in the sex chromosome and insights into its evolutionary processes. Transcript expression data of males and females in the period of sex determination allowed the identification of the coiled-coil domain containing 114 (CCDC114) and the insulin-like growth factor binding protein (IGFBP-6) as sex determination candidate genes of piauçu.

Keywords: Z chromosome, W chromosome, chromosome-level genome assembly, Characiformes, Anostomidae

1. Introdução Geral

A demanda global por proteínas derivadas de pescado tem aumentado, principalmente devido ao crescimento da população e preferência por alimentos saudáveis (BACHER, 2015). A aquicultura é uma das fontes de proteínas mais sustentáveis para humanos e, nas últimas décadas, foi o setor da pecuária que mais se expandiu (YUE; WANG, 2016). Devido a estagnação e insustentabilidade da pesca predatória (FAO, 2018), a aquicultura será responsável por suprir a maior parte do aumento na demanda por pescado, de maneira correta e sustentável (BACHER, 2015).

O Brasil apresenta um grande potencial para a produção de organismos aquáticos; tem uma das maiores bacias hidrográficas do mundo, contém aproximadamente 12% de toda superfície de água doce do planeta e mais de 4 milhões de hectares de barragens artificiais e reservatórios (ROCHA et al., 2013). Além disso, o país apresenta uma extensa zona costeira de 8.700 km, temperaturas altas na maior parte ano e alta diversidade de espécies com potencial para melhoramento genético (ROCHA et al., 2013). Aproximadamente 40 espécies nativas são utilizadas para aquicultura (GODINHO; KYNARD; MARTINEZ, 2007), entretanto, a maioria dos cultivos é realizada com espécies exóticas.

Em 2020, a produção do Brasil de peixes cultivados foi de 802.903 toneladas, e desta, 60% foi referente à produção de tilápia (PEIXE BR, 2020). Devido à ausência de pacote tecnológico para as espécies nativas, seu cultivo ainda não é aderido pelos produtores, que preferem produzir espécies já melhoradas geneticamente e que dispõe de protocolos de manejo, reprodução e nutrição bem definidos (SIDONIO *et al.*, 2012). Este panorama revela a carência por projetos que disponibilizem recursos genéticos para a produção de espécies nativas do Brasil.

O sequenciamento do genoma é uma ferramenta essencial para abordar questões importantes na aquicultura, que são críticas para uma aquicultura sustentável e lucrativa (YUE;

WANG, 2016). Em muitas espécies, o sequenciamento do genoma tem sido utilizado para montagem de genoma de referência de qualidade que serve como base para aplicação de outras técnicas como seleção genômica, estudos de associação genômica ampla (GWAS) e identificação de loci de caráter quantitativo (QTL) com eficiência (YOU; SHAN; SHI, 2020). Devido ao panorama brasileiro de subaproveitamento das espécies nativas, é fundamental que se iniciem esforços para produzir recursos genômicos das nossas espécies. Junto com desenvolvimento de técnicas de reprodução mais eficientes, o melhoramento genético é um dos pilares para alavancar o desenvolvimento de pacote tecnológico das espécies nativas e assim alavancar sua produção.

2. Genomas de Peixes

Os peixes da classe Actinopterygii, são caracterizados por nadadeiras "raiadas" e apresentam mais de 23.500 espécies. Eles representam mais de 95% de todas as espécies de peixes vivas e aproximadamente metade das espécies existentes de vertebrados. Mais de 99,8% dos actinopterígios pertencem aos teleósteos (VOLFF, 2004).

A incrível diversidade dos actinopterígeos/teleósteos pode ser atribuída a um evento de duplicação do genoma (TS-WGD) que ocorreu durante a evolução inicial da linhagem (ALBERT; TAGLIACOLLO; DAGOSTA, 2020). Nesta ocasião, centenas de genes duplicados foram mantidos ao longo de centenas de milhões de anos de evolução (VOLFF, 2004). Os mecanismos evolutivos que atuaram neste cenário, tais como subfuncionalização, neofuncionalização e seleção de dosagem dos genes em duplicata, podem estar envolvidos na origem da variabilidade dos peixes teleósteos (GLASAUER; NEUHAUSS, 2014). Este fato os torna extremamente atraentes para o estudo de muitas questões evolutivas relacionadas a diversos aspectos da biologia.

O genoma do baiacu (*Takifugu rubripes)* foi o segundo genoma de vertebrado a ser sequenciado, logo após o do humano. Seu sequenciamento, por meio da estratégia de *whole-genome shotgun* (WGS), permitiu a primeira comparação do genoma completo entre duas espécies de vertebrados (APARICIO *et al.*, 2002).

Atualmente, os peixes apresentam 678 genomas disponíveis no GenBank Genomas de 678 espécies de Teleostei (17,84% dos eucariotos), e representam o segundo maior grupo de animais sequenciados, atrás apenas dos insetos (1.452 genomas). Dentre esses genomas, 24,19% são montagens em nível cromossômico e 17,26% contêm o mitogenoma. O número cromossômico haploide varia de 10, em *Neoceratodus forsteri*, à 85 em *Petromyzon marinus*, sendo mais frequente espécies com 24 cromossomos (9,44%).

Os Cichliformes (representados pelas tilápias, tucunaré), são a maior ordem entre os vertebrados e a mais diversificada entre os peixes. Provavelmente por isso, representam quase metade (46,48%) dos genomas de peixes disponíveis no GenBank, seguidos dos Cypriniformes (6,48%), grupo que inclui as carpas.

Os Characiformes são teleósteos exclusivos de ambientes de água doce e que tem representantes distribuídos por toda região Neotropical. Eles destacam-se tanto pela grande diversidade de espécies, como também pela grande diversidade morfológica e ecológica. Apesar da riqueza de espécies, são escassamente representados, tendo somente poucas espécies com genomas disponíveis; i.e., piranha *Pygocentrus nattereri* (SCHARTL et al., 2019), African pike *Hepsetus odoe* (DU *et al.*, 2020), peixe-cego *Astyanax mexicanus* (WARREN et al., 2021), e tambaqui *Colossoma macropomum* (HILSDORF et al., 2021).

Os maiores genomas encontrados na base de dados foram de peixes pulmonados; i.e., African lungfish (*Protopterus annectens*), de 40 gigabase (Gb) e Australian lungfish (*Neoceratodus forsteri*), de 34 Gb. Estas espécies são os descendentes dos peixes mais relacionados evolutivamente com os tetrápodes e, notavelmente, possuem os maiores genomas

já sequenciados até hoje (WANG *et al.*, 2021) (MEYER *et al.*, 2021). Com exceção destes genomas, que podem ser considerados discrepantes, o tamanho dos genomas disponíveis na base de dados varia de 0,03 Gb até 4, 47 Gb.

3. Megaleporinus macrocephalus

Megaleporinus macrocephalus, conhecido popularmente como piauçu, é uma espécie de peixe de água doce da família Anostomidae. Ela ocorre na Bacia do Rio Paraná e está presente na América do Sul (no Brasil, Argentina, Bolívia e Paraguai). Apresenta um comprimento de até 60 cm e tem como característica marcante a presença de manchas escuras alongadas na vertical do corpo, que podem variar de uma a quatro. É uma espécie migradora, e pode realizar grandes deslocamentos no período pré-reprodutivo (REYNALTE-TATAJE; ZANIBONI-FILHO; MUELBERT, 2008).

O piauçu apresenta alto valor econômico para a aquicultura brasileira. Sua produção pode alcançar 3,8 mil toneladas/ano (IBGE, 2014) e, dados mais recentes apontaram o crescimento de 378% nas exportações da espécie no período de 2019 a 2020 (PEIXE BR, 2021). É um animal de fácil cultivo, já que se reproduz artificialmente em cativeiro, é tolerante ao manejo, aceita ração e apresenta filé de qualidade. Também é muito apreciado na pesca esportiva (REYNALTE-TATAJE; ZANIBONI-FILHO; MUELBERT, 2008).

A espécie tem sistema de determinação sexual ZZ/ZW, no qual a fêmea é o sexo heterogamético; e apresenta cromossomos sexuais heteromórficos (e. g. citologicamente distintos)(GALETTI, JR. *et al.*, 1981). Apesar da forma mais conhecida de diferenciação sexual envolver cromossomos sexuais estruturalmente diferenciados (CHARLESWORTH, 2021), esta morfologia só foi detectada em 6% das espécies de peixe de água doce estudadas (OLIVEIRA; FORESTI; HILSDORF, 2009). Por isso, além de sua relevância na aquicultura,

M. macrocephalus representa um modelo único para estudos da evolução de cromossomos sexuais polimórficos em peixes, e em vertebrados.

4. Sequenciamento de Terceira Geração

O sequenciamento de leituras longas, também chamado de sequenciamento de terceira geração, teve seu surgimento na década passada, e representou uma revolução na genômica (POLLARD *et al.*, 2018). Apesar das limitações verificadas no surgimento da tecnologia, como, o baixo rendimento do sequenciamento e alta porcentagem de erro das *reads*; o desenvolvimento de sequenciadores e químicas mais eficientes conseguiram superar estas limitações e aumentar seu custo-benefício (LOGSDON; VOLLGER; EICHLER, 2020). Atualmente, o sequenciamento de *long-reads* tem sido amplamente utilizado, e possibilitado a montagem de genomas de alta qualidade para espécies modelos e não-modelos. As duas plataformas que dominam este mercado são a Pacific Biosciences (PacBio); com a tecnologia "*single molecule real-time*" (SMRT) e a Oxford Nanopore Technologies (ONT); com a tecnologia *Nanopore*.

A vantagem deste tipo de sequenciamento sobre o sequenciamento de segunda geração é que enquanto plataformas como Illumina, MGISEQ, BGISEQ, entre outras, produzem *reads* de até 600 pares de bases; as *long-reads* geralmente apresentam comprimento maior que 10 kb (POLLARD *et al.*, 2018). Com isso, as leituras longas permitem melhorar a qualidade da montagem do genoma, porque conseguem melhor resolução em regiões complexas, como as regiões repetitivas; que são dificilmente capturadas com as *short-reads* (POLLARD *et al.*, 2018).

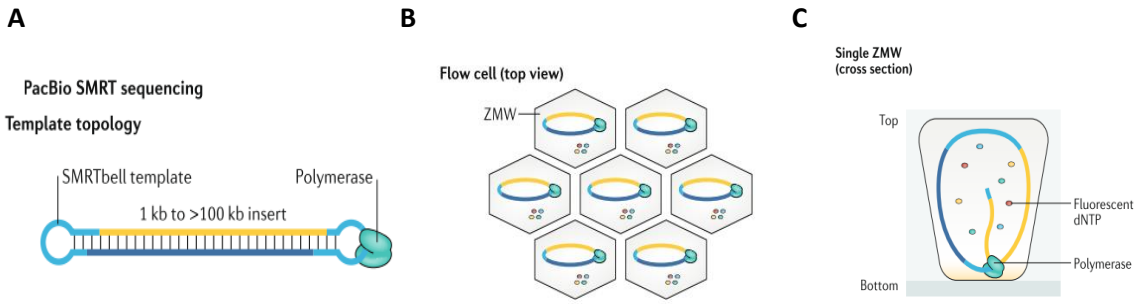


Figure 1. PacBio SMRT sequencing. SMRTbell (A). SMRT Cell (B). Zero Mode Waveguide (ZMW) (C). Adaptado de Logsdon et al. (2020).

No sequenciamento PacBio SMRT, o DNA de alto peso molecular é extraído de amostras de tecido e fragmentado. Em seguida, o DNA dupla-fita é ligado a adaptadores do tipo *hairpin* para formar uma molécula circular, conhecida como SMRTbell (Figura 1A). Uma vez que a SMRTbell é gerada, ela é ligada por uma DNA polimerase e carregada em uma SMRT Cell para o sequenciamento. A SMRT Cell é composta por milhões de câmaras, conhecidas como *Zero Mode Wave Guides* (ZMWs), que é onde a reação de sequenciamento ocorre (Figura 1B). No sistema Sequel II, sequenciador mais recente da Pacific Biosciences, cada SMRT Cell contém cerca de 8 milhões de ZMWs (LOGSDON; VOLLGER; EICHLER, 2020).

Após o carregamento do DNA na SMRT Cell, a SMRTbell adentra na ZMW e a polimerase fica imobilizada em seu fundo para o início da reação (Figura 1C). Idealmente, somente uma molécula deve entrar em cada câmara, porém, pode ocorrer de mais de uma SMRTbell adentrar na mesma câmara, o que irá gerar erros no sequenciamento (EID et al., 2009).

Em seguida, nucleotídeos marcados com fluorescência (dNTPs) são adicionados e a reação se inicia. No momento da incorporação do dNTP pela polimerase, um feixe de luz é emitido pelo fundo da ZMW, que excita o nucleotídeo, e este emite fluorescência. Uma câmera em cada ZMW detecta a posição relativa de cada nucleotídeo incorporado e o comprimento de onda emitido. Como as ZMWs emitem um pequeno feixe de luz de 20 – 30 nm, a área de detecção do sinal de fluorescência é restringida à área de incorporação do nucleotídeo. Isto

evita a detecção de ruído gerado pelos dNTPs ainda não incorporados. A reação só termina quando a polimerase perde sua eficiência (EID et al., 2009).

O sequenciamento PacBio pode ser feito de dois modos: *Continuous Long Read* (CLR) *Sequencing* ou *Circular Consensus Sequencing* (CCS), também conhecido como *Hi-Fi*. O sequenciamento CLR é usado para gerar as *reads* mais longas possíveis. Para isso, o *input* de DNA para gerar as bibliotecas é maior (>30 kb) e durante a reação, a DNA polimerase dá uma ou poucas passadas pela SMRTbell. Já no sequenciamento CCS, o tamanho do *input* de DNA é menor (10 - 30 kb) e por isso, a polimerase dá diversas passadas no *template*. Cada *Hi-Fi read* é gerada pelo consenso das *subreads* geradas por cada ZMW. As *Hi-Fi reads* tem comprimento menor que as CLR *reads*, porém apresentam alta acurácia (~ 99,9%); comparável às leituras de sequenciadores Illumina (LOGSDON; VOLLGER; EICHLER, 2020).

Normalmente, algumas estratégias são empregadas para diminuir a taxa de erro das CLR *reads*. Alguns programas de montagem de genomas, possuem uma etapa de correção das CLR *reads* anterior à montagem. Para isso, as *reads* mais longas ou com maior cobertura (chamadas de *seed-reads*) são selecionadas, e em seguida as *reads* não selecionadas são alinhadas à estas, visando gerar um consenso. Após a montagem, os *contigs* também são "polidos" pelas PacBio *reads*. Isto é, as *reads* são mapeadas aos *contigs* e é gerado um consenso. Da mesma forma, os *contigs* também podem ser polidos por *short-reads*, que apresentam alta acurácia e consequentemente, elevam significativamente a qualidade das bases do genoma.

5. Captura de Conformação da Cromatina (Hi-C)

A cromatina é o complexo de DNA e proteínas que se encontra dentro do núcleo celular das células eucarióticas. Ela se apresenta empacotada em estruturas tridimensionais, que mantêm uma relação direta entre distância física e genômica (BELTON *et al.*, 2012). Por exemplo, dois loci que estão próximos no cromossomo, também estão próximos no espaço

físico tridimensional (BURTON *et al.*, 2009). A técnica de Hi-C, utiliza esta relação entre ligação e proximidade para ordenar e orientar os *contigs* de um genoma em cromossomos (LOGSDON; VOLLGER; EICHLER, 2020b).

No Hi-C, a cromatina do núcleo é transpassada com formaldeído, que se liga covalentemente aos complexos proteína-DNA próximos um do outro. A cromatina transpassada é digerida com uma enzima de restrição ou nuclease, e as pontas do DNA de fita simples são preenchidas e reparadas com nucleotídeos marcados com biotina (LOGSDON; VOLLGER; EICHLER, 2020). Este processo cria junções entre sequências diferentes que estão localizadas próximas no espaço nuclear (i. e. junções quiméricas), e é chamado de ligação por proximidade. Quanto mais vezes a junção entre duas determinadas sequências for observada, mais perto elas estão no espaço genômico. As sequências marcadas com biotina são purificadas e submetidas a sequenciamento *paired-end* (LOGSDON; VOLLGER; EICHLER, 2020). Como resultado, as *reads* Hi-C são mapeadas a um genoma e usadas para agrupar *contigs* em cromossomos, detectar variantes estruturais e erros na montagem.

6. Objetivo Geral

O objetivo deste trabalho foi produzir recursos genômicos para *Megaleporinus macrocephalus*, por meio construção de um genoma de referência para espécie, que irá gerar subsídios para a aquicultura.

7. Referências

ALBERT, J. S.; TAGLIACOLLO, V. A.; DAGOSTA, F. Diversification of Neotropical Freshwater Fishes. 2020.

APARICIO, S.; CHAPMAN, J.; STUPKA, E.; PUTNAM, N.; CHIA, J. ming; DEHAL, P.; CHRISTOFFELS, A.; RASH, S.; HOON, S.; SMIT, A.; SOLLEWIJN GELPKE, M. D.; ROACH, J.; OH, T.; HO, I. Y.; WONG, M.; DETTER, C.; VERHOEF, F.; PREDKI, P.; TAY, A.; LUCAS, S.; RICHARDSON, P.; SMITH, S. F.; CLARK, M. S.; EDWARDS, Y. J. K.; DOGGETT, N.; ZHARKIKH, A.; TAVTIGIAN, S. V.; PRUSS, D.; BARNSTEAD, M.; EVANS, C.; BADEN, H.; POWELL, J.; GLUSMAN, G.; ROWEN, L.; HOOD, L.; TAN, Y.

H.; ELGAR, G.; HAWKINS, T.; VENKATESH, B.; ROKHSAR, D.; BRENNER, S. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. **Science**, v. 297, n. 5585, p. 1301–1310, 23 ago. 2002.

BACHER, K. Perceptions and misconceptions of aquaculture: a global overview. **GLOBEFISH Research Programme**, v. 120, 2015.

BATEMAN, A.; MARTIN, M. J.; ORCHARD, S.; MAGRANE, M.; AGIVETOVA, R.; AHMAD, S.; ALPI, E.; BOWLER-BARNETT, E. H.; BRITTO, R.; BURSTEINAS, B.; BYE-A-JEE, H.; COETZEE, R.; CUKURA, A.; SILVA, A. da; DENNY, P.; DOGAN, T.; EBENEZER, T. G.; FAN, J.; CASTRO, L. G.; GARMIRI, P.; GEORGHIOU, G.; GONZALES, L.; HATTON-ELLIS, E.; HUSSEIN, A.; IGNATCHENKO, A.; INSANA, G.; ISHTIAQ, R.; JOKINEN, P.; JOSHI, V.; JYOTHI, D.; LOCK, A.; LOPEZ, R.; LUCIANI, A.; LUO, J.; LUSSI, Y.; MACDOUGALL, A.; MADEIRA, F.; MAHMOUDY, M.; MENCHI, M.; MISHRA, A.; MOULANG, K.; NIGHTINGALE, A.; OLIVEIRA, C. S.; PUNDIR, S.; QI, G.; RAJ, S.; RICE, D.; LOPEZ, M. R.; SAIDI, R.; SAMPSON, J.; SAWFORD, T.; SPERETTA, E.; TURNER, E.; TYAGI, N.; VASUDEV, P.; VOLYNKIN, V.; WARNER, K.; WATKINS, X.; ZARU, R.; ZELLNER, H.; BRIDGE, A.; POUX, S.; REDASCHI, N.; AIMO, L.; ARGOUD-PUY, G.; AUCHINCLOSS, A.; AXELSEN, K.; BANSAL, P.; BARATIN, D.; BLATTER, M. C.; BOLLEMAN, J.; BOUTET, E.; BREUZA, L.; CASALS-CASAS, C.; DE CASTRO, E.; ECHIOUKH, K. C.; COUDERT, E.; CUCHE, B.; DOCHE, M.; DORNEVIL, D.; ESTREICHER, A.; FAMIGLIETTI, M. L.; FEUERMANN, M.; GASTEIGER, E.; GEHANT, S.; GERRITSEN, V.; GOS, A.; GRUAZ-GUMOWSKI, N.; HINZ, U.; HULO, C.; HYKA-NOUSPIKEL, N.; JUNGO, F.; KELLER, G.; KERHORNOU, A.; LARA, V.; LE MERCIER, P.; LIEBERHERR, D.; LOMBARDOT, T.; MARTIN, X.; MASSON, P.; MORGAT, A.; NETO, T. B.; PAESANO, S.; PEDRUZZI, I.; PILBOUT, S.; POURCEL, L.; POZZATO, M.; PRUESS, M.; RIVOIRE, C.; SIGRIST, C.; SONESSON, K.; STUTZ, A.; SUNDARAM, S.; TOGNOLLI, M.; VERBREGUE, L.; WU, C. H.; ARIGHI, C. N.; ARMINSKI, L.; CHEN, C.; CHEN, Y.; GARAVELLI, J. S.; HUANG, H.; LAIHO, K.; MCGARVEY, P.; NATALE, D. A.; ROSS, K.; VINAYAKA, C. R.; WANG, Q.; WANG, Y.; YEH, L. S.; ZHANG, J. UniProt: the universal protein knowledgebase in 2021. **Nucleic Acids Research**, v. 49, n. D1, p. D480–D489, 8 jan. 2021a. Disponível em: <https://academic.oup.com/nar/article/49/D1/D480/6006196>. Acesso em: 14 fev. 2022.

BATEMAN, A.; MARTIN, M. J.; ORCHARD, S.; MAGRANE, M.; AGIVETOVA, R.; AHMAD, S.; ALPI, E.; BOWLER-BARNETT, E. H.; BRITTO, R.; BURSTEINAS, B.; BYE-A-JEE, H.; COETZEE, R.; CUKURA, A.; SILVA, A. Da; DENNY, P.; DOGAN, T.; EBENEZER, T. G.; FAN, J.; CASTRO, L. G.; GARMIRI, P.; GEORGHIOU, G.; GONZALES, L.; HATTON-ELLIS, E.; HUSSEIN, A.; IGNATCHENKO, A.; INSANA, G.; ISHTIAQ, R.; JOKINEN, P.; JOSHI, V.; JYOTHI, D.; LOCK, A.; LOPEZ, R.; LUCIANI, A.; LUO, J.; LUSSI, Y.; MACDOUGALL, A.; MADEIRA, F.; MAHMOUDY, M.; MENCHI, M.; MISHRA, A.; MOULANG, K.; NIGHTINGALE, A.; OLIVEIRA, C. S.; PUNDIR, S.; QI, G.; RAJ, S.; RICE, D.; LOPEZ, M. R.; SAIDI, R.; SAMPSON, J.; SAWFORD, T.; SPERETTA, E.; TURNER, E.; TYAGI, N.; VASUDEV, P.; VOLYNKIN, V.; WARNER, K.; WATKINS, X.; ZARU, R.; ZELLNER, H.; BRIDGE, A.; POUX, S.; REDASCHI, N.; AIMO, L.; ARGOUD-PUY, G.; AUCHINCLOSS, A.; AXELSEN, K.; BANSAL, P.; BARATIN, D.; BLATTER, M. C.; BOLLEMAN, J.; BOUTET, E.; BREUZA, L.; CASALS-CASAS, C.; DE CASTRO, E.; ECHIOUKH, K. C.; COUDERT, E.; CUCHE, B.; DOCHE, M.; DORNEVIL,

D.; ESTREICHER, A.; FAMIGLIETTI, M. L.; FEUERMANN, M.; GASTEIGER, E.; GEHANT, S.; GERRITSEN, V.; GOS, A.; GRUAZ-GUMOWSKI, N.; HINZ, U.; HULO, C.; HYKA-NOUSPIKEL, N.; JUNGO, F.; KELLER, G.; KERHORNOU, A.; LARA, V.; LE MERCIER, P.; LIEBERHERR, D.; LOMBARDOT, T.; MARTIN, X.; MASSON, P.; MORGAT, A.; NETO, T. B.; PAESANO, S.; PEDRUZZI, I.; PILBOUT, S.; POURCEL, L.; POZZATO, M.; PRUESS, M.; RIVOIRE, C.; SIGRIST, C.; SONESSON, K.; STUTZ, A.; SUNDARAM, S.; TOGNOLLI, M.; VERBREGUE, L.; WU, C. H.; ARIGHI, C. N.; ARMINSKI, L.; CHEN, C.; CHEN, Y.; GARAVELLI, J. S.; HUANG, H.; LAIHO, K.; MCGARVEY, P.; NATALE, D. A.; ROSS, K.; VINAYAKA, C. R.; WANG, Q.; WANG, Y.; YEH, L. S.; ZHANG, J. UniProt: the universal protein knowledgebase in 2021. **Nucleic Acids Research**, v. 49, n. D1, p. D480–D489, 8 jan. 2021b.

BELLOTT, D. W.; SKALETSKY, H.; CHO, T.-J.; BROWN, L.; LOCKE, D.; CHEN, N.; GALKINA, S.; PYNTIKOVA, T.; KOUTSEVA, N.; GRAVES, T.; KREMITZKI, C.; WARREN, W. C.; CLARK, A. G.; GAGINSKAYA, E.; WILSON, R. K.; DAVID, &. Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. **Nature Publishing Group**, v. 49, 2017.

BELTON, J.-M.; MCCORD, R. P.; GIBCUS, J. H.; NAUMOVA, N.; ZHAN, Y.; DEKKER, J. Hi-C: A comprehensive technique to capture the conformation of genomes. 2012.

BERTOLLO, L. A. C.; MOREIRA–FILHO, O.; GALETTI, P. M. Cytogenetics and taxonomy: considerations based on chromosome studies of freshwater fish. **Journal of Fish Biology**, v. 28, n. 2, p. 153–159, 1986.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 1 ago. 2014.

BOUZA, C.; HERMIDA, M.; PARDO, B. G.; FERNÁNDEZ, C.; FORTES, G. G.; CASTRO, J.; SÁNCHEZ, L.; PRESA, P.; PÉREZ, M.; SANJUÁN, A.; DE CARLOS, A.; ÁLVAREZ-DIOS, J. A.; EZCURRA, S.; CAL, R. M.; PIFERRER, F.; MARTÍNEZ, P. A microsatellite genetic map of the turbot (Scophthalmus maximus). **Genetics**, v. 177, n. 4, 2007.

BRAY, N. L.; PIMENTEL, H.; MELSTED, P.; PACHTER, L. Near-optimal probabilistic RNA-seq quantification. **Nature Biotechnology 2016 34:5**, v. 34, n. 5, p. 525–527, 4 abr. 2016.

BRŮNA, T.; HOFF, K. J.; LOMSADZE, A.; STANKE, M.; BORODOVSKY, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. **NAR Genomics and Bioinformatics**, v. 3, n. 1, p. 1–11, 6 jan. 2021a. Disponível em: <https://academic.oup.com/nargab/article/3/1/lqaa108/6066535>. Acesso em: 14 fev. 2022.

BRŮNA, T.; HOFF, K. J.; LOMSADZE, A.; STANKE, M.; BORODOVSKY, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. **NAR Genomics and Bioinformatics**, v. 3, n. 1, p. 1–11, 6 jan. 2021b.

BURTON, D. R.; WALKER, L. M.; PHOGAT, S. K.; CHAN-HUI, P. Y.; WAGNER, D.; PHUNG, P.; GOSS, J. L.; WRIN, T.; SIMEK, M. D.; FLING, S.; MITCHAM, J. L.; LEHRMAN, J. K.; PRIDDY, F. H.; OLSEN, O. A.; FREY, S. M.; HAMMOND, P. W.; PROTOCOL, G.; KAMINSKY, S.; ZAMB, T.; MOYLE, M.; KOFF, W. C.; POIGNARD, P.

Broad and potent neutralizing antibodies from an african donor reveal a new HIV-1 vaccine target. **Science**, v. 326, n. 5950, p. 285–289, 2009. Disponível em: <www.sciencemag.org/cgi/content/full/1178746/DC1>.

CAPELLA-GUTIÉRREZ, S.; SILLA-MARTÍNEZ, J. M.; GABALDÓN, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. **BIOINFORMATICS APPLICATIONS NOTE**, v. 25, n. 15, p. 1972–1973, 2009.

CASTAÑO-SÁNCHEZ, C.; FUJI, K.; OZAKI, A.; HASEGAWA, O.; SAKAMOTO, T.; MORISHIMA, K.; NAKAYAMA, I.; FUJIWARA, A.; MASAOKA, T.; OKAMOTO, H.; HAYASHIDA, K.; TAGAMI, M.; KAWAI, J.; HAYASHIZAKI, Y.; OKAMOTO, N. A second generation genetic linkage map of Japanese flounder (Paralichthys olivaceus). **BMC Genomics**, v. 11, n. 1, 2010.

CATCHEN, J.; AMORES, A.; BASSHAM, S. Chromonomer: A tool set for repairing and enhancing assembled genomes through integration of genetic maps and conserved synteny. **G3: Genes, Genomes, Genetics**, v. 10, n. 11, 2020.

CATCHEN, J.; HOHENLOHE, P. A.; BASSHAM, S.; AMORES, A.; CRESKO, W. A. Stacks: An analysis tool set for population genomics. **Molecular Ecology**, v. 22, n. 11, p. 3124–3140, 2013.

CATCHEN, J. M.; AMORES, A.; HOHENLOHE, P.; CRESKO, W.; POSTLETHWAIT, J. H. Stacks: Building and genotyping loci de novo from short-read sequences. **G3: Genes, Genomes, Genetics**, v. 1, n. 3, p. 171–182, 2011.

CHALOPIN, D.; NAVILLE, M.; PLARD, F.; GALIANA, D.; VOLFF, J.-N. Comparative Analysis of Transposable Elements Highlights Mobilome Diversity and Evolution in Vertebrates. [s.d.]

CHARLESWORTH, B. **The Evoluton of Sex Chromosomes Downloaded from**. [s.l: s.n.].

CHARLESWORTH, B. **The Evoluton of Sex Chromosomes Downloaded from**. [s.l: s.n.].

CHARLESWORTH, D. **Young sex chromosomes in plants and animalsNew Phytologist**2019a.

CHARLESWORTH, D. **Young sex chromosomes in plants and animalsNew Phytologist**2019b.

CHARLESWORTH, D. The timing of genetic degeneration of sex chromosomes. 2021c.

CHARLESWORTH, D. The timing of genetic degeneration of sex chromosomes. 2021d.

CHARLESWORTH, D.; CHARLESWORTH, B.; MARAIS, G. Steps in the evolution of heteromorphic sex chromosomes. **Heredity 2005 95:2**, v. 95, n. 2, p. 118–128, 25 maio 2005.

CHEN, S.; ZHANG, G.; SHAO, C.; HUANG, Q.; LIU, G.; ZHANG, P.; SONG, W.; AN, N.; CHALOPIN, D.; VOLFF, J. N.; HONG, Y.; LI, Q.; SHA, Z.; ZHOU, H.; XIE, M.; YU, Q.; LIU, Y.; XIANG, H.; WANG, N.; WU, K.; YANG, C.; ZHOU, Q.; LIAO, X.; YANG, L.; HU, Q.; ZHANG, J.; MENG, L.; JIN, L.; TIAN, Y.; LIAN, J.; YANG, J.; MIAO, G.; LIU, S.; LIANG, Z.; YAN, F.; LI, Y.; SUN, B.; ZHANG, H.; ZHANG, J.; ZHU, Y.; DU, M.; ZHAO, Y.; SCHARTL, M.; TANG, Q.; WANG, J. Whole-genome sequence of a flatfish provides

insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. **Nature Genetics**, v. 46, n. 3, p. 253–260, 2014.

CHIN, C. S.; PELUSO, P.; SEDLAZECK, F. J.; NATTESTAD, M.; CONCEPCION, G. T.; CLUM, A.; DUNN, C.; O'MALLEY, R.; FIGUEROA-BALDERAS, R.; MORALES-CRUZ, A.; CRAMER, G. R.; DELLEDONNE, M.; LUO, C.; ECKER, J. R.; CANTU, D.; RANK, D. R.; SCHATZ, M. C. Phased diploid genome assembly with single-molecule real-time sequencing. **Nature Methods**, v. 13, n. 12, p. 1050–1054, 2016.

CONTE, M. A.; CLARK, F. E.; ROBERTS, R. B.; XU, L.; TAO, W.; ZHOU, Q.; WANG, D.; KOCHER, T. D. Origin of a Giant Sex Chromosome. **Molecular biology and evolution**, v. 38, n. 4, p. 1554–1569, 2021.

CORNELIO, D.; CASTRO, J. P.; SANTOS, M. H.; VICARI, M. R.; DE ALMEIDA, M. C.; MOREIRA-FILHO, O.; CAMACHO, J. P. M.; ARTONI, R. F. Hermaphroditism can compensate for the sex ratio in the Astyanax scabripinnis species complex (Teleostei: Characidae): expanding the B chromosome study model. **Reviews in Fish Biology and Fisheries**, v. 27, n. 3, p. 681–689, 1 set. 2017. Disponível em: <https://link.springer.com/article/10.1007/s11160-017-9488-8>. Acesso em: 8 mar. 2022.

DE LA SERRANA, D. G.; MACQUEEN, D. J. **Insulin-like growth factor-binding proteins of teleost fishesFrontiers in Endocrinology**2018a.

DE LA SERRANA, D. G.; MACQUEEN, D. J. **Insulin-like growth factor-binding proteins of teleost fishesFrontiers in Endocrinology**2018b.

DE LOS RÍOS-PÉREZ, L.; NGUINKAL, J. A.; VERLEIH, M.; REBL, A.; BRUNNER, R. M.; KLOSA, J.; SCHÄFER, N.; STÜEKEN, M.; GOLDAMMER, T.; WITTENBURG, D. An ultra-high density SNP-based linkage map for enhancing the pikeperch (Sander lucioperca) genome assembly to chromosome-scale. **Scientific Reports**, v. 10, n. 1, 2020.

DOMINGUES, R. R.; MASTROCHIRICO-FILHO, V. A.; MENDES, N. J.; HASHIMOTO, D. T.; COELHO, R.; DA CRUZ, V. P.; ANTUNES, A.; FORESTI, F.; MENDONÇA, F. F. Comparative eye and liver differentially expressed genes reveal monochromatic vision and cancer resistance in the shortfin mako shark (Isurus oxyrinchus). **Genomics**, v. 112, n. 6, 2020.

DU, X.; HONG, X.; FAN, G.; HUANG, X.; SUN, S.; BINGJIE, O.; ZHANG, H.; ZHANG, M.; LIU, S.; LIU, X.; ZHANG, W. Chromosome-level genome assembly of the African pike, Hepsetus odoe. **bioRxiv**, p. 2020.05.13.094987, 14 maio 2020.

DUDCHENKO, O.; BATRA, S. S.; OMER, A. D.; NYQUIST, S. K.; HOEGER, M.; DURAND, N. C.; SHAMIM, M. S.; MACHOL, I.; LANDER, E. S.; AIDEN, A. P.; AIDEN, E. L. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. **Science**, v. 356, n. 6333, p. 92–95, 2017.

DURAND, N. C.; SHAMIM, M. S.; MACHOL, I.; RAO, S. S. P.; HUNTLEY, M. H.; LANDER, E. S.; AIDEN, E. L. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. **Cell Systems**, v. 3, n. 1, 2016.

EID, J.; FEHR, A.; GRAY, J.; LUONG, K.; LYLE, J.; OTTO, G.; PELUSO, P.; RANK, D.; BAYBAYAN, P.; BETTMAN, B.; BIBILLO, A.; BJORNSON, K.; CHAUDHURI, B.; CHRISTIANS, F.; CICERO, R.; CLARK, S.; DALAL, R.; DEWINTER, A.; DIXON, J.;

FOQUET, M.; GAERTNER, A.; HARDENBOL, P.; HEINER, C.; HESTER, K.; HOLDEN, D.; KEARNS, G.; KONG, X.; KUSE, R.; LACROIX, Y.; LIN, S.; LUNDQUIST, P.; MA, C.; MARKS, P.; MAXHAM, M.; MURPHY, D.; PARK, I.; PHAM, T.; PHILLIPS, M.; ROY, J.; SEBRA, R.; SHEN, G.; SORENSON, J.; TOMANEY, A.; TRAVERS, K.; TRULSON, M.; VIECELI, J.; WEGENER, J.; WU, D.; YANG, A.; ZACCARIN, D.; ZHAO, P.; ZHONG, F.; KORLACH, J.; TURNER, S. Real-Time DNA Sequencing from Single Polymerase Molecules. **Science**, v. 323, n. 5910, p. 133–138, 2 jan. 2009a.

EID, J.; FEHR, A.; GRAY, J.; LUONG, K.; LYLE, J.; OTTO, G.; PELUSO, P.; RANK, D.; BAYBAYAN, P.; BETTMAN, B.; BIBILLO, A.; BJORNSON, K.; CHAUDHURI, B.; CHRISTIANS, F.; CICERO, R.; CLARK, S.; DALAL, R.; DEWINTER, A.; DIXON, J.; FOQUET, M.; GAERTNER, A.; HARDENBOL, P.; HEINER, C.; HESTER, K.; HOLDEN, D.; KEARNS, G.; KONG, X.; KUSE, R.; LACROIX, Y.; LIN, S.; LUNDQUIST, P.; MA, C.; MARKS, P.; MAXHAM, M.; MURPHY, D.; PARK, I.; PHAM, T.; PHILLIPS, M.; ROY, J.; SEBRA, R.; SHEN, G.; SORENSON, J.; TOMANEY, A.; TRAVERS, K.; TRULSON, M.; VIECELI, J.; WEGENER, J.; WU, D.; YANG, A.; ZACCARIN, D.; ZHAO, P.; ZHONG, F.; KORLACH, J.; TURNER, S. Real-Time DNA Sequencing from Single Polymerase Molecules. **Science**, v. 323, n. 5910, p. 133–138, 2 jan. 2009b.

EMMS, D. M.; KELLY, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. **Genome Biology**, v. 20, n. 1, p. 1–14, 14 nov. 2019.

FAO. **The state of world fisheries and aquaculture 2018- Meeting the sustainable development goals. Rome**. [s.l: s.n.]v. 35

FLYNN, J. M.; HUBLEY, R.; GOUBERT, C.; ROSEN, J.; CLARK, A. G.; FESCHOTTE, C.; SMIT, A. F. RepeatModeler2 for automated genomic discovery of transposable element families. **PNAS**, v. 117, n. 17, p. 9451–9457, 2020.

FORESTI, F.; TOLEDO, L. F. A.; TOLEDO, S. A. Polymorphic nature of nucleolus organizer regions in fishes. **Cytogenetic and Genome Research**, v. 31, n. 3, 1981.

GABRIEL, L.; HOFF, K. J.; BRŮNA, T.; BORODOVSKY, M.; STANKE, M. TSEBRA: transcript selector for BRAKER. **BMC Bioinformatics**, v. 22, n. 1, p. 1–12, 1 dez. 2021a.

GABRIEL, L.; HOFF, K. J.; BRŮNA, T.; BORODOVSKY, M.; STANKE, M. TSEBRA: transcript selector for BRAKER. **BMC Bioinformatics**, v. 22, n. 1, p. 1–12, 1 dez. 2021b.

GALETTI, JR., P. M.; FORESTI, F.; BERTQLLO, L. A. C.; FILHO, M. Heteromorphic sex chromosomes in three species of the genus Leporinus (Pisces, Anostomidae). **Cytogenetic and Genome Research**, v. 29, n. 3, p. 138–142, 7 maio 1981. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/7226896/>. Acesso em: 18 ago. 2021.

GAO, B.; SHEN, D.; XUE, S.; CHEN, C.; CUI, H.; SONG, C. The contribution of transposable elements to size variations between four teleost genomes. **Mobile DNA**, v. 7, n. 1, 9 fev. 2016.

GLASAUER, S. M. K.; NEUHAUSS, S. C. F. **Whole-genome duplication in teleost fishes and its evolutionary consequencesMolecular Genetics and Genomics**Mol Genet Genomics, 19 nov. 2014.

GODINHO, A. L.; KYNARD, B.; MARTINEZ, C. B. Supplemental water releases for fisheries restoration in a Brazilian floodplain River: A conceptual model. **River Research and Applications**, v. 23, n. 9, 2007.

GÖTZ, S.; GARCÍA-GÓMEZ, J. M.; TEROL, J.; WILLIAMS, T. D.; NAGARAJ, S. H.; NUEDA, M. J.; ROBLES, M.; TALÓN, M.; DOPAZO, J.; CONESA, A. High-throughput functional annotation and data mining with the Blast2GO suite. **Nucleic Acids Research**, v. 36, n. 10, p. 3420–3435, 1 jun. 2008. Disponível em: <https://academic.oup.com/nar/article/36/10/3420/2410320>. Acesso em: 14 fev. 2022.

GUERRERO-CÓZAR, I.; PEREZ-GARCIA, C.; BENZEKRI, H.; SÁNCHEZ, J. J.; SEOANE, P.; CRUZ, F.; GUT, M.; ZAMORANO, M. J.; CLAROS, M. G.; MANCHADO, M. Development of whole-genome multiplex assays and construction of an integrated genetic map using SSR markers in Senegalese sole. **Scientific Reports**, v. 10, n. 1, 2020.

HILSDORF, A. W. S.; ULIANO-SILVA, M.; COUTINHO, L. L.; MONTENEGRO, H.; ALMEIDA-VAL, V. M. F.; PINHAL, D. Genome assembly and annotation of the tambaqui (Colossoma macropomum): an emblematic fish of the Amazon River basin. **bioRxiv**, p. 2021.09.08.459456, 9 set. 2021a.

HILSDORF, A. W. S.; ULIANO-SILVA, M.; COUTINHO, L. L.; MONTENEGRO, H.; ALMEIDA-VAL, V. M. F.; PINHAL, D. Genome assembly and annotation of the tambaqui (Colossoma macropomum): an emblematic fish of the Amazon River basin. **bioRxiv**, p. 2021.09.08.459456, 9 set. 2021b.

HOFF, K. J.; LANGE, S.; LOMSADZE, A.; BORODOVSKY, M.; STANKE, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. **Bioinformatics**, v. 32, n. 5, p. 767–769, 1 mar. 2016a. Disponível em: <https://academic.oup.com/bioinformatics/article/32/5/767/1744611>. Acesso em: 3 out. 2021.

HOFF, K. J.; LANGE, S.; LOMSADZE, A.; BORODOVSKY, M.; STANKE, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. **Bioinformatics**, v. 32, n. 5, p. 767–769, 1 mar. 2016b.

HOFF, K. J.; LOMSADZE, A.; BORODOVSKY, M.; STANKE, M. Whole-Genome Annotation with BRAKER. **Methods in Molecular Biology**, v. 1962, p. 65–95, 2019. Disponível em: <https://link.springer.com/protocol/10.1007/978-1-4939-9173-0_5>. Acesso em: 14 fev. 2022.

HOWE, K. L.; ACHUTHAN, P.; ALLEN, J.; ALLEN, J.; ALVAREZ-JARRETA, J.; RIDWAN AMODE, M.; ARMEAN, I. M.; AZOV, A. G.; BENNETT, R.; BHAI, J.; BILLIS, K.; BODDU, S.; CHARKHCHI, M.; CUMMINS, C.; DA RIN FIORETTO, L.; DAVIDSON, C.; DODIYA, K.; EL HOUDAIGUI, B.; FATIMA, R.; GALL, A.; GIRON, C. G.; GREGO, T.; GUIJARRO-CLARKE, C.; HAGGERTY, L.; HEMROM, A.; HOURLIER, T.; IZUOGU, O. G.; JUETTEMANN, T.; KAIKALA, V.; KAY, M.; LAVIDAS, I.; LE, T.; LEMOS, D.; MARTINEZ, J. G.; MARUGÁN, J. C.; MAUREL, T.; MCMAHON, A. C.; MOHANAN, S.; MOORE, B.; MUFFATO, M.; OHEH, D. N.; PARASCHAS, D.; PARKER, A.; PARTON, A.; PROSOVETSKAIA, I.; SAKTHIVEL, M. P.; ABDUL SALAM, A. I.; SCHMITT, B. M.; SCHUILENBURG, H.; SHEPPARD, D.; STEED, E.; SZPAK, M.; SZUBA, M.; TAYLOR,

K.; THORMANN, A.; THREADGOLD, G.; WALTS, B.; WINTERBOTTOM, A.; CHAKIACHVILI, M.; CHAUBAL, A.; DE SILVA, N.; FLINT, B.; FRANKISH, A.; HUNT, S. E.; IISLEY, G. R.; LANGRIDGE, N.; LOVELAND, J. E.; MARTIN, F. J.; MUDGE, J. M.; MORALES, J.; PERRY, E.; RUFFIER, M.; TATE, J.; THYBERT, D.; TREVANION, S. J.; CUNNINGHAM, F.; YATES, A. D.; ZERBINO, D. R.; FLICEK, P. Ensembl 2021. **Nucleic Acids Research**, v. 49, n. D1, p. D884–D891, 8 jan. 2021.

HUGHES, L. C.; ORTÍ, G.; HUANG, Y.; SUN, Y.; BALDWIN, C. C.; THOMPSON, A. W.; ARCILA, D.; BETANCUR, R.; LI, C.; BECKER, L.; BELLORA, N.; ZHAO, X.; LI, X.; WANG, M.; FANG, C.; XIE, B.; ZHOUI, Z.; HUANG, H.; CHEN, S.; VENKATESH, B.; SHI, Q. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. **Proceedings of the National Academy of Sciences of the United States of America**, v. 115, n. 24, p. 6249–6254, 12 jun. 2018.

IMARAZENE, B.; DU, K.; BEILLE, S.; JOUANNO, E.; FERON, R.; PAN, Q.; TORRES-PAZ, J.; LOPEZ-ROQUES, C.; CASTINEL, A.; GIL, L.; KUCHLY, C.; DONNADIEU, C.; PARRINELLO, H.; JOURNOT, L.; CABAU, C.; ZAHM, M.; KLOPP, C.; PAVLICA, T.; AL-RIKABI, A.; LIEHR, T.; SIMANOVSKY, S. A.; BOHLEN, J.; SEMBER, A.; PEREZ, J.; VEYRUNES, F.; MUELLER, T. D.; POSTLETHWAIT, J. H.; SCHARTL, M.; HERPIN, A.; RÉTAUX, S.; GUIGUEN, Y. A supernumerary “B-sex” chromosome drives male sex determination in the Pachón cavefish, Astyanax mexicanus. **Current Biology**, v. 31, n. 21, p. 4800- 4809.e9, 8 nov. 2021. Disponível em: <https://doi.org/10.1016/j.cub.2021.08.030>. Acesso em: 4 maio. 2022.

KARNOVSKY, M. J. A formaldehyde-glutaraldehyde fixative of high osmolality for use in electron microscopy. **Journal cell biology**, v. 27, p. 137–138, 1985. Disponível em: <https://www.researchgate.net/publication/244955881>. Acesso em: 28 fev. 2022.

KATOH, K.; STANDLEY, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. **Molecular Biology and Evolution**, v. 30, n. 4, p. 772–780, 1 abr. 2013.

KRATOCHVÍL, L.; GAMBLE, T.; ROVATSOS, M. Sex chromosome evolution among amniotes: Is the origin of sex chromosomes non-random? **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 376, n. 1833, 13 set. 2021. . Acesso em: 26 abr. 2022.

LAHN, B. T.; PAGE, D. C. Four evolutionary strata on the human X chromosome. **Science**, v. 286, n. 5441, 1999.

LI, H.; DURBIN, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. **Bioinformatics**, v. 26, n. 5, p. 589–595, 15 jan. 2010.

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, 2009a.

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, 2009b.

LI, M.; ZHANG, R.; FAN, G.; XU, W.; ZHOU, Q.; WANG, L.; LI, W.; PANG, Z.; YU, M.; LIU, Q.; LIU, X.; SCHARTL, M.; CHEN, S. Reconstruction of the Origin of a Neo-Y Sex Chromosome and Its Evolution in the Spotted Knifejaw, Oplegnathus punctatus. **Molecular biology and evolution**, v. 38, n. 6, p. 2615–2626, 19 maio 2021. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/33693787/>. Acesso em: 22 set. 2021.

LINKHART, T. A.; MOHAN, S.; BAYLINK, D. J. Growth factors for bone growth and repair: IGF, TGFβ and BMP. In: Bone, 1 SUPPL., 1996, [...]. 1996. v. 19

LIU, J.; WANG, Z.; LI, J.; XU, L.; LIU, J.; FENG, S.; GUO, C.; CHEN, S.; REN, Z.; RAO, J.; WEI, K.; CHEN, Y.; JARVIS, E. D.; ZHANG, G.; ZHOU, Q. A new emu genome illuminates the evolution of genome configuration and nuclear architecture of avian chromosomes. **Genome Research**, 2021.

LOGSDON, G. A.; VOLLGER, M. R.; EICHLER, E. E. Long-read human genome sequencing and its applications. **Nature Reviews Genetics 2020 21:10**, v. 21, n. 10, p. 597–614, 5 jun. 2020a. Disponível em: <https://www.nature.com/articles/s41576-020-0236-x>. Acesso em: 23 ago. 2021.

LOGSDON, G. A.; VOLLGER, M. R.; EICHLER, E. E. Long-read human genome sequencing and its applications. **Nature Reviews Genetics 2020 21:10**, v. 21, n. 10, p. 597–614, 5 jun. 2020b.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, n. 12, p. 1–21, 5 dez. 2014.

MANNI, M.; BERKELEY, M. R.; SEPPEY, M.; SIMÃO, F. A.; ZDOBNOV, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. **Molecular Biology and Evolution**, 28 jul. 2021a. Disponível em: <https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab199/6329644>. Acesso em: 16 ago. 2021.

MANNI, M.; BERKELEY, M. R.; SEPPEY, M.; SIMÃO, F. A.; ZDOBNOV, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. **Molecular Biology and Evolution**, 28 jul. 2021b.

MARÇAIS, G.; KINGSFORD, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. **Bioinformatics**, v. 27, n. 6, p. 764–770, 15 mar. 2011.

MELO, B. F.; SIDLAUSKAS, B. L.; NEAR, T. J.; ROXO, F. F.; GHEZELAYAGH, A.; OCHOA, L. E.; STIASSNY, M. L. J.; ARROYAVE, J.; CHANG, J.; FAIRCLOTH, B. C.; MACGUIGAN, D. J.; HARRINGTON, R. C.; BENINE, R. C.; BURNS, M. D.; HOEKZEMA, K.; SANCHES, N. C.; MALDONADO-OCAMPO, J. A.; CASTRO, R. M. C.; FORESTI, F.; ALFARO, M. E.; OLIVEIRA, C. Accelerated Diversification Explains the Exceptional Species Richness of Tropical Characoid Fishes. **Systematic Biology**, v. 0, n. 0, p. 1–15, 7 jun. 2021a.

MELO, B. F.; SIDLAUSKAS, B. L.; NEAR, T. J.; ROXO, F. F.; GHEZELAYAGH, A.; OCHOA, L. E.; STIASSNY, M. L. J.; ARROYAVE, J.; CHANG, J.; FAIRCLOTH, B. C.; MACGUIGAN, D. J.; HARRINGTON, R. C.; BENINE, R. C.; BURNS, M. D.; HOEKZEMA,

K.; SANCHES, N. C.; MALDONADO-OCAMPO, J. A.; CASTRO, R. M. C.; FORESTI, F.; ALFARO, M. E.; OLIVEIRA, C. Accelerated Diversification Explains the Exceptional Species Richness of Tropical Characoid Fishes. **Systematic Biology**, v. 0, n. 0, p. 1–15, 7 jun. 2021b.

MEYER, A.; SCHLOISSNIG, S.; FRANCHINI, P.; DU, K.; WOLTERING, J. M.; IRISARRI, I.; WONG, W. Y.; NOWOSHILOW, S.; KNEITZ, S.; KAWAGUCHI, A.; FABRIZIUS, A.; XIONG, P.; DECHAUD, C.; SPAINK, H. P.; VOLFF, J.-N.; SIMAKOV, O.; BURMESTER, T.; TANAKA, E. M.; SCHARTL, M. Giant lungfish genome elucidates the conquest of land by vertebrates. **Nature 2021 590:7845**, v. 590, n. 7845, p. 284–289, 18 jan. 2021.

O'LEARY, N. A.; WRIGHT, M. W.; BRISTER, J. R.; CIUFO, S.; HADDAD, D.; MCVEIGH, R.; RAJPUT, B.; ROBBERTSE, B.; SMITH-WHITE, B.; AKO-ADJEI, D.; ASTASHYN, A.; BADRETDIN, A.; BAO, Y.; BLINKOVA, O.; BROVER, V.; CHETVERNIN, V.; CHOI, J.; COX, E.; ERMOLAEVA, O.; FARRELL, C. M.; GOLDFARB, T.; GUPTA, T.; HAFT, D.; HATCHER, E.; HLAVINA, W.; JOARDAR, V. S.; KODALI, V. K.; LI, W.; MAGLOTT, D.; MASTERSON, P.; MCGARVEY, K. M.; MURPHY, M. R.; O'NEILL, K.; PUJAR, S.; RANGWALA, S. H.; RAUSCH, D.; RIDDICK, L. D.; SCHOCH, C.; SHKEDA, A.; STORZ, S. S.; SUN, H.; THIBAUD-NISSEN, F.; TOLSTOY, I.; TULLY, R. E.; VATSAN, A. R.; WALLIN, C.; WEBB, D.; WU, W.; LANDRUM, M. J.; KIMCHI, A.; TATUSOVA, T.; DICUCCIO, M.; KITTS, P.; MURPHY, T. D.; PRUITT, K. D. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. **Nucleic acids research**, v. 44, n. D1, p. D733–D745, 2016.

OLIVEIRA, C.; FORESTI, F.; HILSDORF, A. W. S. Genetics of neotropical fish: From chromosomes to populations. **Fish Physiology and Biochemistry**, v. 35, n. 1, p. 81–100, 2009.

PAN, Q.; KAY, T.; DEPINCÉ, A.; ADOLFI, M.; SCHARTL, M.; GUIGUEN, Y.; HERPIN, A. Evolution of master sex determiners: TGF-β signalling pathways at regulatory crossroads. **Philosophical Transactions of the Royal Society B**, v. 376, n. 1832, 30 ago. 2021.

PEICHEL, C. L.; MCCANN, S. R.; ROSS, J. A.; NAFTALY, A. F. S.; URTON, J. R.; CECH, J. N.; GRIMWOOD, J.; SCHMUTZ, J.; MYERS, R. M.; KINGSLEY, D. M.; WHITE, M. A. Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. **Genome Biology 2020 21:1**, v. 21, n. 1, p. 1–31, 19 jul. 2020a. Disponível em: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02097-x>. Acesso em: 22 set. 2021.

PEICHEL, C. L.; MCCANN, S. R.; ROSS, J. A.; NAFTALY, A. F. S.; URTON, J. R.; CECH, J. N.; GRIMWOOD, J.; SCHMUTZ, J.; MYERS, R. M.; KINGSLEY, D. M.; WHITE, M. A. Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. **Genome Biology 2020 21:1**, v. 21, n. 1, p. 1–31, 19 jul. 2020b.

PETERSON, B. K.; WEBER, J. N.; KAY, E. H.; FISHER, H. S.; HOEKSTRA, H. E. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. **PLoS ONE**, v. 7, n. 5, 2012.

POLLARD, M. O.; GURDASANI, D.; MENTZER, A. J.; PORTER, T.; SANDHU, M. S. Long reads: their purpose and place. **Human Molecular Genetics**, v. 27, n. R2, p. R234–R241,

1 ago. 2018. Disponível em: <https://academic.oup.com/hmg/article/27/R2/R234/4996216>. Acesso em: 25 ago. 2021.

PORTO-FORESTI, F.; HASHIMOTO, D. T.; ALVES, A. L.; ALMEIDA, R. B. C.; SENHORINI, J. A.; BORTOLOZZI, J.; FORESTI, F. Cytogenetic markers as diagnoses in the identification of the hybrid between Piau&u (Leporinus macrocephalus) and Piapara (Leporinus elongatus). **Genetics and Molecular Biology**, v. 31, n. 1 SUPPL. 1, p. 195–202, 2008.

PRIYANKA, P. P.; YENUGU, S. **Coiled-Coil Domain-Containing (CCDC) Proteins: Functional Roles in General and Male Reproductive PhysiologyReproductive Sciences**2021.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A. R.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P. I. W.; DALY, M. J.; SHAM, P. C. PLINK: A tool set for whole-genome association and population-based linkage analyses. **American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, 2007.

RAFATI, N.; CHEN, J.; HERPIN, A.; PETTERSSON, M. E.; HAN, F.; FENG, C.; WALLERMAN, O.; RUBIN, C.-J.; PÉRON, S.; COCCO, A.; LARSSON, M.; TRÖTSCHEL, C.; POETSCH, A.; KORSCHING, K.; BÖNIGK, W.; KÖRSCHEN, H. G.; BERG, F.; FOLKVORD, A.; KAUPP, U. B.; SCHARTL, M.; ANDERSSON, L. Reconstruction of the birth of a male sex chromosome present in Atlantic herring. **Proceedings of the National Academy of Sciences**, v. 117, n. 39, p. 24359–24368, 29 set. 2020. Disponível em: <https://www.pnas.org/content/117/39/24359>. Acesso em: 22 set. 2021.

RAMIREZ, J. L.; BIRINDELLI, J. L. O.; GALETTI, P. M. A new genus of Anostomidae (Ostariophysi: Characiformes): Diversity, phylogeny and biogeography based on cytogenetic, molecular and morphological data. **Molecular Phylogenetics and Evolution**, v. 107, p. 308–323, 2017. Disponível em: <http://dx.doi.org/10.1016/j.ympev.2016.11.012>.

RAMOS, L.; ANTUNES, A. Decoding sex: Elucidating sex determination and how high-quality genome assemblies are untangling the evolutionary dynamics of sex chromosomes. **Genomics**, v. 114, n. 2, p. 110277, 1 mar. 2022. Disponível em: <https://doi.org/10.1016/j.ygeno.2022.110277>. Acesso em: 27 abr. 2022.

RASTAS, P. Lep-MAP3: Robust linkage mapping even for low-coverage whole genome sequencing data. **Bioinformatics**, v. 33, n. 23, p. 3726–3732, 2017.

RASTAS, P.; PAULIN, L.; HANSKI, I.; LEHTONEN, R.; AUVINEN, P.; BRUDNO, M. Lep-MAP: Fast and accurate linkage map construction for large SNP datasets. **Bioinformatics**, v. 29, n. 24, 2013.

REID, D. P.; SMITH, C. A.; ROMMENS, M.; BLANCHARD, B.; MARTIN-ROBICHAUD, D.; REITH, M. A genetic linkage map of Atlantic halibut (Hippoglossus hippoglossus L.). **Genetics**, v. 177, n. 2, 2007.

REYNALTE-TATAJE, D.; ZANIBONI-FILHO, E.; MUELBERT, B. Estadios do desenvolvimento embrionário do piavuçu Leporinus macrocephalus (Garavello & Britski, 1988). **Acta Scientiarum. Animal Sciences**, v. 23, n. 0, p. 823, 9 maio 2008. Disponível em:

<https://periodicos.uem.br/ojs/index.php/ActaSciAnimSci/article/view/2614>. Acesso em: 26 ago. 2021.

RHIE, A.; MCCARTHY, S. A.; FEDRIGO, O.; DAMAS, J.; FORMENTI, G.; LONDON, S. E.; CLAYTON, D. F.; MELLO, C. v; FRIEDRICH, S. R. Towards complete and error-free genome assemblies of all vertebrate species. p. 1–56, 2020a.

RHIE, A.; WALENZ, B. P.; KOREN, S.; PHILLIPPY, A. M. Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. **Genome Biology**, v. 21, n. 1, p. 1–27, 2020b.

RHIE, A.; WALENZ, B. P.; KOREN, S.; PHILLIPPY, A. M. Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. **Genome Biology**, v. 21, n. 1, p. 1–27, 2020c.

ROCHA, C. M. C. da; RESENDE, E. K. de; ROUTLEDGE, E. A. B.; LUNDSTEDT, L. M. Avanços na pesquisa e no desenvolvimento da aquicultura brasileira. **Pesquisa Agropecuária Brasileira**, v. 48, n. 8, p. iv–vi, 2013a.

ROCHA, C. M. C. da; RESENDE, E. K. de; ROUTLEDGE, E. A. B.; LUNDSTEDT, L. M. Avanços na pesquisa e no desenvolvimento da aquicultura brasileira. **Pesquisa Agropecuária Brasileira**, v. 48, n. 8, p. iv–vi, 2013b.

SCHARTL, M.; KNEITZ, S.; VOLKOFF, H.; ADOLFI, M.; SCHMIDT, C.; FISCHER, P.; MINX, P.; TOMLINSON, C.; MEYER, A.; WARREN, W. C. The Piranha Genome Provides Molecular Insight Associated to Its Unique Feeding Behavior. **Genome Biology and Evolution**, v. 11, n. 8, p. 2099–2106, 1 ago. 2019a.

SCHARTL, M.; KNEITZ, S.; VOLKOFF, H.; ADOLFI, M.; SCHMIDT, C.; FISCHER, P.; MINX, P.; TOMLINSON, C.; MEYER, A.; WARREN, W. C. The Piranha Genome Provides Molecular Insight Associated to Its Unique Feeding Behavior. **Genome Biology and Evolution**, v. 11, n. 8, p. 2099–2106, 1 ago. 2019b.

SHAO, F.; LUDWIG, A.; MAO, Y.; LIU, N.; PENG, Z. Chromosome-level genome assembly of the female western mosquitofish (Gambusia affinis). **GigaScience**, v. 9, n. 8, p. 1–10, 1 ago. 2020. Disponível em: <https://academic.oup.com/gigascience/article/9/8/giaa092/5897807>. Acesso em: 22 set. 2021.

SIDONIO, L.; CAVALCANTI, I.; CAPANEMA, L.; MORCH, R.; LIMA, J.; BURNS, V.; JÚNIOR, A. J. A.; AMARAL, J. V. Experiências internacionais aquícolas e oportunidades de desenvolvimento da aquicultura no Brasil : proposta de inserção do BNDES. **BNDES Setorial**, v. 36, 2012.

SMITH, S. R.; NORMANDEAU, E.; DJAMBAZIAN, H.; NAWARATHNA, P. M.; BERUBE, P.; MUIR, A. M.; RAGOUSSIS, J.; PENNEY, C. M.; SCRIBNER, K. T.; LUIKART, G.; WILSON, C. C.; BERNATCHEZ, L. A chromosome-anchored genome assembly for Lake Trout (Salvelinus namaycush). **Molecular Ecology Resources**, v. 22, n. 2, p. 679–694, 1 fev. 2022.

SONESON, C.; LOVE, M. I.; ROBINSON, M. D.; FLOOR, S. N. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 2; peer review: 2 approved] report report. 2016.

SONG, W.; LI, Y.; ZHAO, Y.; LIU, Y.; NIU, Y.; PANG, R.; MIAO, G.; LIAO, X.; SHAO, C.; GAO, F.; CHEN, S. Construction of a High-Density Microsatellite Genetic Linkage Map and Mapping of Sexual and Growth-Related Traits in Half-Smooth Tongue Sole (Cynoglossus semilaevis). **PLoS ONE**, v. 7, n. 12, 2012.

STAMATAKIS, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. **Bioinformatics**, v. 30, n. 9, p. 1312–1313, 1 maio 2014.

STEINKE, D.; SALZBURGER, W.; MEYER, A. Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. **Journal of Molecular Evolution**, v. 62, n. 6, p. 772–784, 11 jun. 2006.

SUMNER, A. T. A simple technique for demonstrating centromeric heterochromatin. **Experimental Cell Research**, v. 75, n. 1, 1972.

TOMASZKIEWICZ, M.; MEDVEDEV, P.; MAKOVA, K. D. Y and W Chromosome Assemblies: Approaches and Discoveries. **Trends in Genetics**, v. 33, n. 4, p. 266–282, 1 abr. 2017. . Acesso em: 27 abr. 2022.

UTSUNOMIA, R.; SILVA, D. M. Z. de A.; RUIZ-RUANO, F. J.; GOES, C. A. G.; MELO, S.; RAMOS, L. P.; OLIVEIRA, C.; PORTO-FORESTI, F.; FORESTI, F.; HASHIMOTO, D. T. Satellitome landscape analysis of Megaleporinus macrocephalus (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. **Scientific Reports**, v. 9, n. 1, p. 1–10, 2019a.

UTSUNOMIA, R.; SILVA, D. M. Z. de A.; RUIZ-RUANO, F. J.; GOES, C. A. G.; MELO, S.; RAMOS, L. P.; OLIVEIRA, C.; PORTO-FORESTI, F.; FORESTI, F.; HASHIMOTO, D. T. Satellitome landscape analysis of Megaleporinus macrocephalus (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. **Scientific Reports**, v. 9, n. 1, p. 1–10, 2019b.

VARELA, E. S.; BEKAERT, M.; GANECO-KIRSCHNIK, L. N.; TORATI, L. S.; SHIOTSUKI, L.; DE ALMEIDA, F. L.; VILLELA, L. C. V.; REZENDE, F. P.; DA SILVA BARROSO, A.; DE FREITAS, L. E. L.; TAGGART, J. B.; MIGAUD, H. A high-density linkage map and sex-linked markers for the Amazon Tambaqui Colossoma macropomum. **BMC Genomics**, v. 22, n. 1, p. 1–10, 1 dez. 2021.

VOLFF, J.-N. Genome evolution and biodiversity in teleost fish. **Heredity 2005 94:3**, v. 94, n. 3, p. 280–294, 22 dez. 2004a. Disponível em: <https://www.nature.com/articles/6800635>. Acesso em: 22 set. 2021.

VOLFF, J.-N. Genome evolution and biodiversity in teleost fish. **Heredity 2005 94:3**, v. 94, n. 3, p. 280–294, 22 dez. 2004b.

VURTURE, G. W.; SEDLAZECK, F. J.; NATTESTAD, M.; UNDERWOOD, C. J.; FANG, H.; GURTOWSKI, J.; SCHATZ, M. C. GenomeScope: Fast reference-free genome profiling from short reads. **Bioinformatics**, v. 33, n. 14, p. 2202–2204, 2017.

WANG, K.; WANG, J.; ZHU, C.; YANG, L.; REN, Y.; RUAN, J.; FAN, G.; HU, J.; XU, W.; BI, X.; ZHU, Y.; SONG, Y.; CHEN, H.; MA, T.; ZHAO, R.; JIANG, H.; ZHANG, B.; FENG, C.; YUAN, Y.; GAN, X.; LI, Y.; ZENG, H.; LIU, Q.; ZHANG, Y.; SHAO, F.; HAO, S.; ZHANG, H.; XU, X.; LIU, X.; WANG, D.; ZHU, M.; ZHANG, G.; ZHAO, W.; QIU, Q.; HE,

S.; WANG, W. African lungfish genome sheds light on the vertebrate water-to-land transition. **Cell**, v. 184, n. 5, p. 1362- 1376.e18, 4 mar. 2021.

WARREN, W. C.; BOGGS, T. E.; BOROWSKY, R.; CARLSON, B. M.; FERRUFINO, E.; GROSS, J. B.; HILLIER, L.; HU, Z.; KEENE, A. C.; KENZIOR, A.; KOWALKO, J. E.; TOMLINSON, C.; KREMITZKI, M.; LEMIEUX, M. E.; GRAVES-LINDSAY, T.; MCGAUGH, S. E.; MILLER, J. T.; MOMMERSTEEG, M. T. M.; MORAN, R. L.; PEUSS, R.; RICE, E. S.; RIDDLE, M. R.; SIFUENTES-ROMERO, I.; STANHOPE, B. A.; TABIN, C. J.; THAKUR, S.; YAMAMOTO, Y.; ROHNER, N. A chromosome-level genome of Astyanax mexicanus surface fish for comparing population-specific genetic differences contributing to trait evolution. **Nature Communications 2021 12:1**, v. 12, n. 1, p. 1–12, 4 mar. 2021a.

WARREN, W. C.; BOGGS, T. E.; BOROWSKY, R.; CARLSON, B. M.; FERRUFINO, E.; GROSS, J. B.; HILLIER, L.; HU, Z.; KEENE, A. C.; KENZIOR, A.; KOWALKO, J. E.; TOMLINSON, C.; KREMITZKI, M.; LEMIEUX, M. E.; GRAVES-LINDSAY, T.; MCGAUGH, S. E.; MILLER, J. T.; MOMMERSTEEG, M. T. M.; MORAN, R. L.; PEUSS, R.; RICE, E. S.; RIDDLE, M. R.; SIFUENTES-ROMERO, I.; STANHOPE, B. A.; TABIN, C. J.; THAKUR, S.; YAMAMOTO, Y.; ROHNER, N. A chromosome-level genome of Astyanax mexicanus surface fish for comparing population-specific genetic differences contributing to trait evolution. **Nature Communications 2021 12:1**, v. 12, n. 1, p. 1–12, 4 mar. 2021b.

WRIGHT, A. E.; DEAN, R.; ZIMMER, F.; MANK, J. E. **How to make a sex chromosomeNature Communications**2016.

XIAO, Y.; XIAO, Z.; MA, D.; LIU, J.; LI, J. Genome sequence of the barred knifejaw Oplegnathus fasciatus (Temminck & Schlegel, 1844): The first chromosome-level draft genome in the family Oplegnathidae. **GigaScience**, v. 8, n. 3, p. 1–8, 2019.

XUE, L.; GAO, Y.; WU, M.; TIAN, T.; FAN, H.; HUANG, Y.; HUANG, Z.; LI, D.; XU, L. Telomere-to-telomere assembly of a fish Y chromosome reveals the origin of a young sex chromosome pair. **Genome Biology 2021 22:1**, v. 22, n. 1, p. 1–20, 12 jul. 2021a. Disponível em: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02430-y>. Acesso em: 20 out. 2021.

XUE, L.; GAO, Y.; WU, M.; TIAN, T.; FAN, H.; HUANG, Y.; HUANG, Z.; LI, D.; XU, L. Telomere-to-telomere assembly of a fish Y chromosome reveals the origin of a young sex chromosome pair. **Genome Biology 2021 22:1**, v. 22, n. 1, p. 1–20, 12 jul. 2021b.

YAN, Y. L.; BATZEL, P.; TITUS, T.; SYDES, J.; DESVIGNES, T.; BREMILLER, R.; DRAPER, B.; POSTLETHWAIT, J. H. A hormone that lost its receptor: Anti-müllerian hormone (AMH) in Zebrafish gonad development and sex determination. **Genetics**, v. 213, n. 2, 2019a.

YAN, Y. L.; BATZEL, P.; TITUS, T.; SYDES, J.; DESVIGNES, T.; BREMILLER, R.; DRAPER, B.; POSTLETHWAIT, J. H. A hormone that lost its receptor: Anti-müllerian hormone (AMH) in Zebrafish gonad development and sex determination. **Genetics**, v. 213, n. 2, 2019b.

YOU, X.; SHAN, X.; SHI, Q. **Research advances in the genomics and applications for molecular breeding of aquaculture animalsAquaculture**2020.

YUE, G. H.; WANG, L. Current status of genome sequencing and its applications in aquaculture. 2016a.

YUE, G. H.; WANG, L. Current status of genome sequencing and its applications in aquaculture. 2016b.

ZIMIN, A. V.; SALZBERG, S. L. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. **PLoS Computational Biology**, v. 16, n. 6, 2020.

ZIMINID, A. V; SALZBERGID, S. L. The SAMBA tool uses long reads to improve the contiguity of genome assemblies. **PLOS Computational Biology**, v. 18, n. 2, p. e1009860, 4 fev. 2022.

8. Manuscrito I

Unraveling the sex chromosome in *Megaleporinus macrocephalus*, a neotropical fish with ZW sex chromosome system

Abstract

Megaleporinus macrocephalus is a Neotropical fish within Characoidei. The species presents a well-established heteromorphic ZZ/ZW sex-determination system and thus, constitute a good model for studying the evolution of the W chromosome in fishes. We used PacBio long reads, BGISEQ short reads, and Hi-C data to assemble a chromosome-level pseudo-haploid genome for the species. We generated family segregation information to construct a genetic map and whole-genome sequences along with RNA-seq data of males and females to provide insights into the sex determination process and chromosomal evolution of *M. macrocephalus*. The genome comprises 27 chromosomes, which corresponds to the karyotype data of the species. The sex chromosome is composed of 24 Mb of the pseudo-autosomal region of Z and W, and specific regions of Z and W (18 Mb and 3Mb, respectively). We found evidence of recombination suppression within the W-specific region. The Z and W chromosomes comprise at least one and two evolutionary *strata*, respectively. The CCDC114 and IGFBP-6 are candidate sex determination genes of the species. This is the first report so far of a sex chromosome assembly of a Neotropical fish species.

Keywords: Neotropical fish, Anostomidae, Characoidei, Characiformes, PacBio, Hi-C

8.1 Background

Sex chromosomes are present in most vertebrate genomes. Most mammals have male heterogamy (XY males and XX females), and all birds studied have female heterogamy (ZW females and ZZ males). Species of reptiles, amphibians and fish with genetic (rather than environmental) sex determination can have male or female heterogamy (TOMASZKIEWICZ; MEDVEDEV; MAKOVA, 2017).

These chromosomes constitute the most complex regions of the genome to sequence and assemble. The main challenges are associated with its haploid nature and high sequence divergence from heterogametic sex, as well as the accumulation of repetitive sequences (RHIE *et al.*, 2020a). As a result of these disadvantages, sex chromosomes are often disregarded in sequencing designs and when included, homogametic sexes (XX females or ZZ males) are generally preferred. Currently, with the advancement of sequencing technologies and the development of new software, an exponential increase in available sex chromosome assemblies has been observed e. g., spotted knifejaw (XIAO *et al.*, 2019), emu (LIU *et al.*, 2021), threespine stickleback (PEICHEL *et al.*, 2020a), zig-zag eel (XUE *et al.*, 2021a).

The theory of canonical evolution explains that the emergence of a sex chromosome occurs when an ancestral autosome acquires a new sex-determining locus. Eventually, the appearance of a sexually antagonistic gene near the locus of sex determination favors a series of inversions which lead to the suppression of recombination between the X and Y or Z and W chromosomes. The lack of recombination leads to the accumulation of repetitive DNA, which can lead to an increase in the size of the Y or W in the short term, but typically results in large-scale deletions, a reduction in the physical size of the sex-limited chromosome, and highly heteromorphic sex chromosome (WRIGHT *et al.*, 2016).

The new data on sex chromosomes has significantly contributed to the knowledge of the evolution of sex chromosomes, and it is already understood that the process is much more

complex and not unidirectional like the classical canonical theory (KRATOCHVÍL; GAMBLE; ROVATSOS, 2021). As an example, we can mention the emergence of a sex determination gene “immune” to degenerative processes due to its location in a region of low recombination of the centromere of the homomorphic sex chromosome (XUE *et al.*, 2021a); the fusion of a B chromosome with an autosome that possibly gave rise to a giant sex chromosome in Oreochromini (CONTE *et al.*, 2021) and also the occurrence of a “sex B” chromosome (IMARAZENE *et al.*, 2021).

Most genes retained in both sex chromosome systems (XX/XY, ZZ/ZW) are convergent, have regulatory functions, and are widely expressed throughout the body (i.e., in no specific tissue) (RAMOS; ANTUNES, 2022). Furthermore, theories suggest that they have survived strong pressures from purifying selection to remain retained on the sex chromosomes as they are generally gene dosage-sensitive (e.g. embryos carrying one copy of the gene do not express sufficient gene dosage and are unviable) (BELLOTT *et al.*, 2017).

However, Y, X and Z, called male germline chromosomes, in addition to genes widely expressed throughout the body, also acquired and amplified during their evolutionary history families of masculinizing genes, which have a specific expression in the testes and act in spermatogenesis, and other male characteristics (RAMOS; ANTUNES, 2022). W is an exclusively female germline and, surprisingly, genes expressed exclusively in the ovaries or other female tissue and with feminizing action to date have not been found in its repertoire, unlike expected (BELLOTT *et al.*, 2017).

Teleosts represent more than 95% of all living fish species and approximately half of the extant vertebrate species (VOLFF, 2004). The incredible diversity of this group can be attributed to a genome-doubling event (TS-WGD) that occurred during the early evolution of the lineage (ALBERT; TAGLIACOLLO; DAGOSTA, 2020). This makes the group

particularly interesting in the study of evolutionary issues related to various aspects of biology, especially the evolution of sex chromosome systems.

Teleostei Y chromosome assemblies have been frequently reported in the literature, such as from the zig-zag eel (XUE et al., 2021), spotted knifejaw (LI, M. et al., 2021), stickleback (PEICHEL et al., 2021), and Atlantic herring (RAFATI et al., 2020). However, there are only two reports of W chromosome assemblies in fish, e.g., Chinese tongue sole (CHEN et al., 2014) and mosquitofish (SHAO et al., 2020). In addition, the former was assembled with short reads, which is not recommended for sex chromosome assemblies (PEICHEL et al., 2020), and the latter did not present details about the gene repertoire and evolution of W.

The genus *Megaleporinus* (Characiformes, Anostomidae) was established to separate *Leporinus* species that had a ZZ/ZW sexual system as a synapomorphy (RAMIREZ; BIRINDELLI; GALETTI, 2017). *Megaleporinus* sex chromosomes are in an intermediate stage of evolution (CHARLESWORTH, 2019) and emerged at least 12 million years ago after they diverged from *Leporinus* (RAMIREZ; BIRINDELLI; GALETTI, 2017).

The piauçu *Megaleporinus macrocephalus* is a freshwater fish from South America that has a well-established ZZ/ZW heteromorphic sex chromosome system (GALETTI, JR. et al., 1981). The W chromosome of *M. macrocephalus* is a large chromosome with a huge C-positive heterochromatic block occupying entire long arms, while the Z chromosome is a medium-sized metacentric with only portions of heterochromatin at the end of the long arms (GALETTI, JR. et al. ., 1981).

Due to the scarce record of W assemblies in vertebrates and especially in fish, it is necessary to sequence more W models to answer fundamental questions in evolutionary biology such as the possible existence of a feminizing sex-determining gene in W and its gene repertoire (PEICHEL et al., 2020; SMEDS et al., 2015).

The piauçu is an excellent model for the study of the W chromosome, therefore, this project proposes to carry out the genome assembly of *Megaleporinus macrocephalus* as well to promote insights regarding its sex chromosome evolution.

8.2 Ethics Statement

This study was conducted in strict accordance with the recommendations of the National Council for Control of Animal Experimentation (CONCEA) (Brazilian Ministry of Science, Technology, and Innovation) and was approved by the Ethics Committee on Animal Use (CEUA number 4936/20) of Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus Jaboticabal, SP, Brazil.

8.3 Material

8.3.1 Female individual for genome sequencing

Tissue samples for genome sequencing were obtained from an adult ZW female of *Megaleporinus macrocephalus* from the broodstock of the Aquaculture Center of São Paulo State University. Cytogenetic analysis of this animal was performed using the lymphocyte culture technique described by (BERTOLLO; MOREIRA–FILHO; GALETTI, 1986) with some adjustments. C-banding was performed according to (SUMNER, 1972).

8.3.2 Biological material for linkage mapping

The breeders used in this study belong to the breeding population kept at the Aquaculture Center, São Paulo State University (UNESP), Jaboticabal city (São Paulo State, Brazil). Experimental tests were conducted at the Laboratory of Genetics in Aquaculture and Conservation (LaGeAC, UNESP, Brazil). The population was composed of 299 fish belonging to 4 full-sib families generated from single mating (1 female x 1 male) (Table 1). The families were produced during the breeding season of December 2018.

Induced spawning was performed using carp pituitary extract dissolved in saline solution (0.9% NaCl) and applied in two dosages, with a 12 h interval: the first and second dosage of 0.6 and 5.4 mg/kg for females, and a single dosage of 1.5 mg/kg for males, at the same time of the females' second dosage (REYNALTE-TATAJE; ZANIBONI-FILHO; MUELBERT, 2008). After hatching in 20 L conical fiberglass incubators, the larvae were transferred to tanks of 250 L. The larvae were fed with *Artemia nauplii* for 20 days. Gradually, the feed was replaced by 50% of crude protein. In the fingerling stage, 1.2 mm pelleted feeds were used (40% of crude protein) and provided twice daily (commercial feed Nutripiscis Presence).

Each full-sib family was kept separately in individual fiberglass tanks of 1 m3 up to 6 months old. The fish were kept in a water recirculation system, fitted with mechanical and biological filters, an external aeration system, and controlled temperature at 30 °C (standard deviation = 0.5 °C) using a thermal controller connected to heaters (2 × 500 watts). Temperature, dissolved oxygen, and pH were measured with a Multiparameter Water Quality Checker U-50 (Horiba). After this period, we collected blood samples for genomic analyses and the weight of all animals were registered with analytical balance (average weight was 6 g). Fish were then euthanized for sex identification.

Individual sex was verified by a PCR-based protocol using a chromosome W-probe (UTSUNOMIA et al., 2019) as well as by cytogenetic analysis. Chromosome preparations were obtained from kidney tissues using the technique described by (FORESTI; TOLEDO; TOLEDO, 1981).

8.3.3 Biological material for Resequencing (pool-sequencing)

We used resequencing (pool-sequencing, Pool-seq) analyses to contrast whole-genome sex differences in *M. macrocephalus*. For this, we collected samples of 20 males and 20 females originating from four commercial fish farms in Brazil. Briefly, fish were anesthetized with 0.1% benzocaine for blood collection. The sex of each fish was verified by cytogenetic analysis, as detailed above (8.4.3 8.3.2 Biological material for linkage mapping), and samples were clustered in separated male and female pools.

We also performed sequencing of males and females of *Leporinus friderici*, which is closely related to *M. macrocephalus*. One of the main characteristics that distinguish these species is the absence of differentiated sex chromosomes in the former (RAMIREZ; BIRINDELLI; GALETTI, 2017). For Pool-Sex experiments, we collected samples from 28 phenotypic males and 28

phenotypic females of a wild population of *L. friderici*. This natural population belongs to the Sapucaí-Mirim River (São Joaquim da Barra, São Paulo State, Brazil, 20°32'40.7"S 47°48'02.1"W), and the animals were captured under authorization N° 33435-1, issued by ICMBio (Chico Mendes Institute for the Conservation of Biodiversity, Brazilian Ministry of Environment). No animals were kept or transported to the laboratory. Fish were anesthetized with 0.1% benzocaine for the collection of fin clips. Phenotypic sex identification was performed by macroscopic visualization of the gonads during the breeding season when is possible to distinguish the fish phenotypic sex.

8.3.4 Biological material for RNA-seq

For RNA-seq experiments, 60 individuals that comprised the offspring of one full-sib family of *M. macrocephalus* were used. Fish were produced and maintained as described above (8.4.3 8.3.2 Biological material for linkage mapping). At 150 days after fertilization, when the period of sex differentiation recently occurred, according to previous experiments in this species (unpublished data); the two gonads and kidneys of each fish were dissected immediately. Fish were euthanized by benzocaine anesthetic overdose (2%) for sampling. One gonad was stored in RNAlater (Thermo Fischer Scientific) for RNA extraction, and the other was fixed for 24 hours in Karnovsky's solution (KARNOVSKY, 1985) and then stored in ethanol 70% for phenotypic sex identification in microscopy. The sex of each fish was verified by cytogenetic analysis, as detailed above (8.4.3 8.3.2 Biological material for linkage mapping). The phenotypic sex was obtained through gonadal histology as described by CORNELIO et al., 2017.

After phenotypic and genotypic sex identification, the samples were clustered in two pools: ZZ males and ZW females. Each pool had three biological replicates that consisted of 10 gonads, resulting in 6 libraries for RNA-sequencing.

8.4 Methods

8.4.1 Genome sequencing

To generate long reads, high molecular weight (HMW) DNA was extracted from blood using Nanobind CBB Big DNA Kit (Circulomics) and a continuous long read (CLR) library was constructed using SMRTbell Express Template Prep Kit 2.0. The library was sequenced in one single-molecule real-time (SMRT) cell of PacBio Sequel II System (Pacific Biosciences). To improve the accuracy of the long reads, a short read library was produced with MGIEasy PCR-Free Library Prep Set (MGI Tech Co., Ltd.) and sequenced on a BGI MGISEQ-2000 150 bp PE. Finally, to merge the scaffolds into putative chromosomes, we generated a chromatin interaction (Hi-C) library using Proximo Hi-C Library Prep Kit (Phase Genomics) on liver tissue with *in vivo* cross-linking. Sequencing was performed on an Illumina NovaSeq 6000 150 bp PE.

8.4.2 Genome Size Estimate

MGISEQ short reads were used to estimate the haploid genome size, rate of heterozygosity, and abundance of repetitive elements. First, the reads were trimmed with Trimommatic v0.39 (BOLGER; LOHSE; USADEL, 2014) and bases with an average quality < 20 within a sliding window of 4 bp and bases with quality < 20 at the beginning and the end of the reads were removed. Reads with length < 36 were also discarded. After filtering, Jellyfish (MARÇAIS; KINGSFORD, 2011) was used to count canonical *k*-mers (-C flag) of length ranging from 21 to 24. The resulting *k*-mer profile was loaded on GenomeScope (VURTURE *et al.*, 2017).

8.4.3 Genome Assembly

An initial contig assembly was performed with Falcon/Falcon-Unzip (CHIN *et al.*, 2016) with a minimum read length cutoff of 5,000 bp. Falcon (CHIN *et al.*, 2016) was run with default parameters, except for computing the overlaps. Raw read overlaps were computed with daligner parameters -v -k16 -w7 -h64 -e0.70 -s1000 -M27 -H5000 to better reflect the higher error rate in PacBio Sequel II. Preassembled read (pread) overlaps were computed with daligner parameters -v -k20 -w6 -h256 -e0.96 -s1000 -l2500 -M27 -H5000. Falcon-Unzip (CHIN *et al.*, 2016) was run with default parameters and resulted in a set of primary and alternate contigs. False duplications in the contigs were removed with Purge_Dups (https://github.com/dfguan/purge_dups). Short-read polishing was made with Polca (ZIMIN; SALZBERG, 2020). To polish primary and alternate assemblies, we first concatenated them and followed with one round of short-read polishing. To improve the assembly's contiguity, we used PacBio long reads > 10 kb to fill in spanned gaps with SAMBA (ZIMINID; SALZBERGID, 2022). The Juicer/3d-dna pipeline (DUDCHENKO *et al.*, 2017) was used to orient scaffolds into putative chromosomes. First, we generated a file with the location of DpnII enzyme restriction sites in the assembly (*generate_site_positions.py* Juicer script), and a file with scaffolds sizes. Second, Hi-C reads were aligned to the assembly and filtered by Juicer (DURAND *et al.*, 2016) to generate a duplicated-free list of paired alignments (merged_nodups file). At last, 3d-dna (DUDCHENKO *et al.*, 2017) was run with a minimum scaffold size of 10 kb. The resulted contact map was manually curated in Juicebox Assembly Tools (JBAT) following a post curation process. At last, we performed an extra round of polishing with Polca (ZIMIN; SALZBERG, 2020). We also assembled the mitogenome using the MGISEQ reads.

8.4.4 Quality Assessment of Genome

The correctness was evaluated in each assembly step using Merqury (RHIE *et al.*, 2020b). The tool compared assembly *k*-mers to those found in the unassembled highly accurate MGISEQ short

reads to estimate base-level accuracy (consensus quality value, QV) and *k*-mer completeness. The QV represents a log-scaled probability of error for the consensus base calls. Contiguity measures such as contig and scaffold N50 were obtained with the *stats.sh* script of BBMap (https://sourceforge.net/projects/bbmap). To access the completeness of the genome, we performed BUSCO analysis (MANNI *et al.*, 2021a) using the Actinopterygii dataset. The assembly was verified for contamination by the National Center for Biotechnology Information (NCBI) submission protocols. All the contaminated scaffolds identified were removed.

8.4.5 Karyotype Validation

To validate the quality of our assembly, we performed a Pearson's correlation of the estimated size in base pair (bp), based on the average karyotype size in micrometers (µm), and the assembled size (bp) of each chromosome. For this, we measured both arms of each chromosome pair of the female karyotype and calculated an average size (µm) for each chromosome. The estimated chromosome size was calculated using the formula: chromosome average size (µm) x total genome size (bp) / total karyotype size (µm).

8.4.6 Repeat Annotation

We used RepeatModeler2 (FLYNN *et al.*, 2020), with the LTR option enabled, to produce a custom *de novo* library of the repeats present in the genome. Next, Repeat Masker (https://www.repeatmasker.org/) was used to identify, classify, and mask repetitive elements, including low-complexity sequences and interspersed repeats. We used a combined library to run Repeat Masker. First, the RepBase RepeatMasker Edition (version 20181026) was combined with the Dfam library with *addRepBase.pl* and *configure.pl* scripts. Then, only the repeats present in Teleost were selected with *famdb.py*. Finally, the custom de novo library, the Teleost repeat

sequences, and a satellite library of the species (UTSUNOMIA et al., 2019) were concatenated. Repeats elements were soft masked with RepeatMasker.

8.4.7 Gene Prediction and Annotation

We performed gene prediction with *ab initio* and homology-based methods using BRAKER (HOFF *et al.*, 2019) pipeline. First, BRAKER1 (HOFF *et al.*, 2016a) used RNA-seq data (8.4.4 8.3.4 Biological material for RNA-seq) as extrinsic evidence to predict introns. Next, BRAKER2 (BRŮNA *et al.*, 2021a) used protein homology information from Orthodb sequences of Vertebrata. At last, TSEBRA (GABRIEL *et al.*, 2021a) selected the best transcripts from both BRAKER predictions to increase their accuracies. Then, we performed a sanity check on the dataset to include only high-quality predictions. To assign functional annotation to the gene models, we performed searches using the predicted proteins with the Actinopterygii dataset of UniProtKB (BATEMAN *et al.*, 2021a). Search results were loaded into Blast2GO (GÖTZ *et al.*, 2008), mapped, and annotated. The quality of the annotation was evaluated using BUSCO (MANNI *et al.*, 2021b).

8.4.8 Comparative Genomics

To identify gene families among *M. macrocephalus* and other species, we downloaded the whole genome protein sequences from Ensembl release 105 (HOWE *et al.*, 2021) of cavefish *Astyanax mexicanus* (GCA_000372685.2), goldfish *Carassius auratus* (GCA_003368295.1), common carp *Cyprinus carpio* (GCA_000951615.2), zebrafish *Danio rerio* (GCA_000002035.4), European seabass *Dicentrarchus labrax* (GCA_000689215.1), channel catfish *Ictalurus punctatus* (GCA_001660625.1), coelacanth *Latimeria chalumnae* (GCA_000225785.1), rainbow trout *Oncorhynchus mykiss* (GCA_002163495.1), Nile tilapia *Oreochromis niloticus*

(GCA_001858045.3), Amazon molly *Poecilia formosa* (GCA_000485575.1), red-bellied piranha *Pygocentrus nattereri* (GCA_001682695.1), Japanese pufferfish *Takifugu rubripes* (GCA_901000725.2); and tambaqui *Colossoma macropomum* (GCF_904425465.1) from RefSeq (O'LEARY *et al.*, 2016). We extracted the longest transcript variant per gene (*primary_transcript.py*) and then the dataset was used as input to Orthofinder v2.5.4 (EMMS; KELLY, 2019) to access the orthogroups. The single-copy orthologs genes were used to generate a phylogeny. First, the sequences were aligned with MAFFT v7.490 (KATOH; STANDLEY, 2013) and trimmed with trimAL v1.4. rev15 (CAPELLA-GUTIÉRREZ; SILLA-MARTÍNEZ; GABALDÓN, 2009), using the automated1 option. Next, all sequences were concatenated using *geneStitcher.py* (https://github.com/ballesterus/Utensils/blob/master/geneStitcher.py) and used to construct a phylogeny with RAxML v8.2.12 (STAMATAKIS, 2014) using the criteria of Maximum Likelihood (M.L) and 1,000 bootstrap replicates as statistic support.

8.4.9 Linkage mapping

SNP genotyping

DNA was extracted from blood samples with Wizard Genomic DNA Purification kit (Promega) and quality was verified in 1% agarose gel electrophoresis. Purity was accessed in Nanodrop One and concentration (ng/µl) was measured by Qubit fluorometer with Qubit dsDNA HS Assay kit (Invitrogen, USA). We used a modified version of the protocol described by (PETERSON *et al.*, 2012) for the construction of ddRADseq (Double digest restriction-site associated DNA) libraries. Briefly, 75 ng of genomic DNA from each individual was digested (8 U/reaction) using the combination of two restriction enzymes, SphI and MluCI (New England Biolabs), and ligated to specific adapters (P1 and P2, 0.25 µM) using the enzyme T4 DNA ligase, at 23ºC for one and a half hour and 65ºC for 10 minutes to heat kill the enzyme. The P1 adapters have

additional 5 nucleotides that function as individual tags (barcode). The selection of digested fragments was performed using E-Gel Power Snap System (Thermo Fisher Scientific) with fragments of approximately 350 bp. Subsequently, PCR assays were performed to incorporate the identification of each library. In total, 7 libraries were constructed, with an average of 46 samples/library. PCR was performed under the conditions of the Platinum SuperFi DNA Polymerase enzyme (Thermo Fischer Scientific). The reactions were purified with the ProNex Size-Selective Purification System kit (Promega) and the concentration was checked again by fluorometry in the Qubit 3.0 instrument (Thermo Fisher Scientific). Finally, the libraries were sequenced in 2 lanes of Illumina Hiseq2500 150 PE, using 15 % PhiX (Novogene).

The overall quality of raw sequencing data was checked using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Next, the data were analyzed using Stacks (version 2.41) (CATCHEN *et al.,* 2013) for SNPs calling. Briefly, sequences were demultiplexed and filtered using *process_radtags* and individual reads that passed the previous quality filters were aligned to the chromosome-level reference-genome of *M. macrocephalus*. Subsequently, *gstacks* created loci by incorporating the ddRAD aligned reads. Finally, *populations* was used to generate genotype data for the samples. To differentiate putative SNPs from sequencing errors, we used Plink 1.9 (PURCELL *et al.,* 2007) to filter spurious SNPs with more than 10% genotyping error rate (--geno 0.1), minor allele frequencies less than 0.05 (--min-maf 0.05), and Hardy-Weinberg imbalance (p < 5E10-5). Regarding the removal of individuals, samples that had more than 15% (--mind 0.15) of absent genotypes were excluded.

Linkage map

A linkage map was created using Lep-MAP3 (RASTAS, 2017). First, a parenthood test was performed using the *IBD* module, and individuals with more than 10% of Mendelian errors were

removed. The *ParentCall2* module was used to impute possible missing genotypes or to correct erroneous parental genotypes based on progeny data. *Filtering2* module was used to remove markers with significant segregation distortion (dataTolerance = 0.001) and non-informative markers. Markers were assigned to linkage groups (LG) by *SeparateChromosomes2* using the minimum logarithm of odds (LOD) score. The best LOD was selected iteratively and accounted for markers distribution in the first 27 linkage groups, which corresponds to the haploid chromosome number of the species. Next, orphan markers were assigned to existing linkage groups (LOD score lower than in *SeparateChromosomes2*) using *JoinSingles2* and ordered within each linkage group using the *OrderMarkers2* module. Due to the slight stochastic variation in marker distances between runs, the *OrderMarkers2* module was run 15 times and the order with the best likelihood value for each LG was selected. The linkage map graphic representation was constructed using R/*LinkageMapView* .

The reliability of the SNP *loci* attribution on the LG and the respective *loci* ordering within the LGs was verified through comparative genomic synteny analysis with the reference genome using *Circa* (http://omgenomics.com/circa/).

We used the genome scaffolds to generate other chromosome-level genome using the linkage map as a reference in Chromonomer v1.13(CATCHEN; AMORES; BASSHAM, 2020). This was done to verify possible differences between the linkage map ordering (genetic mapping) and the Hi-C ordering (physical mapping). Chromonomer (CATCHEN; AMORES; BASSHAM, 2020) attempts to find the best set of nonconflicting markers that maximizes the number of scaffolds in the resulting genome while minimizing ordering discrepancies. It resulted in a FASTA file (chromonome.fa), the chromosome-level genome oriented according to the genetic map.

8.4.10 Resequencing (pool-sequencing)

DNA was extracted individually and quantified according to the 8.5.9.2 8.4.9 Linkage mapping section. Library construction and sequencing were performed at Laboratoire de Physiologie et Génomique des Poissons (LPGP) of the National Research Institute for Agriculture, Food and Environment (INRAE), France, using an Illumina NovaSeq platform 150 bp PE.

The Pool-Seq dataset was analyzed with the Pooled Sequencing Analysis for Sex Signal (PSASS) pipeline (https://doi.org/10.5281/zenodo.2615936). Briefly, reads from the male and female pools were mapped into the female pseudo-haplotype chromosome-level genome (GCA_021613375.1) using bwa-mem (LI; DURBIN, 2010) with default parameters. As *L. friderici* does not have a published reference genome, we also used the *M. macrocephalus* female genome (GCA_021613375.1). Then, the alignment files were sorted, merged and PCR duplicates were removed with Picard tools (https://github.com/broadinstitute/picard). Reads with mapping quality < 20 and that not mapped uniquely were also removed with samtools (LI *et al.*, 2009a). Next, the two sex BAM files were used to generate a pileup file using samtools mpileup (LI *et al.*, 2009b) with per-base alignment quality disabled (−B). A sync file was created using popoolation mpileup2sync (H. Li et al., 2009)parameters: --min-qual 20), which contained the nucleotide composition of each sex for each position in the reference genome. With this sync file, fixation index (F_{ST}), SNPs, and coverage between the two sexes in all reference positions were calculated in a 50 kb sliding window with an output point every 1,000 bp to identify sex-specific SNPs enriched regions. The plots were constructed with the R packages sgtr (https://github.com/SexGenomicsToolkit/sgtr) and karyoploteR (https://bernatgel.github.io/karyoploter_tutorial/).

8.4.11 RNA-seq

RNA was extracted from each pool with RNeasy Micro Kit (Qiagen). Next, the integrity (RIN > 7) and concentration (ng/µl) were accessed using Bioanalyzer 2100 (Agilent). At last, library construction and sequencing were performed by BGI Genomics (Shenzhen, China) using BGISEQ-500 platform 100 bp PE.

Raw read quality was accessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapters and poor quality reads were trimmed in Trimmomatic (BOLGER; LOHSE; USADEL, 2014) (parameters LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:36). Trimmed reads were pseudo aligned against mRNA sequences obtained from *M. macrocephalus* genome (GCA_021613375.1) with kallisto (BRAY *et al.*, 2016). A matrix with estimated counts of transcripts abundance was exported with R/tximport v1.10.1 (SONESON *et al.*, 2016). Differential expression analysis was performed with R/DESeq2 (LOVE; HUBER; ANDERS, 2014) and transcripts with False Discovery Rate (FDR) adjusted *p*-values ≤ 0.05 were considered differentially expressed. Transcripts with Log Fold Change (LFC) ≥ 2 were considered as up-regulated in males and down-regulated in females and transcripts with LFC ≤ -2 were down-regulated in males and up-regulated in females (DOMINGUES *et al.*, 2020).

8.4.12 Results

8.4.12.1 Genome Sequencing

Genome Assembly

We generated 88.8 Gb of PacBio continuous long reads (CLR), 85 Gb of MGISEQ short reads, and 105 Gb of Hi-C data. Genome coverage based on final assembly size was 69.4x, 66.4x, and 82x, respectively. The unique molecular yield of PacBio reads was 56 Gb and the subread N50 length was 32 kb. Regarding the MGISEQ reads, after the removal of poor-quality sequences, we kept 82 Gb of clean data. This dataset was used to generate *k*-mer spectrum plots to estimate the overall characteristics of the genome. All *k*-mer plots were similar and showed a profile correspondent with a low heterozygosity rate (Figure 2). The estimated genome size (21-mer) was 1.04 Gb with heterozygosity of 0.48% and 18% of repeat content.

PacBio long reads were assembled into 2,770 primary contigs (33 contigs > 5 Mb) with N50 of 1.53 Mb. After gap filling, the initial contigs were clustered in 1,227 scaffolds with N50 of 5.0 Mb. The scaffolds were ordered and oriented into 27 chromosomes (Figure 3), which is consistent with the haploid chromosome number of the species (PORTO-FORESTI *et al.*, 2008); and 73 unplaced scaffolds (< 250 kb). The 27 chromosomes comprised 99.56 % of genome assembly. The final *M. macrocephalus* pseudo-haploid genome contains the 27 expected chromosomes, 73 unplaced scaffolds, and the mitogenome. It has contig and scaffold N50 of 5.0 and 45.03 Mb, respectively, and 1.28 Gb (Table 2. Statistics for genome assembly of Megaleporinus macrocephalus.).

We used the highly-accurate MGISEQ reads to plot Merqury (RHIE *et al.*, 2020c) evaluation against the genome *k*-mers. Figure 2B, shows that (i) the distribution of the *k*-mers in the assembly is correspondent to the short read profile (Figure 2A), (ii) there are two peaks demonstrating that 1-copy (heterozygous) and 2-copy (homozygous) *k*-mers were found once in the assembly, as expected for a pseudo-haplotype genome (RHIE *et al.*, 2020c), (iii) most of the assembly *k*-mers

(in red) are unique, indicating that the assembly has low content of artificial duplications (*i.e.*, *k*-mers found twice, in blue) (iv) there are missing *k*-mers in the assembly (black peak), which is compatible with haploid genomes, (v) the 1-copy *k*-mer peak (red) is greater than its missing sequences (black), this suggests that FALCON-Unzip erroneously included sequences from both haplotypes into the primary pseudo-haplotype (RHIE *et al.*, 2020c). Also, this possibly led to an assembled genome size greater the estimated (1.04 Gb). The accuracy of the base calls (QV), which is calculated using the *k*-mers found only in the assembly (bar at the beginning of Figure 2B), was 37. 53 (Table 2) and represents a base accuracy > 99.9% (*e.g.*, QV = 30 means 99.9% accuracy). The completeness score shows that 93.05% of *k*-mers in the MGISEQ reads are present in the assembly, which is a good recovery of *k*-mers for a species with 0.5% heterozygosity.

We identified the chromosome 13 as the sex chromosome of *M. macrocephalus.* That was possible through the difference in coverage between the sex chromosome and the autosomes on the Hi-C contact map (Figure 3). The sex chromosome upper part has less coverage than its end and the autosomes. Our hypothesis is that this part of the sex chromosome corresponds to the specific region of the W chromosome. Therefore, the final part of the sex chromosome, that has normal coverage, would correspond to the pseudo-autosomal region of Z and W.

The Pearson's correlation between the autosomes assembled size with its actual karyotypic size (Table 4. Assembled and estimated chromosomes sizes (bp) calculated using karyotype data.) was 99%, demonstrating the high quality of the assembled *M. macrocephalus* genome. We also measured Z and W chromosomes, including the heterochromatic and euchromatic regions (Table 5 and Figure 4). The estimated size of Z and W were approximately 50 and 71 Mb, respectively. Considering the assembled sex chromosome was ~ 45 Mb, we can verify that some sequences are missing.

Repeat Annotation

Using the *de novo* prediction model, 2,544 new families of repeats were found in the genome. Along with the homology-based prediction, the repeat content found in *M. macrocephalus* accounted for 46.71% of the genome (598 Mb). Among the repeats, transposable elements (TEs) were the most common representing 37.49% of the genome. DNA transposons were the most abundant TE (11.82%), following 3.02% of long terminal repeats (LTR), 3.42% of long interspersed nuclear elements (LINE), and 0.33% of short interspersed nuclear elements (SINE) (Table 6). A great percentage (18.89%) of the interspersed repeats remained unclassified. Satellite DNA and simple repeats accounted for 4.40% and 4.06% of the genome, respectively (Tables 6 and 7).

Figure 5 shows older TE copies located on the right side of the graph and rather recent ones, that not diverged much from the consensus TE sequence, on the left side. Most of the interspersed repeat content found in the *M. macrocephalus* genome is recent (*K*-values < 25). Also, it is possible to observe two bursts of transposition dominated by DNA transposon. Interestingly, we found a W-specific satellite MmaSat97 (UTSUNOMIA *et al.*, 2019b) on chromosome 13, which corroborates its status as the sex chromosome of *M. macrocephalus*. The repeat content found in the sex chromosome was slightly higher than in the autosomes (4.24%). Total interspersed repeats and satellites were the classes that presented the major difference, 2.37% and 2.24%, respectively (Table *7*).

Gene Prediction and Annotation

For *ab initio* gene prediction, BRAKER1 (HOFF *et al.*, 2016b) used 28.26 Gb of RNA-seq data as extrinsic evidence to predicted 60,482 genes. For homology-based gene prediction,

BRAKER2 (BRŮNA *et al.*, 2021b) generated 57,574 hints and predicted genes. TSEBRA (GABRIEL *et al.*, 2021b) combined BRAKER runs and selected the best predictions, totalizing 44,054 gene predictions. We performed quality control of the selected predicted genes. Of all the 44,054 predicted genes, 66.94% (29,490) were annotated in the Actinopterygii database of Eggnog or UniProtKB (BATEMAN *et al.*, 2021b); and 33.06% (13, 525) were not annotated. We kept the annotated (29,490) and the non-annotated predicted genes with more than 150 aa (1,039) for the final dataset, summarizing 30,501 protein-coding predicted genes (Table 8). The final dataset had 94.1% complete BUSCOs, 89% complete and single-copy, 5.1% duplicated, 2.2% fragmented, and 3.7% missing BUSCOs. For the functional annotation, we performed blast searches against the Actinopterygii database of UniProtKB (BATEMAN *et al.*, 2021b). Of all the predicted genes, only 3.34% (1,018) were not annotated.

The most representative gene onthology (GO) terms (> 15% of genes, Figure 6) according to the three-domain were the following: cell, cell part, membrane, membrane part, organelle, organelle part, protein-containing complex (Cellular Component); binding, catalytic activity (Molecular Function); cellular process, metabolic process, biological regulation, regulation of biological process, response to stimulus, developmental process, multicellular organismal process, signaling, localization, cellular component organization or biogenesis (Biological Process).

Comparative Phylogenomics

Of all the genes (406,287) from the 14 species, 95.1% (386,569) were assigned to 24,971 orthogroups. We found 11,105 orthogroups with all the species present, of which 336 were single-copy. The single-copy orthologs dataset was used to generate a phylogeny (Figure 7) using the Maximum Likelihood criteria and 1,000 bootstrap replicates.

8.4.12.2 Genetic map

In total, 1,307,500,332 raw reads were sequenced by ddRADseq, resulting in approximately 200 Gb of data (approximately 28 Gb per library). After filtering (removal of low-quality sequences and reads with missing or ambiguous barcodes), an average of 11% of the reads were removed from each library, *i.e.*, 89% of the reads were retained for analysis. Furthermore, 24 individuals were excluded due to the low number of reads (< 1 M). The average number of reads per sample was 4.3 M. Raw sequencing data and filtered reads for each library are shown in Table 9.

After mapping the ddRAD reads to the chromosome-level genome, Stacks analysis (CATCHEN *et al.*, 2011) resulted in 41,033 SNPs from 85,167 loci for 281 individuals. In Plink (PURCELL *et al.*, 2007), after applying the mind and geno filters, 56 individuals and 8,971 SNPs were excluded. At last, the maf filter excluded 3,733 SNPs. Thus, 225 individuals and 28,329 SNPs passed all quality controls (total genotyping rate of 0.96) and were used for the linkage mapping.

After calling possible missing or erroneous parental genotypes in the *ParentalCall* module, a total of 9,997 SNPs were joined to linkage groups (LGs). We computed several LOD scores between markers and selected the best marker distribution according to the karyotype characteristics of the species. Although the *M. macrocephalus* haploid chromosome number is 27, we did not detect a specific LOD resulting in 27 LGs. Thus, LOD 12 was applied because it resulted in 28 LGs, (considering that LOD 11 or 13 totalized in 26 LGs, Figure 8). The remaining markers were assigned to the existing LGs using LOD 10, which recovered 1,234 markers. 18,098 markers were discarded because no association with the linkage map was detected. For each LG, the orders of the markers with the best likelihood were combined to produce the final linkage map.

A total of 11,231 SNPs were assigned to 28 LGs (Figure 8). We constructed a male, a female, and a sex-averaged map (average position between male and female map).

The number of SNPs in the LGs varied from 710 (LG1) to 203 (LG28). In the sex-averaged map, LGs length ranged from 143.08 (LG22) to 43.25 (LG24) cM, with an average of 3320.36 cM and an average distance between markers of 0.29 cM (SD = 0.12). The highest and lowest marker densities were found on LG1 and LG22, with an average of 0.18 and 0.61 cM, respectively (Table 11).

Concerning sex-specific differences, the average distance between markers in male and female maps were 0.31 and 0.29, respectively. Therefore, the male map (3,518.24 cM) was longer than the female (3,301.97 cM). The male:female genetic length ratio over the entire genome was 1.07. The ratios varied from 0.54 (LG24) to 1.37 (LG12). The highest density of recombination was detected at the proximal region of the centromeres (considering that this species has metacentric/submetacentric chromosomes), although some exceptions occurred in the terminal region of the LG11 (Figure 9).

In LG24, recombination was distributed differently between the sexes (heterochiasmy). In the female map, LG24 was almost double the size (87.81 cM) of the same LG in the male map (47.37 cM). The comparative synteny analysis revealed a correspondence of LG24 with the sex chromosome 13 (Figure 11), particularly in regions < 20 Mb and > 40 Mb. Also, zero recombination clusters (suppression of recombination) were observed in this LG for both the sexes (Figure 10), *i.e.*, blocks of markers that vary in the physical distance (bp), but do not vary in the genetic distance (cM). Besides, in the genomic synteny between LG24 and the sex chromosome 13 (Figure 8 and 9), the same chromosome was also attributed to LG27, probably in the pseudo-autosomal region (PAR). This indicates that the best LOD value resulted in 28 LGs (n = 27

chromosomes) because both LG24 and LG27 corresponded to the same chromosome (different regions of the sex chromosome).

After ordering the genome scaffolds using the linkage map as a reference, Chromonomer resulted a final chromosome-level genome with 1,575 markers grouped in 352 scaffolds. The remaining SNP loci were not used for genome anchoring because they were not aligned to the piauçu scaffolds, represented markers mapping to multiple regions, or represented loci where the orientation could not be suitably assigned. These results allowed the construction of a chromosome-level genome that clustered 75% (1,221,855,406 bp) of the initial scaffold data into 27 pseudomolecules (chromosomes) totalizing 977 Mb of length. 320 Mb were not anchored in pseudomolecules (chromosomes). LG24 anchored a low number of scaffolds, resulting in a pseudomolecule with poor scaffolding and small size (~ 4 Mb). This can be explained by the region of sex conflict between the Z and W chromosomes and, consequently, the suppression of recombination between them.

The dotplot synteny analysis demonstrated a high degree of concordance between the chromosomes scaffolded with Hi-C data (physical mapping) and the linkage groups of the genetic map (Figure 12, illustrated by chromosomes 5, 8, and 20). Insertions and deletions were observed in all chromosomes (*e.g.*, chromosome 2, since 25% of the initial scaffold data was missing). Beyond that, structural differences between the linkage map and scaffolds were noted in some chromosomes, revealing relocations (chromosomes 1 and 21) and major inversions (chromosomes 3, 18, and 19).

8.4.12.3 Resequencing (pool-sequencing)

We sequenced 20 males and 20 female individuals of *M. macrocephalus* with an average depth of ~28-fold. Pool-sequencing raw data are summarized in Table 12**.**

Reads of *M. macrocephalus* male and female pools were mapped to the chromosome-level female genome to characterize genomic regions enriched for sex-biased signals. *i.e.*, sex coverage differences or sex-biased SNPs. Whole-genome analysis of SNPs distribution **(**Figure 13) revealed a strong sex-linked signal in females on chromosome 13. In addition, high F_{ST} values (Figure 13) were observed on the same chromosome. This profile illustrates a female heterogametic system (ZW/ZZ), as previously reported by cytogenetic data (GALETTI, JR. *et al.*, 1981).

The genomic regions with a high density of female-specific SNPs and high F_{ST} values (Figure 13) are sex conflict regions that could contain the *M. macrocephalus* master sex-determining gene. Chromosome 13 is the only one with these characteristics, and, therefore, is the sex chromosome (Figure 13). The Hi-C contact map pattern, the differences between the sexes observed in LG24, and the repeat annotation of the *M. macrocephalus* genome corroborate its status.

The assembled sex chromosome of *M. macrocephalus* is approximately 45 Mb in length and could be divided into three distinct regions according to the pattern of sex-specific SNPs, F_{ST} and read coverage. The first region comprises the beginning of the chromosome, from 0 to ~ 3 Mb and it is characterized by high F_{ST} values (> 0.04). Notably, the major F_{ST} peak (F_{ST} = 0.17) is located within this region (Figure 14). The plots of depth ratio and absolute depth as well as the visualization of the read alignments showed high coverage in females (> 200 reads; major peak on position 2,404,000 bp with 689 read depth), and low coverage in males (depth ratio < 1, Figure 15). The absence of coverage in males was detected in some areas. The patterns observed here suggest that chromosome W-specific sequences were assembled in this region, which was corroborated by the recombination suppression in the LG24, particularly up to 3 Mb. Hence, we denominated this region as putative W-specific.

The second region comprises ~3 to 20 Mb and the opposite terminal region of the sex chromosome (~ 44 to 45 Mb Figure 13 and Figure 14). This area is characterized by a high density of females-specific SNPs (W pool reads aligning to the Z reference sequence, and *vice-versa*, although in lower amount), with peaks summarizing more than 2,000 SNPs. The depth ratio peaks were constantly more superior in males (1>) than in females (< 1) (Figure 14). Because of these characteristics, we denominated this region as chimeric Z and W, where there is some degree of homology between the sequences of each chromosome. This probably allowed that Z-specific or W-specific sequences were scaffolded at the same region, comprising a chimeric genomic sequence (Figure 16); although it is composed mostly by Z-specific sequences, which was corroborated by the occurrence of higher recombination frequency in males than females in this region of the LG24 (3 to 20 Mb, and 44 to 45 Mb, as demonstrated by the Figure 10).

The last region comprises from ~20 to 44 Mb and is characterized by a lack of sex-specific SNPs (Figure 14). In this genomic locus is also seen an almost equal absolute depth between males and females (depth ratio ≈ 1, Figure 15). This points out homology between the male and female sequences in this zone and that could indicate normal recombination rates as seen in pseudo-autosomal regions (PAR) of sex chromosomes (LG27). Therefore, we denominated this locus as PAR.

Concerning the pool-sequencing of *L. friderici* (Figure 17), the SNPs were equally distributed throughout the reference genome of *M. macrocephalus*, without any sex-specific region on chromosome 13, nor in any other LG. We also observed isolated points of high F_{ST} values along the genome (highest $F_{ST} = 0.12$).

Although the sex locus was not found in *L. friderici*, our results indicated a different sex system in genomic level between the species.

8.4.12.34 RNA-seq

A total amount of 28.26 Gb of gonadal paired-end RNA-seq data were pseudo-aligned with 30,500 transcripts of *M. macrocephalus*. Approximately 99.9% (30,460) of RNA-Seq transcripts were successfully pseudo-aligned, and, after low counts were filtered (≤ 1), 27,120 transcripts remained for the differential expression analysis of males and females. Principal component analysis (PCA) of the differentially expressed transcripts showed that 78% of the variance in the data was explained by Principal component 1 (PC1). Throughout PC1 the data were clustered in two groups, males ZZ and females ZW, as expected (Figure 18A). The heatmap of the Euclidean distances matrix between samples demonstrated the same pattern (Figure 18B).

The analysis resulted in 9,740 differentially expressed transcripts (*p*adj ≤ 0.05); of which 39.83% (3,879) were up-regulated in males; and 27.55% (2,683) up-regulated in females (Figure 18C and Table 14). Most of the transcripts were not significantly expressed (64. 09 %). We found 3.19% (311) of the differentially expressed transcripts of *M. macrocephalus* on the sex chromosome. Of these, 170 were up-regulated in males (LFC ≥ 2) and 58 in females (LFC ≤ -2) (Table 14). In addition, the sex chromosome presented a lower number of differentially expressed transcripts than the average per chromosome (311 *versus* 360, Table 14). In general, there were more up-regulated transcripts in males than in females (> 12.28%;Table 14).

Figure 19 is illustrating the transcript counts of males and females throughout the sex chromosome. Transcript expression occurred mainly in females at the putative W-specific region. Regarding the chimeric region, the higher expression rates observed in males in this area could be explained by a higher concentration of Z-specific sequences, which are duplicated in ZZ. In

general, a similar expression between the sexes in the pseudo-autosomal region was observed. Awkwardly, a few peaks had higher expression in females than in males in this area.

To identify sex determination-related genes (SDG), we searched on the sex chromosome of *M. macrocephalus* for differentially expressed genes already reported in the sex determination pathways of fish (Pan et al., 2021;Table 15). We also considered two hypothesis (Figure 20): 1) SDG is located in a putative Z-specific locus (chimeric region); then the sex determination would be based on a sex chromosome gene dosage mechanism; 2) SDG is located in a putative W-specific locus; then the sex determination would be based on the presence/absence of the gene.

Most of the identified genes were up-regulated in males (LFC ≥ 2), except BMP7 and CCDC114 (Table 15). CCDC114 was the only gene in the W-specific region, the most suitable location to find the SDG. IGFBP6 was within the chimeric region, a potential SDG-containing area.

8.4.13 Discussion

8.4.13.1 Genome

This is the first sex chromosome assembly of a neotropical fish species. To date, despite several fish Y chromosome assemblies have been reported, *e.g.*, stickleback (PEICHEL *et al.*, 2020a), zig-zag eel (XUE *et al.*, 2021b), spotted knifejaw (M. Li et al., 2021) and Atlantic herring (RAFATI *et al.*, 2020); there are only a few reports of fish W chromosome assemblies, *e.g.*, half-smooth tongue sole (CHEN *et al.*, 2014) and mosquitofish (SHAO *et al.*, 2020). Yet, the former was assembled with short-reads (which is not recommended for sex chromosomes assemblies, (PEICHEL et al., 2020) and the latter did not present details about the W content and evolution.

Piauçu sex chromosome assembly provides information about a W heterochromatic chromosome in an intermediate stage of evolution (CHARLESWORTH, 2019) and represents a step advance in the understanding of the evolution of sex chromosomes in fish and vertebrates.

The *M. macrocephalus* genome size was similar to those of other Neotropical fish and had contiguity metrics comparable to chromosome-level assemblies of the group, such as the *Astyanax mexicanus* (GCA_000372685.2) and *Pygocentrus nattereri* (GCA_015220715.1) reference genomes (Table *3*)

As some parts of the assembled sex chromosome are a mix of Z and W sequences, we estimated the true size that unique Z and W assemblies should have (Table *5*). Moreover, we correlated the heterochromatic and euchromatic regions of Z and W to the poolsex-determined regions of the assembled sex chromosome (Table *5*) The euchromatin region probably corresponds to the PAR, which is shared between both Z and W; the W- specific and chimeric regions correspond to the heterochromatic region of W and most of the chimeric region corresponds to the heterochromatic region of Z (the majority of the sequences in the chimeric region are Z-specific, despite not being possible to estimate the exact amount, Figure 15).

Due to the lack of divergence between the sexes in the pseudo-autosomal region (PAR) (almost equal coverage, no sex-specific SNPs, and no high FST values), we assumed that this region was entirely assembled. Its estimated size was 20.68 Mb on Z and W, close to the assembled size (24 Mb). PAR possibly occupies 33.70% and 47.42% of W and Z, respectively, which means that 66,3% (47.2 Mb) of W is constituted by W-specific sequences while Z is constituted by 52,58% (26.6 Mb) of Z-specific sequences (calculated using PAR = 24 Mb). Considering that the whole assembled sex chromosome size was 45 Mb; there were missing approximately 8.6 Mb of Z sequences (Z = PAR + chimeric) and 26.2 Mb of W (W = PAR + chimeric + W-specific).

The repeat content found (46.71%) in the piauçu genome was intermediate to what has been reported in other Neotropical fish species, such as *C. macropomum* (52.49%) (HILSDORF *et al.*, 2021b) and *A. mexicanus* (41%) (WARREN *et al.*, 2021b). DNA transposons were the most abundant type of TE (11.82%), corroborating what has been observed in teleost fish (GAO *et al.*, 2016). A great percentage (18.89%) of the interspersed repeats remained unclassified. This was also reported in tambaqui (HILSDORF *et al.*, 2021b) and red-bellied piranha (SCHARTL *et al.*, 2019b) genomes (39.15% and 28.3% of unclassified sequences, respectively). In the piauçu repeat landscape, (Figure 5), it is possible to observe two bursts of transposition dominated by DNA transposon, as reported in other teleost fish *e.g.*, Nile tilapia (CHALOPIN et al., n.d.).

Satellite sequences are remarkably rich throughout the genome of *M. macrocephalus* (13.47% female and 11.99% male) and even more in the sex chromosomes, especially the W (UTSUNOMIA *et al.*, 2019). In the assembly, repeat elements had a slightly higher abundance in the sex chromosome compared to the autosomes. However, satellites accounted for only 4.40 percent of the repeats found in the repeat annotation. Thus, due to the large number of satellites sequences found in the sex chromosomes of piauçu (UTSUNOMIA *et al.*, 2019) and the presence of the large blocks of heterochromatin in W/Z; we assume that these satDNAs constitute the missing sequences of the W and Z chromosome.

The genome annotation resulted in 30,501 protein-coding predicted genes (Table *8*), which is consistent with other related neotropical fish genomes such as tambaqui and red-bellied piranha (31,149 and 30,575, respectively), but differently of was found in cavefish (248,235; (WARREN et al., 2021). In the phylogeny analysis performed with single-copy orthologs genes (Figure 7), *M. macrocephalus* was placed in a clade with other Characoidei (MELO *et al.*, 2021) species (*A.*

mexicanus, *P. nattereri*, and *C. macropomum*). The tree topology was consistent with previous studies (HUGHES et al., 2018; MELO et al., 2021; STEINKE et al., 2006).

8.4.13.2 Genetic map

In this study, we developed the first dense genetic map available so far for piauçu. We obtained a similar resolution of other linkage maps constructed for related Neotropical fish species that present analogous karyotype characteristics (haploid chromosome number, morphology, and size) (Table 10).

Interestingly, we detected an additional linkage group, differently from what was expected (27 linkage groups as the haploid chromosome number). This extra linkage group (LG24) was associated with the sex chromosome in the synteny analysis. This pattern was previously reported in the butterfly *Melitaea cinxia* (RASTAS *et al.*, 2013), which also presents a ZW sex chromosome system . The additional LG (LG24) was composed of markers that followed a Z chromosomal inheritance, *i.e.*, female offspring are homozygous of one of the fathers´s alleles. This explains the strong heterochiasmy in this LG, in which higher recombination was observed in the male map. Otherwise, although being separated by the LG24 linkage pattern, LG27 is physically merged (in the genome) to this LG and corresponds to the pseudo-autosomal region of the sex chromosome *i.e.*, homologous regions between the Z and W chromosomes, as similarly observed in *Melitaea cinxia* (RASTAS *et al.*, 2013).

An interesting sex-specific difference was also revealed by the linkage mapping analysis in piauçu, the male map was longer than the female, with a genetic-length ratio of 1.07. Recently, the tambaqui *C. macropomum* was described as having an XY hypothetical sex determination system (VARELA *et al.*, 2021), despite not presenting heteromorphic sex chromosomes. Interesting, the tambaqui female linkage map was larger than the male (1.55x) (Ariede et al., in press). Differences

in map length can result from a variation in the number of recombination events in the two parents as well as variations in the number and location of the mapped loci. It is common to find a difference in the recombination ratio between the two sexes in most aquatic species. For instance, the male/ female recombination ratios are 1:8.26 in Atlantic salmon, 1:3.25 in rainbow trout, 1:1.43 in Japanese flounder and 1:2 in halibut . Despite this being a common phenomenon, the mechanism responsible for the different recombination rates between the genders is still not well understood (SONG *et al.*, 2012). This explains the opposite sex-specific differences observed between piauçu and tambaqui and suggests that the heterogametic sex presents smaller maps due to recombination suppression (REID *et al.*, 2007). The influence of the sex-determination system in the sex-specific recombination patterns was also described for other fish lineages. In flatfish, turbot (BOUZA *et al.*, 2007), Senegalese sole (GUERRERO-CÓZAR *et al.*, 2020), and Atlantic halibut (REID *et al.*, 2007) female maps were larger (1.36, 1.32, and 1.07 times); while in the flounder *Paralichthys olivaceus* (CASTAÑO-SÁNCHEZ *et al.*, 2010) and tongue sole *Cynoglossus semilaevis* (SONG *et al.*, 2012), the male maps were slightly larger (1.03 and 1.09 times, respectively).

In addition, our linkage map was successfully used as a reference to anchor the genome scaffolds into a chromosome scale, evidencing its high quality. The chromosome-scale genome anchored using the linkage map presented a high correspondence with the chromosome-level genome (scaffolded with Hi-C physical mapping). These results were similar to those obtained by other chromosome-level genomes anchored with linkage maps, such as in *Astyanax mexicanus* (WARREN *et al.*, 2021a) and *Sander lucioperca* (DE LOS RÍOS-PÉREZ *et al.*, 2020).

The inconsistencies revealed by structural differences (relocations and inversions) between the linkage groups and chromosomes ordered by Hi-C (physical mapping) were also reported in the Lake Trout (*Salvelinus namaycush*) (SMITH *et al.*, 2022) and probably will need further

investigation using other techniques, such as physical mapping of specific DNA into chromosome spreads with FISH (Fluorescence *in situ* hybridization).

8.4.13.3Pool-seq

Different rates of recombination suppression are often observed within sex chromosomes. As in zig zal eel (LIU et al., 2021); stickleback, (PEICHEL et al., 2020) and human, (LAHN & PAGE, 1999). These events are denominated as evolutionary *strata* and can be first identified by sequence divergence (evidentiated by high FST values) and a high concentration of sex-specific SNPs (CHARLESWORTH, 2021). Based on the whole genome sequences of males and females of piauçu, we could discriminate distinct regions in the sex chromosome assembly that comprehend different evolutionary *strata*.

The W chromosome is composed of two evolutionary *strata* concerning the W-specific region (high FST, evidence of recombination suppression, lower depth ratio in males) and the chimeric region (high concentration of female-specific SNPs). The former is in a more advanced stage of recombination suppression, while the latter is in a younger state due to partial similarity with Z (CHARLESWORTH, 2021). The younger *stratum* is located closest to the PAR while the older is further away. Hence, we can suppose that W/Z differentiation started first in this region (CHARLESWORTH; CHARLESWORTH; MARAIS, 2005). The comparison of TE density between the evolutionary *strata* can corroborate which one suffered recombination suppression first (PEICHEL *et al.*, 2020b). The true extent of the strata could not be delimited with our data because of the missing sequences in the W, but according to the size of the heterochromatin blocks of the cytogenetic data, they should comprehend together ~ 50.2 Mb. The missing satellite sequences probably constitute the oldest evolutionary *stratum* of W, due to its degenerative nature.

The chimeric region is also a different evolutionary *stratum* within Z. The presence of female-specific-SNPs and a few peaks of high FST corroborate this hypothesis (CHARLESWORTH, 2021), and indicates a young stratum composed of sequences that still carry some similarities with W. The presence of heterochromatin in the long arm of the Z chromosome, although in a lower amount than in the W, suggests that its missing sequences could also comprehend older evolutionary *strata*.

The *M. macrocephalus* sex chromosomes are in an intermediate stage of degeneration (CHARLESWORTH, 2019) and arose at least 12 MYA, after the divergence of *Megaleporinus* and *Leporinus*, as reported by phylogenetic analysis (RAMIREZ; BIRINDELLI; GALETTI, 2017). Nonetheless, it is necessary to quantify the ages of *strata* in nº. of generations, which is a more appropriate scale of theoretical modelling of degeneration (CHARLESWORTH, 2021). Sequence divergence estimates (*K*s values for synonymous sites in coding sequences) should be used for this purpose (CHARLESWORTH, 2021).

We identified regions with different patterns in the sex chromosome of piauçu and estimated their sizes: the PAR that spans ~ 24 Mb; at least two evolutionary *strata* in W, an older *stratum* of ~ 26.6 Mb and a younger *stratum* of ~20.6 Mb; and one *stratum* (young) in Z of ~31.6 Mb. To delineate the total number of s*trata* and the true extension of them and especially to quantify degeneration (the extent to which genes within the region have lost their functions), completely phased sequences of Z and W are necessary (CHARLESWORTH, 2021).

8.4.13.4 RNAseq

To date, more than 20 SDG genes have been identified on different sex chromosomes of mammals, birds, frogs, and fish (PAN *et al.*, 2021). In mammals and other groups, the sex master genes are conserved while in fish, a lot of different genes have been already reported (Dmrt1,

Amhb1, SdY, Hsd17b1, AmhY, Gdf6-Y). Despite the rapid evolution of different sex-determining genes, only a limited group of factors/signaling pathways keep showing as master genes, such as transcription factors (Dmrt1-or Sox3- related) and others belonging to the TGF-β signaling pathway (Amh, Amhr2, Bmpr1b, Gsdf, and Gdf6) (PAN *et al.*, 2021).

The sex determination processes have two main vias, it can be due to the presence of a sex-determining gene in one sex and absence in the other; or through dosage compensation, where different expressions of the sex-determining gene differentiate the sexes (in this case, half of the male expression in females).

If sex-determination is due to presence/absence, the SDG would probably be in the W-specific region/older *stratum*, which theoretically suffered recombination suppression first (CHARLESWORTH, 2019). However, we could not find any previously sex-determination-associated gene in this region. Due to a large amount of W (~26.2 Mb) and Z (~8.6 Mb) missing sequences, the SDG may be missing from the assembly.

One of the classic genes related to sex determination (PEICHEL *et al.*, 2020b), the Amhr2 anti-Mullerian hormone gene, was localized in the sex chromosome of piauçu presenting high male up-regulated expression. However, due to its position inside the PAR (on 38.82 Mb), it was not elevated as a candidate gene in piauçu.

The coiled-coil domain containing 114 gene (CCDC114) is located on a peak of female expression within the W-specific region and because of its position in the older *stratum*, it has been considered here as a sex-determination candidate gene. The coiled-coil domain containing proteins are implicated in male gametogenesis, which is evidentiated by their stage-specific expression in the gametes, their tissue-/cellular-specific localization in the testis, and their altered levels in infertile men (PRIYANKA; YENUGU, 2021). In zebrafish, genes encoding coiled-coil domain

proteins (CCDC39, CCDC103, CCDC114, and CCDC151) were down-regulated in mutant fish without the anti-Mullerian hormone gene (YAN *et al.,* 2019a). The mutant fish failed to produce mature sperm with fully developed tails, which likely contributed to the observed loss of fertility as animals aged (YAN *et al.*, 2019b).

The insulin-like growth factor binding protein gene (*Igfbp-6*) was the only within the chimeric region, a potential SDG-containing area, despite not being the main region we would look for (CHARLESWORTH, 2019).

The insulin-like growth factors (IGFs), transforming growth factor-βs (TGFβs), and bone morphogenetic proteins (BMPs) are three growth factor systems on bone formation. Each growth factor family consists of multiple related growth factor genes. IGFs, TGFβs, and BMPs are produced by osteoblasts and other bone cells and affect osteoblast proliferation and differentiation (LINKHART; MOHAN; BAYLINK, 1996).

The IGF-binding proteins regulate IGF-I and IGF-II action. A tandem duplication followed by whole genome duplications in the ancestral of teleost fish, led to the retention of duplicated copies (paralogs) of IGFBP (DE LA SERRANA; MACQUEEN, 2018a), resulting in at least six copies in the lineage. IGFBP-6 regulates processes where IGF-II is involved, such as proliferation, survival, migration, and differentiation. In teleost fish, IGFBP-6 is rather understudied, and there are no conclusions about its roles and functions (DE LA SERRANA; MACQUEEN, 2018b).

IGFBP-6 of piauçu is supposedly located in the Z-specific region because of its position in the chimeric region in addition to male up-regulation. In this case, the most likely scenario is that differential expression of this gene between males and females would determine sex (dosage compensation). In this scenario, the females (that present one copy of Z) should have half of the male expression (two copies of Z), which did not occur. However, the pattern of male up-regulation

in the sex chromosome could indicate incomplete Z chromosome dosage compensation (WRIGHT *et al.*, 2016). Moreover, our RNAseq data was composed of pools of gonads from different individuals. Although the data were distributed as expected (female ZW x male ZZ), a degree of variation within the male expression was observed (Figure 18). This means that males in different developmental stages could have been mixed, which could be masking the true expression of IGFBP-6. Males with different levels of gonadal differentiation were also verified during the testis histology.

RNA-seq experiments are useful to evidentiate a landscape of gene expression in a specific time-lapse. Thus, our experimental design was not suitable for precisely defining gene expression in the sex-determination period. Therefore, we also elect IGFBP-6 as a sex-determination candidate gene of piauçu.

Further extensive research is necessary to corroborate one of these candidates; or identify others; *e. g.*, in the W- and Z- missing sequences. The first step, along with the assembly of the whole Z and W-specific regions, would be designing primers to quantify expression levels at different time points of the sexual determination period of males and females (qPCR).

8.4.14 Data Availability

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAJQXZ000000000. The version described in this paper is version JAJQXZ010000000. The assembled genome is available at the NCBI with the accession number GCA_021613375.1.

8.4.15 References

ALBERT, J. S.; TAGLIACOLLO, V. A.; DAGOSTA, F. Diversification of Neotropical Freshwater Fishes. 2020.

APARICIO, S.; CHAPMAN, J.; STUPKA, E.; PUTNAM, N.; CHIA, J. ming; DEHAL, P.; CHRISTOFFELS, A.; RASH, S.; HOON, S.; SMIT, A.; SOLLEWIJN GELPKE, M. D.; ROACH, J.; OH, T.; HO, I. Y.; WONG, M.; DETTER, C.; VERHOEF, F.; PREDKI, P.; TAY, A.; LUCAS, S.; RICHARDSON, P.; SMITH, S. F.; CLARK, M. S.; EDWARDS, Y. J. K.; DOGGETT, N.; ZHARKIKH, A.; TAVTIGIAN, S. V.; PRUSS, D.; BARNSTEAD, M.; EVANS, C.; BADEN, H.; POWELL, J.; GLUSMAN, G.; ROWEN, L.; HOOD, L.; TAN, Y. H.; ELGAR, G.; HAWKINS, T.; VENKATESH, B.; ROKHSAR, D.; BRENNER, S. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. **Science**, v. 297, n. 5585, p. 1301–1310, 23 ago. 2002.

BACHER, K. Perceptions and misconceptions of aquaculture: a global overview. **GLOBEFISH Research Programme**, v. 120, 2015.

BATEMAN, A.; MARTIN, M. J.; ORCHARD, S.; MAGRANE, M.; AGIVETOVA, R.; AHMAD, S.; ALPI, E.; BOWLER-BARNETT, E. H.; BRITTO, R.; BURSTEINAS, B.; BYE-A-JEE, H.; COETZEE, R.; CUKURA, A.; SILVA, A. da; DENNY, P.; DOGAN, T.; EBENEZER, T. G.; FAN, J.; CASTRO, L. G.; GARMIRI, P.; GEORGHIOU, G.; GONZALES, L.; HATTON-ELLIS, E.; HUSSEIN, A.; IGNATCHENKO, A.; INSANA, G.; ISHTIAQ, R.; JOKINEN, P.; JOSHI, V.; JYOTHI, D.; LOCK, A.; LOPEZ, R.; LUCIANI, A.; LUO, J.; LUSSI, Y.; MACDOUGALL, A.; MADEIRA, F.; MAHMOUDY, M.; MENCHI, M.; MISHRA, A.; MOULANG, K.; NIGHTINGALE, A.; OLIVEIRA, C. S.; PUNDIR, S.; QI, G.; RAJ, S.; RICE, D.; LOPEZ, M. R.; SAIDI, R.; SAMPSON, J.; SAWFORD, T.; SPERETTA, E.; TURNER, E.; TYAGI, N.; VASUDEV, P.; VOLYNKIN, V.; WARNER, K.; WATKINS, X.; ZARU, R.; ZELLNER, H.; BRIDGE, A.; POUX, S.; REDASCHI, N.; AIMO, L.; ARGOUD-PUY, G.; AUCHINCLOSS, A.; AXELSEN, K.; BANSAL, P.; BARATIN, D.; BLATTER, M. C.; BOLLEMAN, J.; BOUTET, E.; BREUZA, L.; CASALS-CASAS, C.; DE CASTRO, E.; ECHIOUKH, K. C.; COUDERT, E.; CUCHE, B.; DOCHE, M.; DORNEVIL, D.; ESTREICHER, A.; FAMIGLIETTI, M. L.; FEUERMANN, M.; GASTEIGER, E.; GEHANT, S.; GERRITSEN, V.; GOS, A.; GRUAZ-GUMOWSKI, N.; HINZ, U.; HULO, C.; HYKA-NOUSPIKEL, N.; JUNGO, F.; KELLER, G.; KERHORNOU, A.; LARA, V.; LE MERCIER, P.; LIEBERHERR, D.; LOMBARDOT, T.; MARTIN, X.; MASSON, P.; MORGAT, A.; NETO, T. B.; PAESANO, S.; PEDRUZZI, I.; PILBOUT, S.; POURCEL, L.; POZZATO, M.; PRUESS, M.; RIVOIRE, C.; SIGRIST, C.; SONESSON, K.; STUTZ, A.; SUNDARAM, S.; TOGNOLLI, M.; VERBREGUE, L.; WU, C. H.; ARIGHI, C. N.; ARMINSKI, L.; CHEN, C.; CHEN, Y.; GARAVELLI, J. S.; HUANG, H.; LAIHO, K.; MCGARVEY, P.; NATALE, D. A.; ROSS, K.; VINAYAKA, C. R.; WANG, Q.; WANG, Y.; YEH, L. S.; ZHANG, J. UniProt: the universal protein knowledgebase in 2021. **Nucleic Acids Research**, v. 49, n. D1, p. D480–D489, 8 jan. 2021a. Disponível em: <https://academic.oup.com/nar/article/49/D1/D480/6006196>. Acesso em: 14 fev. 2022.

BATEMAN, A.; MARTIN, M. J.; ORCHARD, S.; MAGRANE, M.; AGIVETOVA, R.; AHMAD, S.; ALPI, E.; BOWLER-BARNETT, E. H.; BRITTO, R.; BURSTEINAS, B.; BYE-A-JEE, H.; COETZEE, R.; CUKURA, A.; SILVA, A. Da; DENNY, P.; DOGAN, T.; EBENEZER, T. G.; FAN, J.; CASTRO, L. G.; GARMIRI, P.; GEORGHIOU, G.; GONZALES, L.; HATTON-

ELLIS, E.; HUSSEIN, A.; IGNATCHENKO, A.; INSANA, G.; ISHTIAQ, R.; JOKINEN, P.; JOSHI, V.; JYOTHI, D.; LOCK, A.; LOPEZ, R.; LUCIANI, A.; LUO, J.; LUSSI, Y.; MACDOUGALL, A.; MADEIRA, F.; MAHMOUDY, M.; MENCHI, M.; MISHRA, A.; MOULANG, K.; NIGHTINGALE, A.; OLIVEIRA, C. S.; PUNDIR, S.; QI, G.; RAJ, S.; RICE, D.; LOPEZ, M. R.; SAIDI, R.; SAMPSON, J.; SAWFORD, T.; SPERETTA, E.; TURNER, E.; TYAGI, N.; VASUDEV, P.; VOLYNKIN, V.; WARNER, K.; WATKINS, X.; ZARU, R.; ZELLNER, H.; BRIDGE, A.; POUX, S.; REDASCHI, N.; AIMO, L.; ARGOUD-PUY, G.; AUCHINCLOSS, A.; AXELSEN, K.; BANSAL, P.; BARATIN, D.; BLATTER, M. C.; BOLLEMAN, J.; BOUTET, E.; BREUZA, L.; CASALS-CASAS, C.; DE CASTRO, E.; ECHIOUKH, K. C.; COUDERT, E.; CUCHE, B.; DOCHE, M.; DORNEVIL, D.; ESTREICHER, A.; FAMIGLIETTI, M. L.; FEUERMANN, M.; GASTEIGER, E.; GEHANT, S.; GERRITSEN, V.; GOS, A.; GRUAZ-GUMOWSKI, N.; HINZ, U.; HULO, C.; HYKA-NOUSPIKEL, N.; JUNGO, F.; KELLER, G.; KERHORNOU, A.; LARA, V.; LE MERCIER, P.; LIEBERHERR, D.; LOMBARDOT, T.; MARTIN, X.; MASSON, P.; MORGAT, A.; NETO, T. B.; PAESANO, S.; PEDRUZZI, I.; PILBOUT, S.; POURCEL, L.; POZZATO, M.; PRUESS, M.; RIVOIRE, C.; SIGRIST, C.; SONESSON, K.; STUTZ, A.; SUNDARAM, S.; TOGNOLLI, M.; VERBREGUE, L.; WU, C. H.; ARIGHI, C. N.; ARMINSKI, L.; CHEN, C.; CHEN, Y.; GARAVELLI, J. S.; HUANG, H.; LAIHO, K.; MCGARVEY, P.; NATALE, D. A.; ROSS, K.; VINAYAKA, C. R.; WANG, Q.; WANG, Y.; YEH, L. S.; ZHANG, J. UniProt: the universal protein knowledgebase in 2021. **Nucleic Acids Research**, v. 49, n. D1, p. D480–D489, 8 jan. 2021b.

BELLOTT, D. W.; SKALETSKY, H.; CHO, T.-J.; BROWN, L.; LOCKE, D.; CHEN, N.; GALKINA, S.; PYNTIKOVA, T.; KOUTSEVA, N.; GRAVES, T.; KREMITZKI, C.; WARREN, W. C.; CLARK, A. G.; GAGINSKAYA, E.; WILSON, R. K.; DAVID, &. Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. **Nature Publishing Group**, v. 49, 2017.

BELTON, J.-M.; MCCORD, R. P.; GIBCUS, J. H.; NAUMOVA, N.; ZHAN, Y.; DEKKER, J. Hi-C: A comprehensive technique to capture the conformation of genomes. 2012.

BERTOLLO, L. A. C.; MOREIRA–FILHO, O.; GALETTI, P. M. Cytogenetics and taxonomy: considerations based on chromosome studies of freshwater fish. **Journal of Fish Biology**, v. 28, n. 2, p. 153–159, 1986.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 1 ago. 2014.

BOUZA, C.; HERMIDA, M.; PARDO, B. G.; FERNÁNDEZ, C.; FORTES, G. G.; CASTRO, J.; SÁNCHEZ, L.; PRESA, P.; PÉREZ, M.; SANJUÁN, A.; DE CARLOS, A.; ÁLVAREZ-DIOS, J. A.; EZCURRA, S.; CAL, R. M.; PIFERRER, F.; MARTÍNEZ, P. A microsatellite genetic map of the turbot (Scophthalmus maximus). **Genetics**, v. 177, n. 4, 2007.

BRAY, N. L.; PIMENTEL, H.; MELSTED, P.; PACHTER, L. Near-optimal probabilistic RNA-seq quantification. **Nature Biotechnology 2016 34:5**, v. 34, n. 5, p. 525–527, 4 abr. 2016.

BRŮNA, T.; HOFF, K. J.; LOMSADZE, A.; STANKE, M.; BORODOVSKY, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. **NAR Genomics and Bioinformatics**, v. 3, n. 1, p. 1–11, 6 jan. 2021a. Disponível em: <https://academic.oup.com/nargab/article/3/1/lqaa108/6066535>. Acesso em: 14 fev. 2022.

BRŮNA, T.; HOFF, K. J.; LOMSADZE, A.; STANKE, M.; BORODOVSKY, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. **NAR Genomics and Bioinformatics**, v. 3, n. 1, p. 1–11, 6 jan. 2021b.

BURTON, D. R.; WALKER, L. M.; PHOGAT, S. K.; CHAN-HUI, P. Y.; WAGNER, D.; PHUNG, P.; GOSS, J. L.; WRIN, T.; SIMEK, M. D.; FLING, S.; MITCHAM, J. L.; LEHRMAN, J. K.; PRIDDY, F. H.; OLSEN, O. A.; FREY, S. M.; HAMMOND, P. W.; PROTOCOL, G.; KAMINSKY, S.; ZAMB, T.; MOYLE, M.; KOFF, W. C.; POIGNARD, P. Broad and potent neutralizing antibodies from an african donor reveal a new HIV-1 vaccine target. **Science**, v. 326, n. 5950, p. 285–289, 2009. Disponível em: <www.sciencemag.org/cgi/content/full/1178746/DC1>.

CAPELLA-GUTIÉRREZ, S.; SILLA-MARTÍNEZ, J. M.; GABALDÓN, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. **BIOINFORMATICS APPLICATIONS NOTE**, v. 25, n. 15, p. 1972–1973, 2009.

CASTAÑO-SÁNCHEZ, C.; FUJI, K.; OZAKI, A.; HASEGAWA, O.; SAKAMOTO, T.; MORISHIMA, K.; NAKAYAMA, I.; FUJIWARA, A.; MASAOKA, T.; OKAMOTO, H.; HAYASHIDA, K.; TAGAMI, M.; KAWAI, J.; HAYASHIZAKI, Y.; OKAMOTO, N. A second generation genetic linkage map of Japanese flounder (Paralichthys olivaceus). **BMC Genomics**, v. 11, n. 1, 2010.

CATCHEN, J.; AMORES, A.; BASSHAM, S. Chromonomer: A tool set for repairing and enhancing assembled genomes through integration of genetic maps and conserved synteny. **G3: Genes, Genomes, Genetics**, v. 10, n. 11, 2020.

CATCHEN, J.; HOHENLOHE, P. A.; BASSHAM, S.; AMORES, A.; CRESKO, W. A. Stacks: An analysis tool set for population genomics. **Molecular Ecology**, v. 22, n. 11, p. 3124–3140, 2013.

CATCHEN, J. M.; AMORES, A.; HOHENLOHE, P.; CRESKO, W.; POSTLETHWAIT, J. H. Stacks: Building and genotyping loci de novo from short-read sequences. **G3: Genes, Genomes, Genetics**, v. 1, n. 3, p. 171–182, 2011.

CHALOPIN, D.; NAVILLE, M.; PLARD, F.; GALIANA, D.; VOLFF, J.-N. Comparative Analysis of Transposable Elements Highlights Mobilome Diversity and Evolution in Vertebrates. [s.d.]

CHARLESWORTH, B. **The Evoluton of Sex Chromosomes Downloaded from**. [s.l: s.n.].

CHARLESWORTH, B. **The Evoluton of Sex Chromosomes Downloaded from**. [s.l: s.n.].

CHARLESWORTH, D. **Young sex chromosomes in plants and animalsNew Phytologist**2019a.

CHARLESWORTH, D. **Young sex chromosomes in plants and animalsNew Phytologist**2019b.

CHARLESWORTH, D. The timing of genetic degeneration of sex chromosomes. 2021c.

CHARLESWORTH, D. The timing of genetic degeneration of sex chromosomes. 2021d.

CHARLESWORTH, D.; CHARLESWORTH, B.; MARAIS, G. Steps in the evolution of heteromorphic sex chromosomes. **Heredity 2005 95:2**, v. 95, n. 2, p. 118–128, 25 maio 2005.

CHEN, S.; ZHANG, G.; SHAO, C.; HUANG, Q.; LIU, G.; ZHANG, P.; SONG, W.; AN, N.; CHALOPIN, D.; VOLFF, J. N.; HONG, Y.; LI, Q.; SHA, Z.; ZHOU, H.; XIE, M.; YU, Q.; LIU, Y.; XIANG, H.; WANG, N.; WU, K.; YANG, C.; ZHOU, Q.; LIAO, X.; YANG, L.; HU, Q.; ZHANG, J.; MENG, L.; JIN, L.; TIAN, Y.; LIAN, J.; YANG, J.; MIAO, G.; LIU, S.; LIANG, Z.; YAN, F.; LI, Y.; SUN, B.; ZHANG, H.; ZHANG, J.; ZHU, Y.; DU, M.; ZHAO, Y.; SCHARTL, M.; TANG, Q.; WANG, J. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. **Nature Genetics**, v. 46, n. 3, p. 253–260, 2014.

CHIN, C. S.; PELUSO, P.; SEDLAZECK, F. J.; NATTESTAD, M.; CONCEPCION, G. T.; CLUM, A.; DUNN, C.; O'MALLEY, R.; FIGUEROA-BALDERAS, R.; MORALES-CRUZ, A.; CRAMER, G. R.; DELLEDONNE, M.; LUO, C.; ECKER, J. R.; CANTU, D.; RANK, D. R.; SCHATZ, M. C. Phased diploid genome assembly with single-molecule real-time sequencing. **Nature Methods**, v. 13, n. 12, p. 1050–1054, 2016.

CONTE, M. A.; CLARK, F. E.; ROBERTS, R. B.; XU, L.; TAO, W.; ZHOU, Q.; WANG, D.; KOCHER, T. D. Origin of a Giant Sex Chromosome. **Molecular biology and evolution**, v. 38, n. 4, p. 1554–1569, 2021.

CORNELIO, D.; CASTRO, J. P.; SANTOS, M. H.; VICARI, M. R.; DE ALMEIDA, M. C.; MOREIRA-FILHO, O.; CAMACHO, J. P. M.; ARTONI, R. F. Hermaphroditism can compensate for the sex ratio in the Astyanax scabripinnis species complex (Teleostei: Characidae): expanding the B chromosome study model. **Reviews in Fish Biology and Fisheries**, v. 27, n. 3, p. 681–689, 1 set. 2017. Disponível em: <https://link.springer.com/article/10.1007/s11160-017-9488-8>. Acesso em: 8 mar. 2022.

DE LA SERRANA, D. G.; MACQUEEN, D. J. **Insulin-like growth factor-binding proteins of teleost fishesFrontiers in Endocrinology**2018a.

DE LA SERRANA, D. G.; MACQUEEN, D. J. **Insulin-like growth factor-binding proteins of teleost fishesFrontiers in Endocrinology**2018b.

DE LOS RÍOS-PÉREZ, L.; NGUINKAL, J. A.; VERLEIH, M.; REBL, A.; BRUNNER, R. M.; KLOSA, J.; SCHÄFER, N.; STÜEKEN, M.; GOLDAMMER, T.; WITTENBURG, D. An ultra-high density SNP-based linkage map for enhancing the pikeperch (Sander lucioperca) genome assembly to chromosome-scale. **Scientific Reports**, v. 10, n. 1, 2020.

DOMINGUES, R. R.; MASTROCHIRICO-FILHO, V. A.; MENDES, N. J.; HASHIMOTO, D. T.; COELHO, R.; DA CRUZ, V. P.; ANTUNES, A.; FORESTI, F.; MENDONÇA, F. F.

Comparative eye and liver differentially expressed genes reveal monochromatic vision and cancer resistance in the shortfin mako shark (Isurus oxyrinchus). **Genomics**, v. 112, n. 6, 2020.

DU, X.; HONG, X.; FAN, G.; HUANG, X.; SUN, S.; BINGJIE, O.; ZHANG, H.; ZHANG, M.; LIU, S.; LIU, X.; ZHANG, W. Chromosome-level genome assembly of the African pike, Hepsetus odoe. **bioRxiv**, p. 2020.05.13.094987, 14 maio 2020.

DUDCHENKO, O.; BATRA, S. S.; OMER, A. D.; NYQUIST, S. K.; HOEGER, M.; DURAND, N. C.; SHAMIM, M. S.; MACHOL, I.; LANDER, E. S.; AIDEN, A. P.; AIDEN, E. L. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. **Science**, v. 356, n. 6333, p. 92–95, 2017.

DURAND, N. C.; SHAMIM, M. S.; MACHOL, I.; RAO, S. S. P.; HUNTLEY, M. H.; LANDER, E. S.; AIDEN, E. L. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. **Cell Systems**, v. 3, n. 1, 2016.

EID, J.; FEHR, A.; GRAY, J.; LUONG, K.; LYLE, J.; OTTO, G.; PELUSO, P.; RANK, D.; BAYBAYAN, P.; BETTMAN, B.; BIBILLO, A.; BJORNSON, K.; CHAUDHURI, B.; CHRISTIANS, F.; CICERO, R.; CLARK, S.; DALAL, R.; DEWINTER, A.; DIXON, J.; FOQUET, M.; GAERTNER, A.; HARDENBOL, P.; HEINER, C.; HESTER, K.; HOLDEN, D.; KEARNS, G.; KONG, X.; KUSE, R.; LACROIX, Y.; LIN, S.; LUNDQUIST, P.; MA, C.; MARKS, P.; MAXHAM, M.; MURPHY, D.; PARK, I.; PHAM, T.; PHILLIPS, M.; ROY, J.; SEBRA, R.; SHEN, G.; SORENSON, J.; TOMANEY, A.; TRAVERS, K.; TRULSON, M.; VIECELI, J.; WEGENER, J.; WU, D.; YANG, A.; ZACCARIN, D.; ZHAO, P.; ZHONG, F.; KORLACH, J.; TURNER, S. Real-Time DNA Sequencing from Single Polymerase Molecules. **Science**, v. 323, n. 5910, p. 133–138, 2 jan. 2009a.

EID, J.; FEHR, A.; GRAY, J.; LUONG, K.; LYLE, J.; OTTO, G.; PELUSO, P.; RANK, D.; BAYBAYAN, P.; BETTMAN, B.; BIBILLO, A.; BJORNSON, K.; CHAUDHURI, B.; CHRISTIANS, F.; CICERO, R.; CLARK, S.; DALAL, R.; DEWINTER, A.; DIXON, J.; FOQUET, M.; GAERTNER, A.; HARDENBOL, P.; HEINER, C.; HESTER, K.; HOLDEN, D.; KEARNS, G.; KONG, X.; KUSE, R.; LACROIX, Y.; LIN, S.; LUNDQUIST, P.; MA, C.; MARKS, P.; MAXHAM, M.; MURPHY, D.; PARK, I.; PHAM, T.; PHILLIPS, M.; ROY, J.; SEBRA, R.; SHEN, G.; SORENSON, J.; TOMANEY, A.; TRAVERS, K.; TRULSON, M.; VIECELI, J.; WEGENER, J.; WU, D.; YANG, A.; ZACCARIN, D.; ZHAO, P.; ZHONG, F.; KORLACH, J.; TURNER, S. Real-Time DNA Sequencing from Single Polymerase Molecules. **Science**, v. 323, n. 5910, p. 133–138, 2 jan. 2009b.

EMMS, D. M.; KELLY, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. **Genome Biology**, v. 20, n. 1, p. 1–14, 14 nov. 2019.

FAO. **The state of world fisheries and aquaculture 2018- Meeting the sustainable development goals. Rome**. [s.l: s.n.]v. 35

FLYNN, J. M.; HUBLEY, R.; GOUBERT, C.; ROSEN, J.; CLARK, A. G.; FESCHOTTE, C.; SMIT, A. F. RepeatModeler2 for automated genomic discovery of transposable element families. **PNAS**, v. 117, n. 17, p. 9451–9457, 2020.

FORESTI, F.; TOLEDO, L. F. A.; TOLEDO, S. A. Polymorphic nature of nucleolus organizer regions in fishes. **Cytogenetic and Genome Research**, v. 31, n. 3, 1981.

GABRIEL, L.; HOFF, K. J.; BRŮNA, T.; BORODOVSKY, M.; STANKE, M. TSEBRA: transcript selector for BRAKER. **BMC Bioinformatics**, v. 22, n. 1, p. 1–12, 1 dez. 2021a.

GABRIEL, L.; HOFF, K. J.; BRŮNA, T.; BORODOVSKY, M.; STANKE, M. TSEBRA: transcript selector for BRAKER. **BMC Bioinformatics**, v. 22, n. 1, p. 1–12, 1 dez. 2021b.

GALETTI, JR., P. M.; FORESTI, F.; BERTQLLO, L. A. C.; FILHO, M. Heteromorphic sex chromosomes in three species of the genus Leporinus (Pisces, Anostomidae). **Cytogenetic and Genome Research**, v. 29, n. 3, p. 138–142, 7 maio 1981. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/7226896/>. Acesso em: 18 ago. 2021.

GAO, B.; SHEN, D.; XUE, S.; CHEN, C.; CUI, H.; SONG, C. The contribution of transposable elements to size variations between four teleost genomes. **Mobile DNA**, v. 7, n. 1, 9 fev. 2016.

GLASAUER, S. M. K.; NEUHAUSS, S. C. F. **Whole-genome duplication in teleost fishes and its evolutionary consequencesMolecular Genetics and Genomics**Mol Genet Genomics, 19 nov. 2014.

GODINHO, A. L.; KYNARD, B.; MARTINEZ, C. B. Supplemental water releases for fisheries restoration in a Brazilian floodplain River: A conceptual model. **River Research and Applications**, v. 23, n. 9, 2007.

GÖTZ, S.; GARCÍA-GÓMEZ, J. M.; TEROL, J.; WILLIAMS, T. D.; NAGARAJ, S. H.; NUEDA, M. J.; ROBLES, M.; TALÓN, M.; DOPAZO, J.; CONESA, A. High-throughput functional annotation and data mining with the Blast2GO suite. **Nucleic Acids Research**, v. 36, n. 10, p. 3420–3435, 1 jun. 2008. Disponível em: <https://academic.oup.com/nar/article/36/10/3420/2410320>. Acesso em: 14 fev. 2022.

GUERRERO-CÓZAR, I.; PEREZ-GARCIA, C.; BENZEKRI, H.; SÁNCHEZ, J. J.; SEOANE, P.; CRUZ, F.; GUT, M.; ZAMORANO, M. J.; CLAROS, M. G.; MANCHADO, M. Development of whole-genome multiplex assays and construction of an integrated genetic map using SSR markers in Senegalese sole. **Scientific Reports**, v. 10, n. 1, 2020.

HILSDORF, A. W. S.; ULIANO-SILVA, M.; COUTINHO, L. L.; MONTENEGRO, H.; ALMEIDA-VAL, V. M. F.; PINHAL, D. Genome assembly and annotation of the tambaqui (Colossoma macropomum): an emblematic fish of the Amazon River basin. **bioRxiv**, p. 2021.09.08.459456, 9 set. 2021a.

HILSDORF, A. W. S.; ULIANO-SILVA, M.; COUTINHO, L. L.; MONTENEGRO, H.; ALMEIDA-VAL, V. M. F.; PINHAL, D. Genome assembly and annotation of the tambaqui (Colossoma macropomum): an emblematic fish of the Amazon River basin. **bioRxiv**, p. 2021.09.08.459456, 9 set. 2021b.

HOFF, K. J.; LANGE, S.; LOMSADZE, A.; BORODOVSKY, M.; STANKE, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS.

Bioinformatics, v. 32, n. 5, p. 767–769, 1 mar. 2016a. Disponível em: <https://academic.oup.com/bioinformatics/article/32/5/767/1744611>. Acesso em: 3 out. 2021.

HOFF, K. J.; LANGE, S.; LOMSADZE, A.; BORODOVSKY, M.; STANKE, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. **Bioinformatics**, v. 32, n. 5, p. 767–769, 1 mar. 2016b.

HOFF, K. J.; LOMSADZE, A.; BORODOVSKY, M.; STANKE, M. Whole-Genome Annotation with BRAKER. **Methods in Molecular Biology**, v. 1962, p. 65–95, 2019. Disponível em: <https://link.springer.com/protocol/10.1007/978-1-4939-9173-0_5>. Acesso em: 14 fev. 2022.

HOWE, K. L.; ACHUTHAN, P.; ALLEN, J.; ALLEN, J.; ALVAREZ-JARRETA, J.; RIDWAN AMODE, M.; ARMEAN, I. M.; AZOV, A. G.; BENNETT, R.; BHAI, J.; BILLIS, K.; BODDU, S.; CHARKHCHI, M.; CUMMINS, C.; DA RIN FIORETTO, L.; DAVIDSON, C.; DODIYA, K.; EL HOUDAIGUI, B.; FATIMA, R.; GALL, A.; GIRON, C. G.; GREGO, T.; GUIJARRO-CLARKE, C.; HAGGERTY, L.; HEMROM, A.; HOURLIER, T.; IZUOGU, O. G.; JUETTEMANN, T.; KAIKALA, V.; KAY, M.; LAVIDAS, I.; LE, T.; LEMOS, D.; MARTINEZ, J. G.; MARUGÁN, J. C.; MAUREL, T.; MCMAHON, A. C.; MOHANAN, S.; MOORE, B.; MUFFATO, M.; OHEH, D. N.; PARASCHAS, D.; PARKER, A.; PARTON, A.; PROSOVETSKAIA, I.; SAKTHIVEL, M. P.; ABDUL SALAM, A. I.; SCHMITT, B. M.; SCHUILENBURG, H.; SHEPPARD, D.; STEED, E.; SZPAK, M.; SZUBA, M.; TAYLOR, K.; THORMANN, A.; THREADGOLD, G.; WALTS, B.; WINTERBOTTOM, A.; CHAKIACHVILI, M.; CHAUBAL, A.; DE SILVA, N.; FLINT, B.; FRANKISH, A.; HUNT, S. E.; IISLEY, G. R.; LANGRIDGE, N.; LOVELAND, J. E.; MARTIN, F. J.; MUDGE, J. M.; MORALES, J.; PERRY, E.; RUFFIER, M.; TATE, J.; THYBERT, D.; TREVANION, S. J.; CUNNINGHAM, F.; YATES, A. D.; ZERBINO, D. R.; FLICEK, P. Ensembl 2021. **Nucleic Acids Research**, v. 49, n. D1, p. D884–D891, 8 jan. 2021.

HUGHES, L. C.; ORTÍ, G.; HUANG, Y.; SUN, Y.; BALDWIN, C. C.; THOMPSON, A. W.; ARCILA, D.; BETANCUR, R.; LI, C.; BECKER, L.; BELLORA, N.; ZHAO, X.; LI, X.; WANG, M.; FANG, C.; XIE, B.; ZHOUI, Z.; HUANG, H.; CHEN, S.; VENKATESH, B.; SHI, Q. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. **Proceedings of the National Academy of Sciences of the United States of America**, v. 115, n. 24, p. 6249–6254, 12 jun. 2018.

IMARAZENE, B.; DU, K.; BEILLE, S.; JOUANNO, E.; FERON, R.; PAN, Q.; TORRES-PAZ, J.; LOPEZ-ROQUES, C.; CASTINEL, A.; GIL, L.; KUCHLY, C.; DONNADIEU, C.; PARRINELLO, H.; JOURNOT, L.; CABAU, C.; ZAHM, M.; KLOPP, C.; PAVLICA, T.; AL-RIKABI, A.; LIEHR, T.; SIMANOVSKY, S. A.; BOHLEN, J.; SEMBER, A.; PEREZ, J.; VEYRUNES, F.; MUELLER, T. D.; POSTLETHWAIT, J. H.; SCHARTL, M.; HERPIN, A.; RÉTAUX, S.; GUIGUEN, Y. A supernumerary “B-sex” chromosome drives male sex determination in the Pachón cavefish, Astyanax mexicanus. **Current Biology**, v. 31, n. 21, p. 4800- 4809.e9, 8 nov. 2021. Disponível em: <https://doi.org/10.1016/j.cub.2021.08.030>. Acesso em: 4 maio. 2022.

KARNOVSKY, M. J. A formaldehyde-glutaraldehyde fixative of high osmolality for use in electron microscopy. **Journal cell biology**, v. 27, p. 137–138, 1985. Disponível em: <https://www.researchgate.net/publication/244955881>. Acesso em: 28 fev. 2022.

KATOH, K.; STANDLEY, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. **Molecular Biology and Evolution**, v. 30, n. 4, p. 772–780, 1 abr. 2013.

KRATOCHVÍL, L.; GAMBLE, T.; ROVATSOS, M. Sex chromosome evolution among amniotes: Is the origin of sex chromosomes non-random? **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 376, n. 1833, 13 set. 2021. . Acesso em: 26 abr. 2022.

LAHN, B. T.; PAGE, D. C. Four evolutionary strata on the human X chromosome. **Science**, v. 286, n. 5441, 1999.

LI, H.; DURBIN, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. **Bioinformatics**, v. 26, n. 5, p. 589–595, 15 jan. 2010.

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, 2009a.

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, 2009b.

LI, M.; ZHANG, R.; FAN, G.; XU, W.; ZHOU, Q.; WANG, L.; LI, W.; PANG, Z.; YU, M.; LIU, Q.; LIU, X.; SCHARTL, M.; CHEN, S. Reconstruction of the Origin of a Neo-Y Sex Chromosome and Its Evolution in the Spotted Knifejaw, Oplegnathus punctatus. **Molecular biology and evolution**, v. 38, n. 6, p. 2615–2626, 19 maio 2021. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/33693787/>. Acesso em: 22 set. 2021.

LINKHART, T. A.; MOHAN, S.; BAYLINK, D. J. Growth factors for bone growth and repair: IGF, TGFβ and BMP. In: Bone, 1 SUPPL., 1996, [...]. 1996. v. 19

LIU, J.; WANG, Z.; LI, J.; XU, L.; LIU, J.; FENG, S.; GUO, C.; CHEN, S.; REN, Z.; RAO, J.; WEI, K.; CHEN, Y.; JARVIS, E. D.; ZHANG, G.; ZHOU, Q. A new emu genome illuminates the evolution of genome configuration and nuclear architecture of avian chromosomes. **Genome Research**, 2021.

LOGSDON, G. A.; VOLLGER, M. R.; EICHLER, E. E. Long-read human genome sequencing and its applications. **Nature Reviews Genetics 2020 21:10**, v. 21, n. 10, p. 597–614, 5 jun. 2020a. Disponível em: <https://www.nature.com/articles/s41576-020-0236-x>. Acesso em: 23 ago. 2021.

LOGSDON, G. A.; VOLLGER, M. R.; EICHLER, E. E. Long-read human genome sequencing and its applications. **Nature Reviews Genetics 2020 21:10**, v. 21, n. 10, p. 597–614, 5 jun. 2020b.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, n. 12, p. 1–21, 5 dez. 2014.

MANNI, M.; BERKELEY, M. R.; SEPPEY, M.; SIMÃO, F. A.; ZDOBNOV, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. **Molecular Biology and Evolution**, 28 jul. 2021a. Disponível em: <https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab199/6329644>. Acesso em: 16 ago. 2021.

MANNI, M.; BERKELEY, M. R.; SEPPEY, M.; SIMÃO, F. A.; ZDOBNOV, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. **Molecular Biology and Evolution**, 28 jul. 2021b.

MARÇAIS, G.; KINGSFORD, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. **Bioinformatics**, v. 27, n. 6, p. 764–770, 15 mar. 2011.

MELO, B. F.; SIDLAUSKAS, B. L.; NEAR, T. J.; ROXO, F. F.; GHEZELAYAGH, A.; OCHOA, L. E.; STIASSNY, M. L. J.; ARROYAVE, J.; CHANG, J.; FAIRCLOTH, B. C.; MACGUIGAN, D. J.; HARRINGTON, R. C.; BENINE, R. C.; BURNS, M. D.; HOEKZEMA, K.; SANCHES, N. C.; MALDONADO-OCAMPO, J. A.; CASTRO, R. M. C.; FORESTI, F.; ALFARO, M. E.; OLIVEIRA, C. Accelerated Diversification Explains the Exceptional Species Richness of Tropical Characoid Fishes. **Systematic Biology**, v. 0, n. 0, p. 1–15, 7 jun. 2021a.

MELO, B. F.; SIDLAUSKAS, B. L.; NEAR, T. J.; ROXO, F. F.; GHEZELAYAGH, A.; OCHOA, L. E.; STIASSNY, M. L. J.; ARROYAVE, J.; CHANG, J.; FAIRCLOTH, B. C.; MACGUIGAN, D. J.; HARRINGTON, R. C.; BENINE, R. C.; BURNS, M. D.; HOEKZEMA, K.; SANCHES, N. C.; MALDONADO-OCAMPO, J. A.; CASTRO, R. M. C.; FORESTI, F.; ALFARO, M. E.; OLIVEIRA, C. Accelerated Diversification Explains the Exceptional Species Richness of Tropical Characoid Fishes. **Systematic Biology**, v. 0, n. 0, p. 1–15, 7 jun. 2021b.

MEYER, A.; SCHLOISSNIG, S.; FRANCHINI, P.; DU, K.; WOLTERING, J. M.; IRISARRI, I.; WONG, W. Y.; NOWOSHILOW, S.; KNEITZ, S.; KAWAGUCHI, A.; FABRIZIUS, A.; XIONG, P.; DECHAUD, C.; SPAINK, H. P.; VOLFF, J.-N.; SIMAKOV, O.; BURMESTER, T.; TANAKA, E. M.; SCHARTL, M. Giant lungfish genome elucidates the conquest of land by vertebrates. **Nature 2021 590:7845**, v. 590, n. 7845, p. 284–289, 18 jan. 2021.

O'LEARY, N. A.; WRIGHT, M. W.; BRISTER, J. R.; CIUFO, S.; HADDAD, D.; MCVEIGH, R.; RAJPUT, B.; ROBBERTSE, B.; SMITH-WHITE, B.; AKO-ADJEI, D.; ASTASHYN, A.; BADRETDIN, A.; BAO, Y.; BLINKOVA, O.; BROVER, V.; CHETVERNIN, V.; CHOI, J.; COX, E.; ERMOLAEVA, O.; FARRELL, C. M.; GOLDFARB, T.; GUPTA, T.; HAFT, D.; HATCHER, E.; HLAVINA, W.; JOARDAR, V. S.; KODALI, V. K.; LI, W.; MAGLOTT, D.; MASTERSON, P.; MCGARVEY, K. M.; MURPHY, M. R.; O'NEILL, K.; PUJAR, S.; RANGWALA, S. H.; RAUSCH, D.; RIDDICK, L. D.; SCHOCH, C.; SHKEDA, A.; STORZ, S. S.; SUN, H.; THIBAUD-NISSEN, F.; TOLSTOY, I.; TULLY, R. E.; VATSAN, A. R.; WALLIN, C.; WEBB, D.; WU, W.; LANDRUM, M. J.; KIMCHI, A.; TATUSOVA, T.; DICUCCIO, M.; KITTS, P.; MURPHY, T. D.; PRUITT, K. D. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. **Nucleic acids research**, v. 44, n. D1, p. D733–D745, 2016.

OLIVEIRA, C.; FORESTI, F.; HILSDORF, A. W. S. Genetics of neotropical fish: From chromosomes to populations. **Fish Physiology and Biochemistry**, v. 35, n. 1, p. 81–100, 2009.

PAN, Q.; KAY, T.; DEPINCÉ, A.; ADOLFI, M.; SCHARTL, M.; GUIGUEN, Y.; HERPIN, A. Evolution of master sex determiners: TGF-β signalling pathways at regulatory crossroads. **Philosophical Transactions of the Royal Society B**, v. 376, n. 1832, 30 ago. 2021.

PEICHEL, C. L.; MCCANN, S. R.; ROSS, J. A.; NAFTALY, A. F. S.; URTON, J. R.; CECH, J. N.; GRIMWOOD, J.; SCHMUTZ, J.; MYERS, R. M.; KINGSLEY, D. M.; WHITE, M. A. Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. **Genome Biology 2020 21:1**, v. 21, n. 1, p. 1–31, 19 jul. 2020a. Disponível em: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02097-x>. Acesso em: 22 set. 2021.

PEICHEL, C. L.; MCCANN, S. R.; ROSS, J. A.; NAFTALY, A. F. S.; URTON, J. R.; CECH, J. N.; GRIMWOOD, J.; SCHMUTZ, J.; MYERS, R. M.; KINGSLEY, D. M.; WHITE, M. A. Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. **Genome Biology 2020 21:1**, v. 21, n. 1, p. 1–31, 19 jul. 2020b.

PETERSON, B. K.; WEBER, J. N.; KAY, E. H.; FISHER, H. S.; HOEKSTRA, H. E. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. **PLoS ONE**, v. 7, n. 5, 2012.

POLLARD, M. O.; GURDASANI, D.; MENTZER, A. J.; PORTER, T.; SANDHU, M. S. Long reads: their purpose and place. **Human Molecular Genetics**, v. 27, n. R2, p. R234–R241, 1 ago. 2018. Disponível em: <https://academic.oup.com/hmg/article/27/R2/R234/4996216>. Acesso em: 25 ago. 2021.

PORTO-FORESTI, F.; HASHIMOTO, D. T.; ALVES, A. L.; ALMEIDA, R. B. C.; SENHORINI, J. A.; BORTOLOZZI, J.; FORESTI, F. Cytogenetic markers as diagnoses in the identification of the hybrid between Piau&u (Leporinus macrocephalus) and Piapara (Leporinus elongatus). **Genetics and Molecular Biology**, v. 31, n. 1 SUPPL. 1, p. 195–202, 2008.

PRIYANKA, P. P.; YENUGU, S. **Coiled-Coil Domain-Containing (CCDC) Proteins: Functional Roles in General and Male Reproductive PhysiologyReproductive Sciences**2021.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A. R.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P. I. W.; DALY, M. J.; SHAM, P. C. PLINK: A tool set for whole-genome association and population-based linkage analyses. **American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, 2007.

RAFATI, N.; CHEN, J.; HERPIN, A.; PETTERSSON, M. E.; HAN, F.; FENG, C.; WALLERMAN, O.; RUBIN, C.-J.; PÉRON, S.; COCCO, A.; LARSSON, M.; TRÖTSCHEL, C.; POETSCH, A.; KORSCHING, K.; BÖNIGK, W.; KÖRSCHEN, H. G.; BERG, F.; FOLKVORD, A.; KAUPP, U. B.; SCHARTL, M.; ANDERSSON, L. Reconstruction of the birth of a male sex chromosome present in Atlantic herring. **Proceedings of the National Academy of Sciences**, v.

117, n. 39, p. 24359–24368, 29 set. 2020. Disponível em: <https://www.pnas.org/content/117/39/24359>. Acesso em: 22 set. 2021.

RAMIREZ, J. L.; BIRINDELLI, J. L. O.; GALETTI, P. M. A new genus of Anostomidae (Ostariophysi: Characiformes): Diversity, phylogeny and biogeography based on cytogenetic, molecular and morphological data. **Molecular Phylogenetics and Evolution**, v. 107, p. 308–323, 2017. Disponível em: <http://dx.doi.org/10.1016/j.ympev.2016.11.012>.

RAMOS, L.; ANTUNES, A. Decoding sex: Elucidating sex determination and how high-quality genome assemblies are untangling the evolutionary dynamics of sex chromosomes. **Genomics**, v. 114, n. 2, p. 110277, 1 mar. 2022. Disponível em: <https://doi.org/10.1016/j.ygeno.2022.110277>. Acesso em: 27 abr. 2022.

RASTAS, P. Lep-MAP3: Robust linkage mapping even for low-coverage whole genome sequencing data. **Bioinformatics**, v. 33, n. 23, p. 3726–3732, 2017.

RASTAS, P.; PAULIN, L.; HANSKI, I.; LEHTONEN, R.; AUVINEN, P.; BRUDNO, M. Lep-MAP: Fast and accurate linkage map construction for large SNP datasets. **Bioinformatics**, v. 29, n. 24, 2013.

REID, D. P.; SMITH, C. A.; ROMMENS, M.; BLANCHARD, B.; MARTIN-ROBICHAUD, D.; REITH, M. A genetic linkage map of Atlantic halibut (Hippoglossus hippoglossus L.). **Genetics**, v. 177, n. 2, 2007.

REYNALTE-TATAJE, D.; ZANIBONI-FILHO, E.; MUELBERT, B. Estadios do desenvolvimento embrionário do piavuçu Leporinus macrocephalus (Garavello & Britski, 1988). **Acta Scientiarum. Animal Sciences**, v. 23, n. 0, p. 823, 9 maio 2008. Disponível em: <https://periodicos.uem.br/ojs/index.php/ActaSciAnimSci/article/view/2614>. Acesso em: 26 ago. 2021.

RHIE, A.; MCCARTHY, S. A.; FEDRIGO, O.; DAMAS, J.; FORMENTI, G.; LONDON, S. E.; CLAYTON, D. F.; MELLO, C. v; FRIEDRICH, S. R. Towards complete and error-free genome assemblies of all vertebrate species. p. 1–56, 2020a.

RHIE, A.; WALENZ, B. P.; KOREN, S.; PHILLIPPY, A. M. Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. **Genome Biology**, v. 21, n. 1, p. 1–27, 2020b.

RHIE, A.; WALENZ, B. P.; KOREN, S.; PHILLIPPY, A. M. Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. **Genome Biology**, v. 21, n. 1, p. 1–27, 2020c.

ROCHA, C. M. C. da; RESENDE, E. K. de; ROUTLEDGE, E. A. B.; LUNDSTEDT, L. M. Avanços na pesquisa e no desenvolvimento da aquicultura brasileira. **Pesquisa Agropecuária Brasileira**, v. 48, n. 8, p. iv–vi, 2013a.

ROCHA, C. M. C. da; RESENDE, E. K. de; ROUTLEDGE, E. A. B.; LUNDSTEDT, L. M. Avanços na pesquisa e no desenvolvimento da aquicultura brasileira. **Pesquisa Agropecuária Brasileira**, v. 48, n. 8, p. iv–vi, 2013b.

SCHARTL, M.; KNEITZ, S.; VOLKOFF, H.; ADOLFI, M.; SCHMIDT, C.; FISCHER, P.; MINX, P.; TOMLINSON, C.; MEYER, A.; WARREN, W. C. The Piranha Genome Provides Molecular Insight Associated to Its Unique Feeding Behavior. **Genome Biology and Evolution**, v. 11, n. 8, p. 2099–2106, 1 ago. 2019a.

SCHARTL, M.; KNEITZ, S.; VOLKOFF, H.; ADOLFI, M.; SCHMIDT, C.; FISCHER, P.; MINX, P.; TOMLINSON, C.; MEYER, A.; WARREN, W. C. The Piranha Genome Provides Molecular Insight Associated to Its Unique Feeding Behavior. **Genome Biology and Evolution**, v. 11, n. 8, p. 2099–2106, 1 ago. 2019b.

SHAO, F.; LUDWIG, A.; MAO, Y.; LIU, N.; PENG, Z. Chromosome-level genome assembly of the female western mosquitofish (Gambusia affinis). **GigaScience**, v. 9, n. 8, p. 1–10, 1 ago. 2020. Disponível em: <https://academic.oup.com/gigascience/article/9/8/giaa092/5897807>. Acesso em: 22 set. 2021.

SIDONIO, L.; CAVALCANTI, I.; CAPANEMA, L.; MORCH, R.; LIMA, J.; BURNS, V.; JÚNIOR, A. J. A.; AMARAL, J. V. Experiências internacionais aquícolas e oportunidades de desenvolvimento da aquicultura no Brasil : proposta de inserção do BNDES. **BNDES Setorial**, v. 36, 2012.

SMITH, S. R.; NORMANDEAU, E.; DJAMBAZIAN, H.; NAWARATHNA, P. M.; BERUBE, P.; MUIR, A. M.; RAGOUSSIS, J.; PENNEY, C. M.; SCRIBNER, K. T.; LUIKART, G.; WILSON, C. C.; BERNATCHEZ, L. A chromosome-anchored genome assembly for Lake Trout (Salvelinus namaycush). **Molecular Ecology Resources**, v. 22, n. 2, p. 679–694, 1 fev. 2022.

SONESON, C.; LOVE, M. I.; ROBINSON, M. D.; FLOOR, S. N. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 2; peer review: 2 approved] report report. 2016.

SONG, W.; LI, Y.; ZHAO, Y.; LIU, Y.; NIU, Y.; PANG, R.; MIAO, G.; LIAO, X.; SHAO, C.; GAO, F.; CHEN, S. Construction of a High-Density Microsatellite Genetic Linkage Map and Mapping of Sexual and Growth-Related Traits in Half-Smooth Tongue Sole (Cynoglossus semilaevis). **PLoS ONE**, v. 7, n. 12, 2012.

STAMATAKIS, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. **Bioinformatics**, v. 30, n. 9, p. 1312–1313, 1 maio 2014.

STEINKE, D.; SALZBURGER, W.; MEYER, A. Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. **Journal of Molecular Evolution**, v. 62, n. 6, p. 772–784, 11 jun. 2006.

SUMNER, A. T. A simple technique for demonstrating centromeric heterochromatin. **Experimental Cell Research**, v. 75, n. 1, 1972.

TOMASZKIEWICZ, M.; MEDVEDEV, P.; MAKOVA, K. D. Y and W Chromosome Assemblies: Approaches and Discoveries. **Trends in Genetics**, v. 33, n. 4, p. 266–282, 1 abr. 2017. . Acesso em: 27 abr. 2022.

UTSUNOMIA, R.; SILVA, D. M. Z. de A.; RUIZ-RUANO, F. J.; GOES, C. A. G.; MELO, S.; RAMOS, L. P.; OLIVEIRA, C.; PORTO-FORESTI, F.; FORESTI, F.; HASHIMOTO, D. T. Satellitome landscape analysis of Megaleporinus macrocephalus (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. **Scientific Reports**, v. 9, n. 1, p. 1–10, 2019a.

UTSUNOMIA, R.; SILVA, D. M. Z. de A.; RUIZ-RUANO, F. J.; GOES, C. A. G.; MELO, S.; RAMOS, L. P.; OLIVEIRA, C.; PORTO-FORESTI, F.; FORESTI, F.; HASHIMOTO, D. T. Satellitome landscape analysis of Megaleporinus macrocephalus (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. **Scientific Reports**, v. 9, n. 1, p. 1–10, 2019b.

VARELA, E. S.; BEKAERT, M.; GANECO-KIRSCHNIK, L. N.; TORATI, L. S.; SHIOTSUKI, L.; DE ALMEIDA, F. L.; VILLELA, L. C. V.; REZENDE, F. P.; DA SILVA BARROSO, A.; DE FREITAS, L. E. L.; TAGGART, J. B.; MIGAUD, H. A high-density linkage map and sex-linked markers for the Amazon Tambaqui Colossoma macropomum. **BMC Genomics**, v. 22, n. 1, p. 1–10, 1 dez. 2021.

VOLFF, J.-N. Genome evolution and biodiversity in teleost fish. **Heredity 2005 94:3**, v. 94, n. 3, p. 280–294, 22 dez. 2004a. Disponível em: <https://www.nature.com/articles/6800635>. Acesso em: 22 set. 2021.

VOLFF, J.-N. Genome evolution and biodiversity in teleost fish. **Heredity 2005 94:3**, v. 94, n. 3, p. 280–294, 22 dez. 2004b.

VURTURE, G. W.; SEDLAZECK, F. J.; NATTESTAD, M.; UNDERWOOD, C. J.; FANG, H.; GURTOWSKI, J.; SCHATZ, M. C. GenomeScope: Fast reference-free genome profiling from short reads. **Bioinformatics**, v. 33, n. 14, p. 2202–2204, 2017.

WANG, K.; WANG, J.; ZHU, C.; YANG, L.; REN, Y.; RUAN, J.; FAN, G.; HU, J.; XU, W.; BI, X.; ZHU, Y.; SONG, Y.; CHEN, H.; MA, T.; ZHAO, R.; JIANG, H.; ZHANG, B.; FENG, C.; YUAN, Y.; GAN, X.; LI, Y.; ZENG, H.; LIU, Q.; ZHANG, Y.; SHAO, F.; HAO, S.; ZHANG, H.; XU, X.; LIU, X.; WANG, D.; ZHU, M.; ZHANG, G.; ZHAO, W.; QIU, Q.; HE, S.; WANG, W. African lungfish genome sheds light on the vertebrate water-to-land transition. **Cell**, v. 184, n. 5, p. 1362- 1376.e18, 4 mar. 2021.

WARREN, W. C.; BOGGS, T. E.; BOROWSKY, R.; CARLSON, B. M.; FERRUFINO, E.; GROSS, J. B.; HILLIER, L.; HU, Z.; KEENE, A. C.; KENZIOR, A.; KOWALKO, J. E.; TOMLINSON, C.; KREMITZKI, M.; LEMIEUX, M. E.; GRAVES-LINDSAY, T.; MCGAUGH, S. E.; MILLER, J. T.; MOMMERSTEEG, M. T. M.; MORAN, R. L.; PEUSS, R.; RICE, E. S.; RIDDLE, M. R.; SIFUENTES-ROMERO, I.; STANHOPE, B. A.; TABIN, C. J.; THAKUR, S.; YAMAMOTO, Y.; ROHNER, N. A chromosome-level genome of Astyanax mexicanus surface

fish for comparing population-specific genetic differences contributing to trait evolution. **Nature Communications 2021 12:1**, v. 12, n. 1, p. 1–12, 4 mar. 2021a.

WARREN, W. C.; BOGGS, T. E.; BOROWSKY, R.; CARLSON, B. M.; FERRUFINO, E.; GROSS, J. B.; HILLIER, L.; HU, Z.; KEENE, A. C.; KENZIOR, A.; KOWALKO, J. E.; TOMLINSON, C.; KREMITZKI, M.; LEMIEUX, M. E.; GRAVES-LINDSAY, T.; MCGAUGH, S. E.; MILLER, J. T.; MOMMERSTEEG, M. T. M.; MORAN, R. L.; PEUSS, R.; RICE, E. S.; RIDDLE, M. R.; SIFUENTES-ROMERO, I.; STANHOPE, B. A.; TABIN, C. J.; THAKUR, S.; YAMAMOTO, Y.; ROHNER, N. A chromosome-level genome of Astyanax mexicanus surface fish for comparing population-specific genetic differences contributing to trait evolution. **Nature Communications 2021 12:1**, v. 12, n. 1, p. 1–12, 4 mar. 2021b.

WRIGHT, A. E.; DEAN, R.; ZIMMER, F.; MANK, J. E. **How to make a sex chromosomeNature Communications**2016.

XIAO, Y.; XIAO, Z.; MA, D.; LIU, J.; LI, J. Genome sequence of the barred knifejaw Oplegnathus fasciatus (Temminck & Schlegel, 1844): The first chromosome-level draft genome in the family Oplegnathidae. **GigaScience**, v. 8, n. 3, p. 1–8, 2019.

XUE, L.; GAO, Y.; WU, M.; TIAN, T.; FAN, H.; HUANG, Y.; HUANG, Z.; LI, D.; XU, L. Telomere-to-telomere assembly of a fish Y chromosome reveals the origin of a young sex chromosome pair. **Genome Biology 2021 22:1**, v. 22, n. 1, p. 1–20, 12 jul. 2021a. Disponível em: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02430-y>. Acesso em: 20 out. 2021.

XUE, L.; GAO, Y.; WU, M.; TIAN, T.; FAN, H.; HUANG, Y.; HUANG, Z.; LI, D.; XU, L. Telomere-to-telomere assembly of a fish Y chromosome reveals the origin of a young sex chromosome pair. **Genome Biology 2021 22:1**, v. 22, n. 1, p. 1–20, 12 jul. 2021b.

YAN, Y. L.; BATZEL, P.; TITUS, T.; SYDES, J.; DESVIGNES, T.; BREMILLER, R.; DRAPER, B.; POSTLETHWAIT, J. H. A hormone that lost its receptor: Anti-müllerian hormone (AMH) in Zebrafish gonad development and sex determination. **Genetics**, v. 213, n. 2, 2019a.

YAN, Y. L.; BATZEL, P.; TITUS, T.; SYDES, J.; DESVIGNES, T.; BREMILLER, R.; DRAPER, B.; POSTLETHWAIT, J. H. A hormone that lost its receptor: Anti-müllerian hormone (AMH) in Zebrafish gonad development and sex determination. **Genetics**, v. 213, n. 2, 2019b.

YOU, X.; SHAN, X.; SHI, Q. **Research advances in the genomics and applications for molecular breeding of aquaculture animalsAquaculture**2020.

YUE, G. H.; WANG, L. Current status of genome sequencing and its applications in aquaculture. 2016a.

YUE, G. H.; WANG, L. Current status of genome sequencing and its applications in aquaculture. 2016b.

ZIMIN, A. V.; SALZBERG, S. L. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. **PLoS Computational Biology**, v. 16, n. 6, 2020.

ZIMINID, A. V; SALZBERGID, S. L. The SAMBA tool uses long reads to improve the contiguity of genome assemblies. **PLOS Computational Biology**, v. 18, n. 2, p. e1009860, 4 fev. 2022.

8.4.16 Tables

Table 1. Summary of sampling number per family and sex.

Family	Males	Females	Total
1	49	44	93
2	4	20	24
3	49	44	93
4	41	48	89
Total	143	156	299

Table 2. Statistics for genome assembly of *Megaleporinus macrocephalus*.

Characteristic	Value
No. scaffolds	101
No. contigs	1,353
Main genome scaffold sequence total (bp)	1,282,030,339
Main genome contig sequence total (bp)	1,280,781,659
Scaffold N50 (bp)	45,034,219
Contig N50 (bp)	5,013,076
Max. scaffold length (bp)	73,843,892
Max. contig length (bp)	25,940,738
% main genome in scaffolds > 50 kb	99.9%
BUSCO complete	96.2%
BUSCO complete and single copy	95.1%
BUSCO complete and duplicated	1.1%
BUSCO fragmented	0.6%
BUSCO missing	3.2%
Consensus quality value (QV)	37.53
Merqury completeness	93.05%

Table 3. Comparison between available genome assemblies of Neotropical fish species

Species	**Genome size*(bp)**	**Scaffold N50 (bp)**	**Contig N50 (bp)**	**Haploid chromosome number (n)**
Megaleporinus macrocephalus (GCA_021613375.1)	1,280,781,66	45,034,219	5,013,076	27
Colossoma macropomum (GCA_904425465.1)	1,221,809,066	40,163,545	5,645,235	-
Pygocentrus nattereri (GCA_015220715.1)	1,222,050,449	42,283,192	12,898,870	30
Astyanax mexicanus (GCA_000372685.2)	1,291,596,431	35,377,769	1,767,240	25

* Ungapped length

Table 4. Assembled and estimated chromosomes sizes (bp) calculated using karyotype data.

Chromosome	Assembled size (bp)	Estimated size (bp)
1	73,843,892	73,524,246
2	57,073,771	58,279,251
3	54,318,282	52,872,252
4	53,925,970	52,480,849
5	53,011,304	51,891,237
6	52,262,800	51,390,179
7	51,642,936	49,819,898
8	49,518,186	49,300,333
9	49,298,696	48,760,706
10	48,705,945	48,068,703
11	48,550,980	48,064,379
12	47,357,970	47,656,718
14	44,903,196	47,083,191
15	44,847,474	47,013,143
16	44,428,828	46,953,300
17	44,393,724	46,065,509

18	44,386,724	44,600,904
19	43,994,550	43,546,728
20	43,858,395	42,646,484
21	43,700,133	41,890,833
22	41,427,154	41,198,311
23	40,492,453	41,016,705
24	40,120,662	40,253,271
25	40,052,844	40,168,695
26	38,727,815	39,137,695
27	36,538,851	38,981,687
Pearson	**0,99**	

Table 5. Assembled and estimated chromosomes sizes (bp) of the sex chromosomes calculated using karyotype data.

Estimated size (bp)	**Z**	**W**
Euchromatic region	20,676,381	20,676,381
Heterochromatic region	29,933,760	50,525,046
Total	50,610,141	71,201,427

Table 6. Repeat annotation statistics for *Megaleporinus macrocephalus* genome.

	No. of elements	**Length (bp)**	**% of sequence**
Retroelements	325,354	86,900,511	6.78
SINEs:	32,263	4,249,898	0.33
Penelope	1,992	200,769	0.02
LINEs:	135,677	43,882,452	3.42
L2/CR1/Rex	109,233	36,159,292	2.82
R1/LOA/Jockey	538	129,887	0.01
R2/R4/NeSL	222	69,099	0.01
RTE/Bov-B	10,276	2,603,408	0.20
L1/CIN4	10,838	3,907,815	0.30
LTR elements:	157,414	38,768,161	3.02
BEL/Pao	2,206	845,394	0.07
Ty1/Copia	421	198,340	0.02
Gypsy/DIRS1	24,957	7,950,002	0.62
Retroviral	10,995	2,687,534	0.21
DNA transposons	817,030	151,501,452	11.82
hobo-Activator	295,340	56,805,942	4.43
Tc1-IS630-Pogo	253,381	56,461,503	4.40
PiggyBac	3,214	595,853	0.05
Tourist/Harbinger	34,271	6,807,497	0.53

Other (Mirage, P-element, Transib)	333	20,280	0.00
Rolling-circles	18,420	4248,326	0.33
Unclassified:	1379224	242,219,098	18.89
Total interspersed repeats:		480,621,061	37.49
Small RNA:	2375	321,707	0.03
Satellites:	88059	56,424,759	4.40
Simple repeats:	580511	52,069,796	4.06
Low complexity:	55387	5,150,938	0.40

Table 7. Comparison between the repeat content in the sex chromosomes and the autosomes of the *Megaleporinus macrocephalus* genome.

	% Sex chromosome	% Autosomes
Repeat content	50.95	46.71
Retroelements	7.86	6.78
DNA transposons	12.07	11.82
Unclassified	19.92	18.89
Total interspersed repeats	39.86	37.49
Small RNA	0.02	0.03
Satellites	6.64	4.40
Simple repeats	3.72	4.06
Low complexity	0.38	0.40

Table 8. Summary of the annotated features of *Megaleporinus macrocephalus* genome.

Feature	***Megaleporinus macrocephalus***
Genes and Pseudogenes	30,501
Protein-Coding	30,501
Exon	248,235
Intron	217,739
CDS	248,235
mRNA	30,501
Start codon	30,487
Stop codon	30, 488
Mean intron per gene	7.14
Mean exon per gene	8.14

Table 9. Summary of ddRAD sequencing statistics.

Library	Total sequences	Filtered sequences (%)				Retained reads (%)	Average retained reads/ ind. (million)
		Adapter sequence	No barcode	Low quality	No rad cutside		
1	198,679,472	1.19	3.67	2.37	1.33	91.44	3.95
2	191,095,010	1.18	4.92	2.43	1.66	89.82	3.73
3	183,389,588	1.15	4.36	2.59	1.86	90.04	3.59
4	188,028,608	1.17	4.38	2.53	1.81	90.12	3.68
5	165,196,674	1.09	4.53	2.42	1.70	90.27	3.24
6	207,410,674	1.15	6.49	2.32	1.86	88.18	3.98
7	173,700,306	1.15	6.88	2.76	2.07	87.14	3.29
Total	1,307,500,332	1.15	5.03	2.48	1.75	89.58	25.46

Table 10. Statistics of a few Neotropical fish species linkage maps.

Species	**Technique**	**No. of SNPs**	**Length (cM)**	**Average marker interval (cM)**	**Reference**
Megaleporinus macrocephalus	ddRADseq	11,231	3,320.36	0.29	This study
Colossoma macropomum	GBS*	7,734	2,811	0.39	Nunes *et al*., 2017
	RADseq	14,805	2,752	0.51	Varela et al., 2021
Piaractus mesopotamicus	RADseq	17,453	2,755.60	0.47	Mastrochirico-Filho *et al*., 2020

*Genotype-by-sequencing

Table 11. Summary of the genetic map of piauçu. Chr represents the chromosome which the linkage group had synteny with. Size is related to the length in bp, after scaffolding with Chromonomer. N is the number of markers. Length in cM, Density in cM/Locus.

LG	Chr	Size	n	Sex-averaged		Male		Female		M:F
				Length	Density	Length	Density	Length	Density	
1	6	23,599,497	710	126.42	0.18	135.77	0.19	131.68	0.19	1.03
2	4	45,953,104	688	130.95	0.19	127.07	0.18	107.22	0.16	1.19
3	1	49,978,048	625	133.56	0.21	139.93	0.22	130.11	0.21	1.08
4	15	40,070,618	543	120.91	0.22	139.69	0.26	133.3	0.25	1.05
5	10	43,303,417	476	122.43	0.26	106.81	0.22	100.3	0.21	1.06
6	18	34,104,434	484	112.29	0.23	135.78	0.28	115.72	0.24	1.17
7	2	50,146,178	489	129.75	0.27	156.05	0.32	121.16	0.25	1.29
8	9	44,463,537	503	107.79	0.21	116.85	0.23	115.32	0.23	1.01
9	25	36,288,099	429	135.49	0.32	137.46	0.32	135.06	0.31	1.02
10	23	36,576,158	501	122.64	0.24	152.38	0.3	117.92	0.24	1.29
11	16	32,187,476	444	111.72	0.25	108.1	0.24	80.47	0.18	1.34
12	8	43,378,694	442	127.77	0.29	136.22	0.31	99.65	0.23	1.37

13	11	36,977,433	473	121.11	0.26	128.1	0.27	124.15	0.26	1.03
14	12	44,454,369	406	141.52	0.35	142.35	0.35	135.23	0.33	1.05
15	27	35,428,719	381	119.88	0.31	120.71	0.32	119.61	0.31	1.01
16	17	38,667,919	372	139.89	0.38	139.95	0.38	141.94	0.38	0.99
17	19	25,433,675	385	121.22	0.31	124.13	0.32	133.74	0.35	0.93
18	24	19,470,523	339	119.57	0.35	121.55	0.36	121.27	0.36	1
19	22	27,071,565	351	119.33	0.34	137.5	0.39	121.55	0.35	1.13
20	3	38,970,815	296	82.15	0.28	94.36	0.32	95.58	0.32	0.99
21	7	18,150,482	253	81.57	0.32	96.9	0.38	92.98	0.37	1.04
22	20	39,059,067	236	143.08	0.61	142.45	0.6	130.32	0.55	1.09
23	21	26,787,905	250	107.64	0.43	111.77	0.45	104.11	0.42	1.07
24	13	4,059,682	225	43.25	0.19	47.37	0.21	87.81	0.39	0.54
25	5	23,599,497	237	132.10	0.56	150.95	0.64	139.78	0.59	1.08
26	14	45,953,104	239	136.79	0.57	135.83	0.57	147.43	0.62	0.92
27	13	22,889,721	251	120.66	0.48	117.89	0.47	113.62	0.45	1.04
28	26	29,510,817	203	108.88	0.54	114.34	0.56	104.98	0.52	1.09

Total	977,082,963	11,231	3320.36	0.29	3,518.24	0.31	3,301.97	0.29	1.07

Table 12. Summary of raw sequencing data produced through pool-sequencing of males and females of *Megaleporinus macrocephalus* and *Leporinus friderici*.

Species	Pool	Total reads	Total data (Gb)
M. macrocephalus	Male	236,328,878	35.45
	Female	246,536,180	36.98
L. friderici	Male	432,269,160	64.84
	Female	426,001,092	63.90

Table 13. Summary of RNA-sequencing statistics.

Pool	Replicate	Total reads	Retained reads (%)	Total data (Gb)
	1	46,331,866	97.72	4.53
Male	2	50,141,272	97.93	4.91
	3	41,915,706	100.00	4.44
	1	50,886,670	97.58	4.97
Female	2	48,466,868	97.21	4.71
	3	48,337,546	97.33	4.70
	Total	286,079,928	97.91	28.26

Table 14. Summary of the differentially expressed, up-regulated (LFC > 2) and down-regulated (LFC < -2) transcripts in gonads of males and females of *Megaleporinus macrocephalus*.

vs	**D.E** (***p*adj ≤ 0.05**)	**Up-regulated** (**LFC ≥ 2**)	**Down-regulated** (**LFC ≤ -2**)
All	9,740	3,879	2,683
Autosomes	9,429	3,701	2,621
Sex chromosome	311	170	58

Table 15. Differentially expressed sex-determination-related genes on the sex chromosome of *Megaleporinus macrocephalus*.

Gene ID	Position (Mb)	LFC	*padj*	Transcript count	
				Male	Female
CCDC114 Coiled-coil domain containing 114	1.54	-5.50	3.92E-2	2,745	9,149
IGFBP6 Insulin-like growth factor binding protein	6.38	5.19	2.21E-20	7,154	195
SOX13 SRY (sex-determining region Y)-box 13	12.57	-0.50	0.49	331	404
SOX12 SRY (sex-determining region Y)-box 12	20.98	3.45	4.46E-03	69	7
FOXP4 Forkhead box	23.57	2.72	6.59E-05	142	22
BMP7 Bone morphogenetic protein	27.72	-1.64	1.37E-02	77	246
AMHR2	38.82	3.80	4.28E-12	2,372	178

Anti-Mullerian hormone receptor, type II

8.4.17 Figures

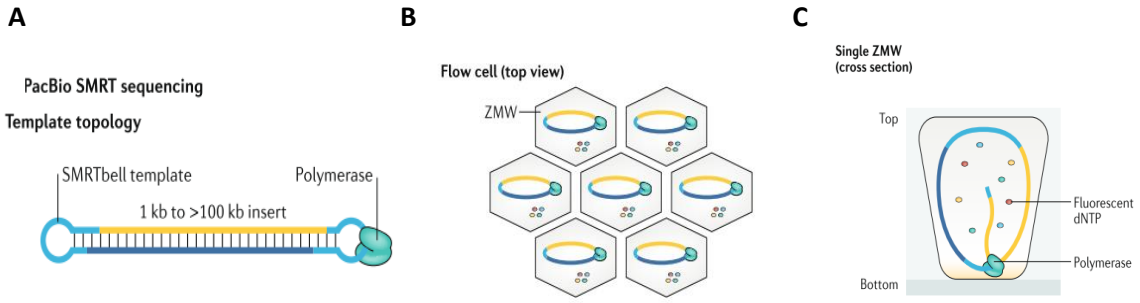


Figure 1. PacBio SMRT sequencing. SMRTbell (A). SMRT Cell (B). Zero Mode Waveguide (ZMW) (C). Adaptado de Logsdon et al. (2020).

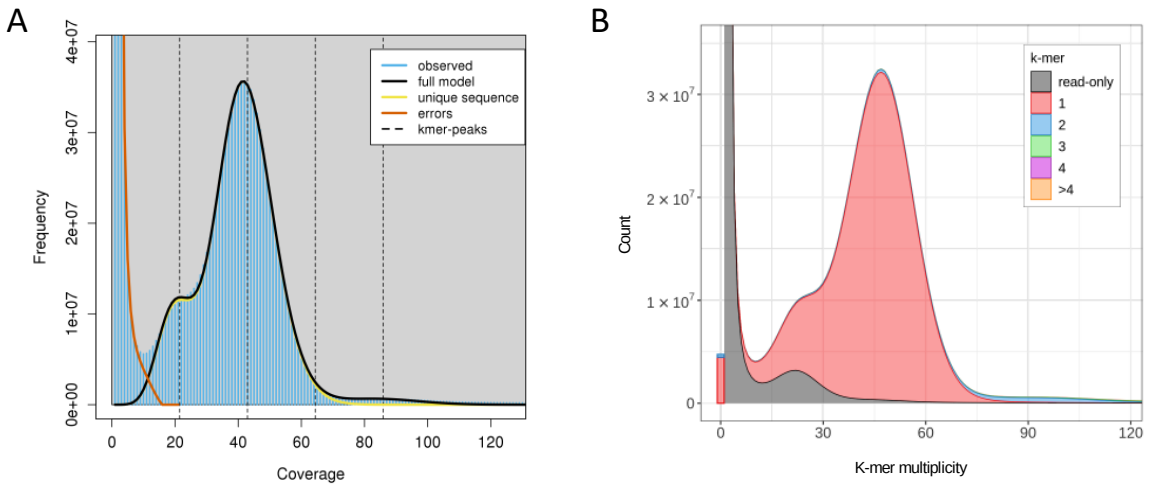


Figure 2. K-mer profile of MGISEQ short reads (A). A k-mer analysis of the *Megaleporinus macrocephalus* genome bases against its sequenced MGISEQ reads (B).

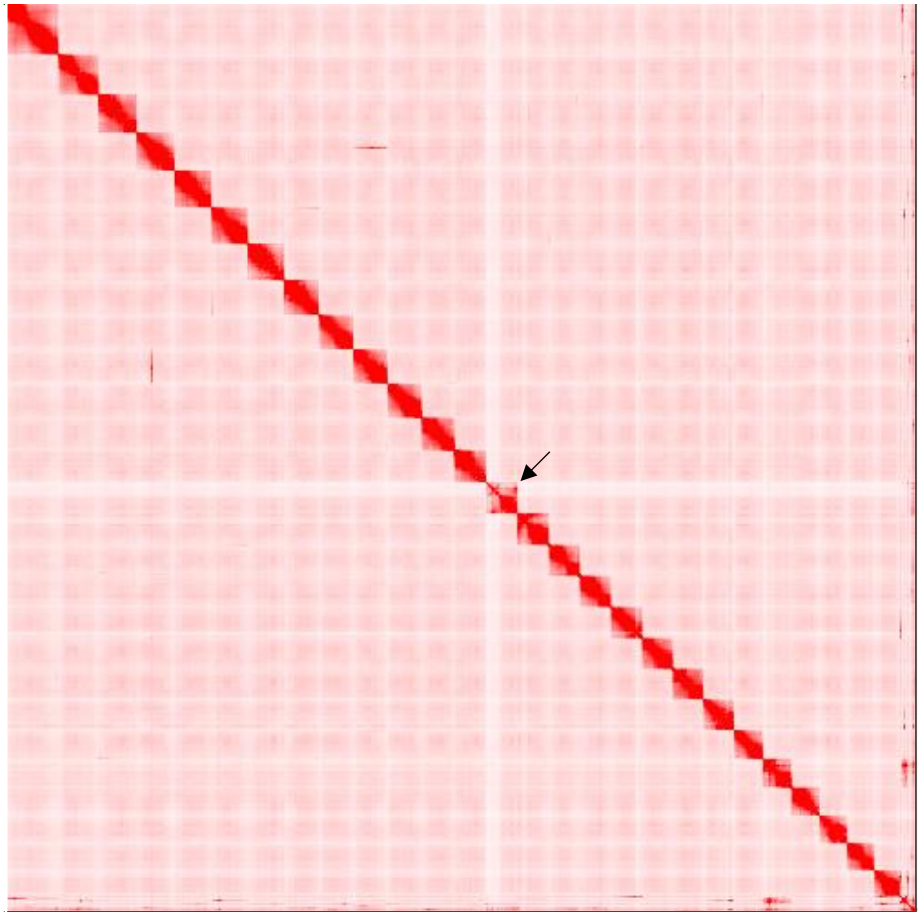


Figure 3. Hi-C contact map highlighting the sex chromosome (arrow) of *Megaleporinus macrocephalus.*

Figure 4. Karyotype of a female of *Megaleporinus macrocephalus* under C-banding.

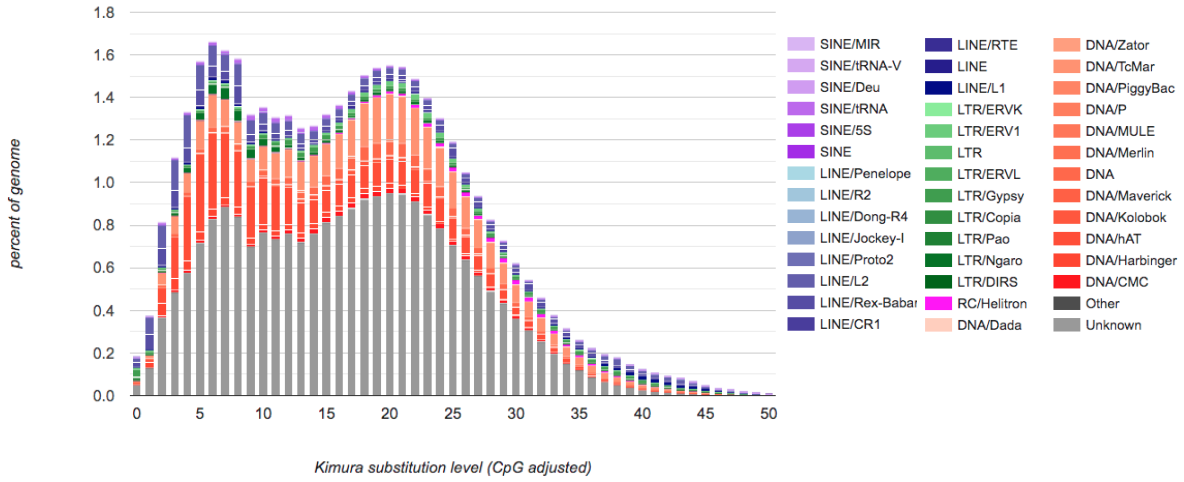


Figure 5. Interspersed repeat landscape of *Megaleporinus macrocephalus* genome. The graph represents genome coverage (*y* axis) for each type of TEs (DNA transposons, SINE, LINE, and LTR retrotransposons), clustered according to Kimura distances (*K*-value; Kimura, 1980) to their corresponding consensus sequence (*x* axis, *K*-values from 0 to 50).

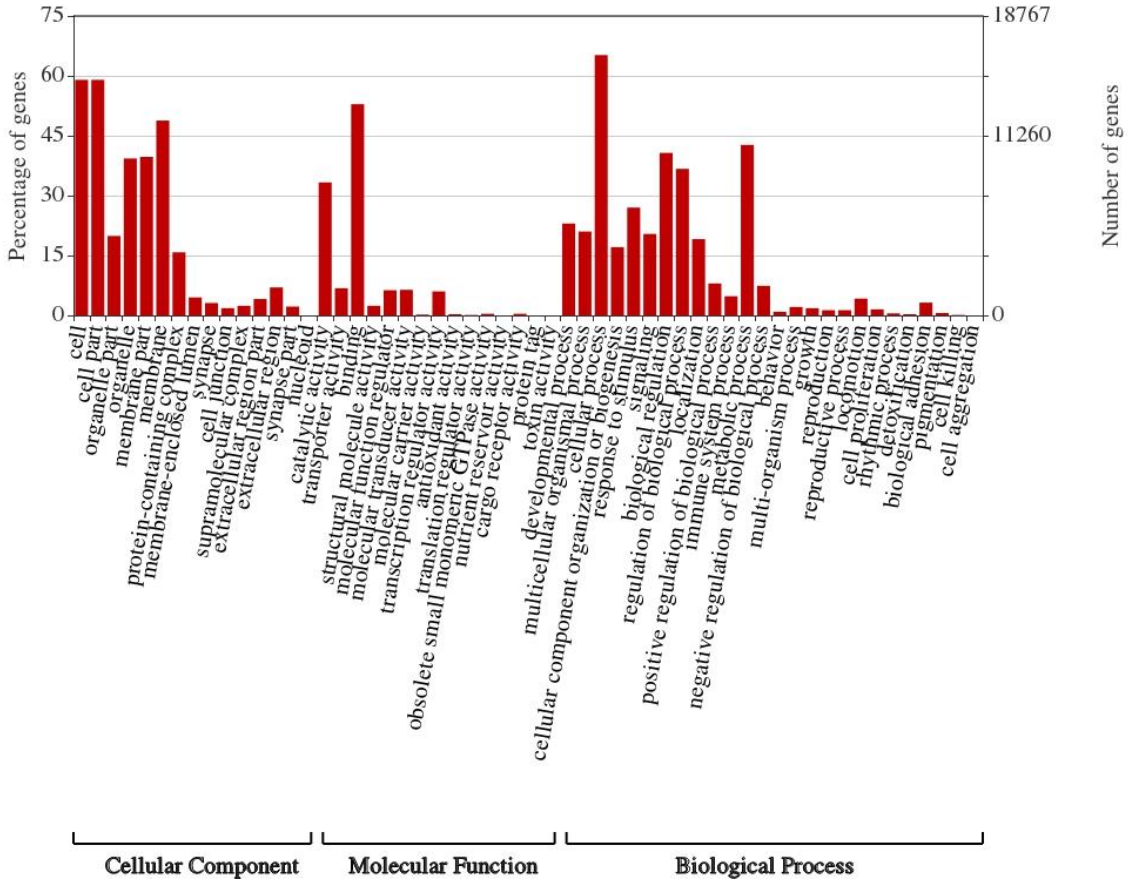


Figure 6. Gene Ontology (G.O) Terms of Cellular Component, Molecular Function and Biological Process domains of *Megaleporinus macrocephalus* genome.

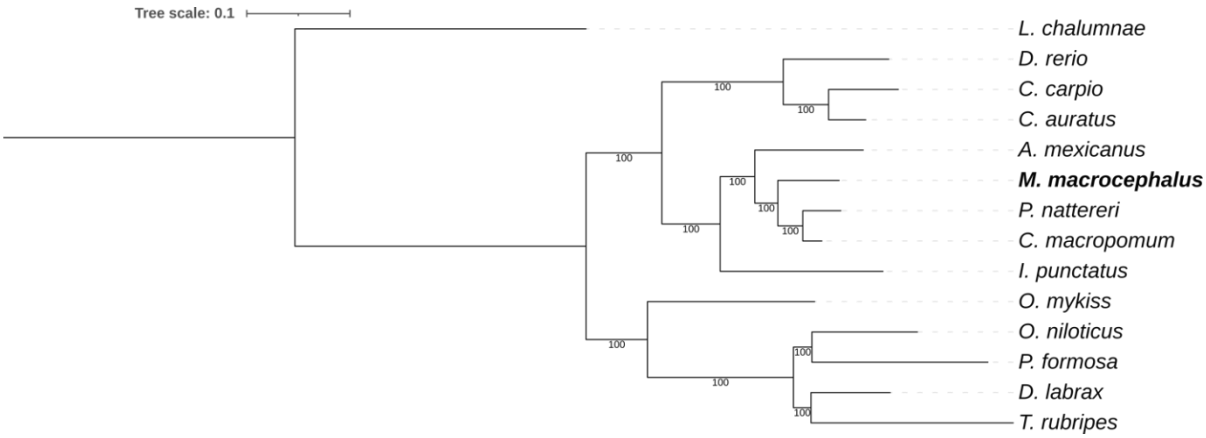


Figure 7. Phylogeny of single-copy orthologs genes (n=336) of *Megaleporinus macrocepha*lus and other fish species using the criteria of Maximum Likelihood.

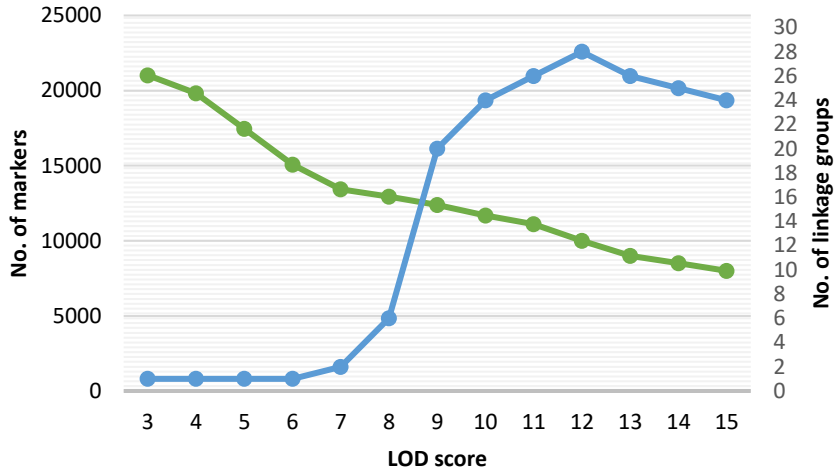


Figure 8. The average number of markers in linkage groups (left y axis) and the number of linkage groups (right *y* axis) according to LOD score (*x* axis).

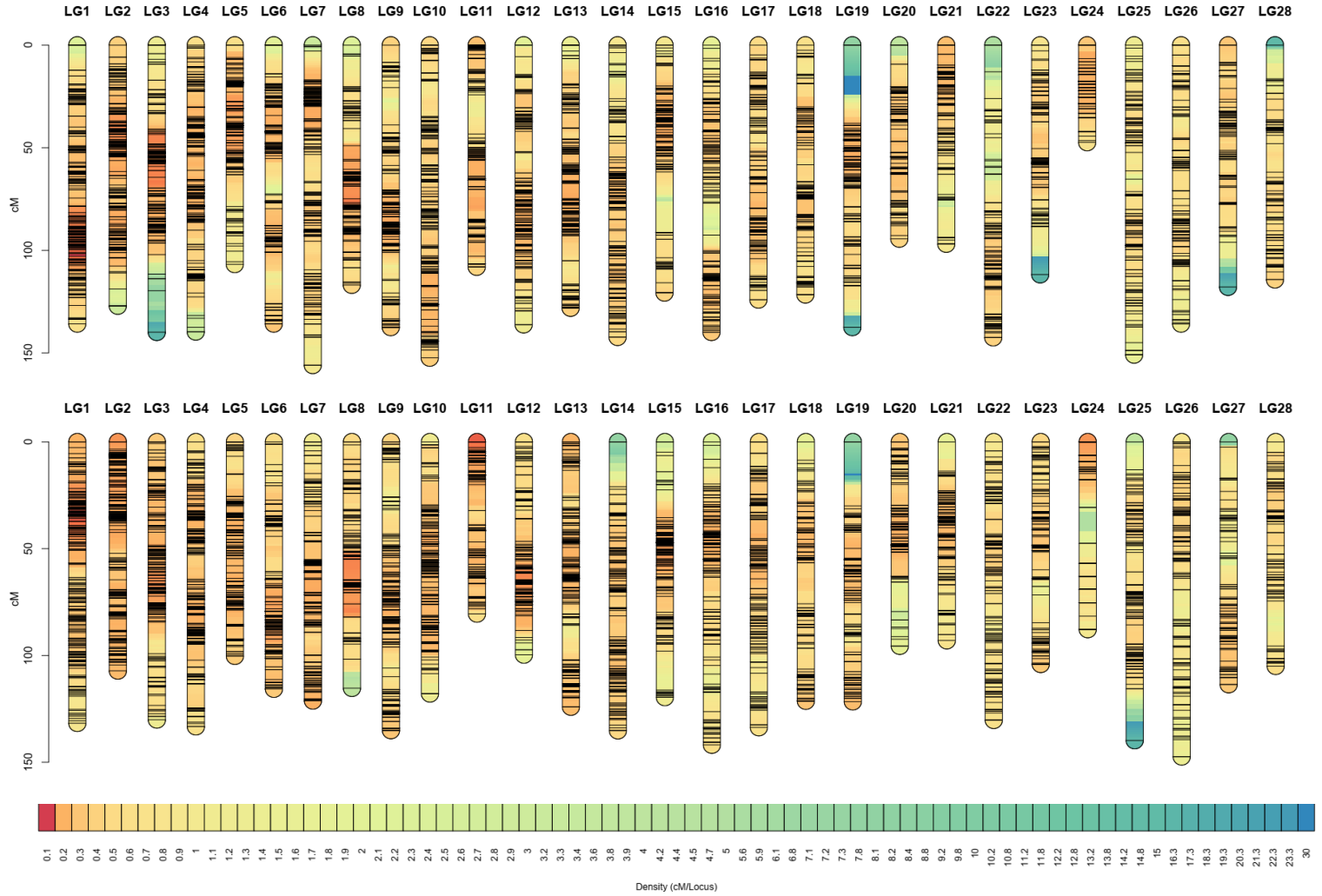
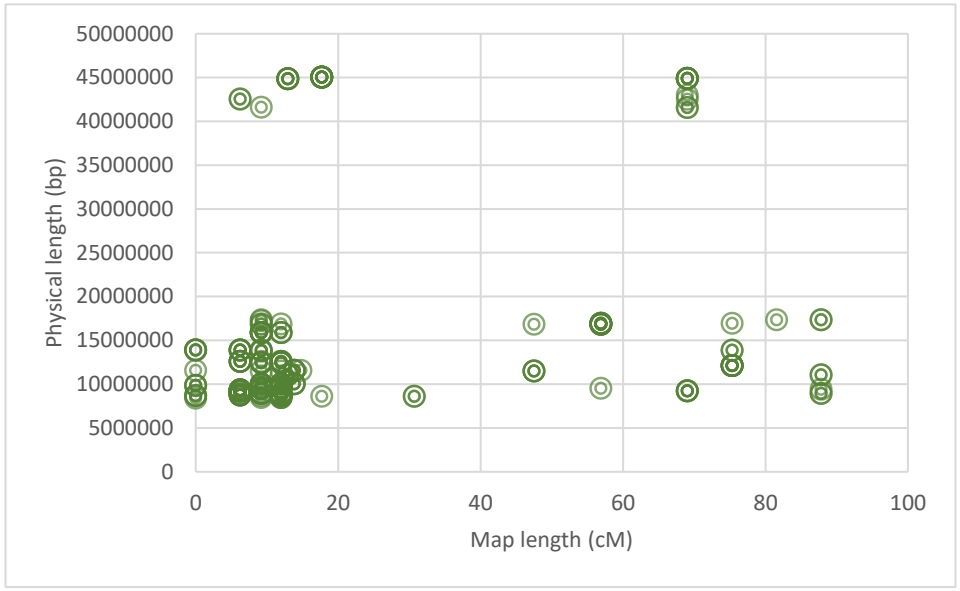


Figure 9. Male (A) and female (B) linkage maps of *Megaleporinus macrocephalus* showing 28 linkage groups and 11,231 SNPs. The density of markers is represented by a range of different colours that vary from blue (low density regions) to red (high density regions).



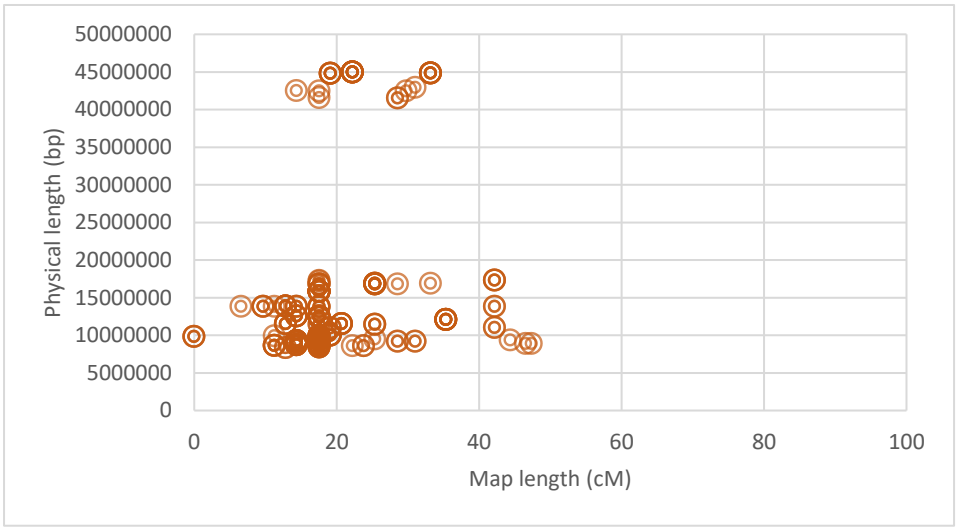


Figure 10. Female and male SNP markers of *Megaleporinus macrocephalus* distributed throughout LG24 based on the genetic map (cM) versus physical location in the genome (bp). Darker and lighter circles mean higher and lower abundance of SNPs.

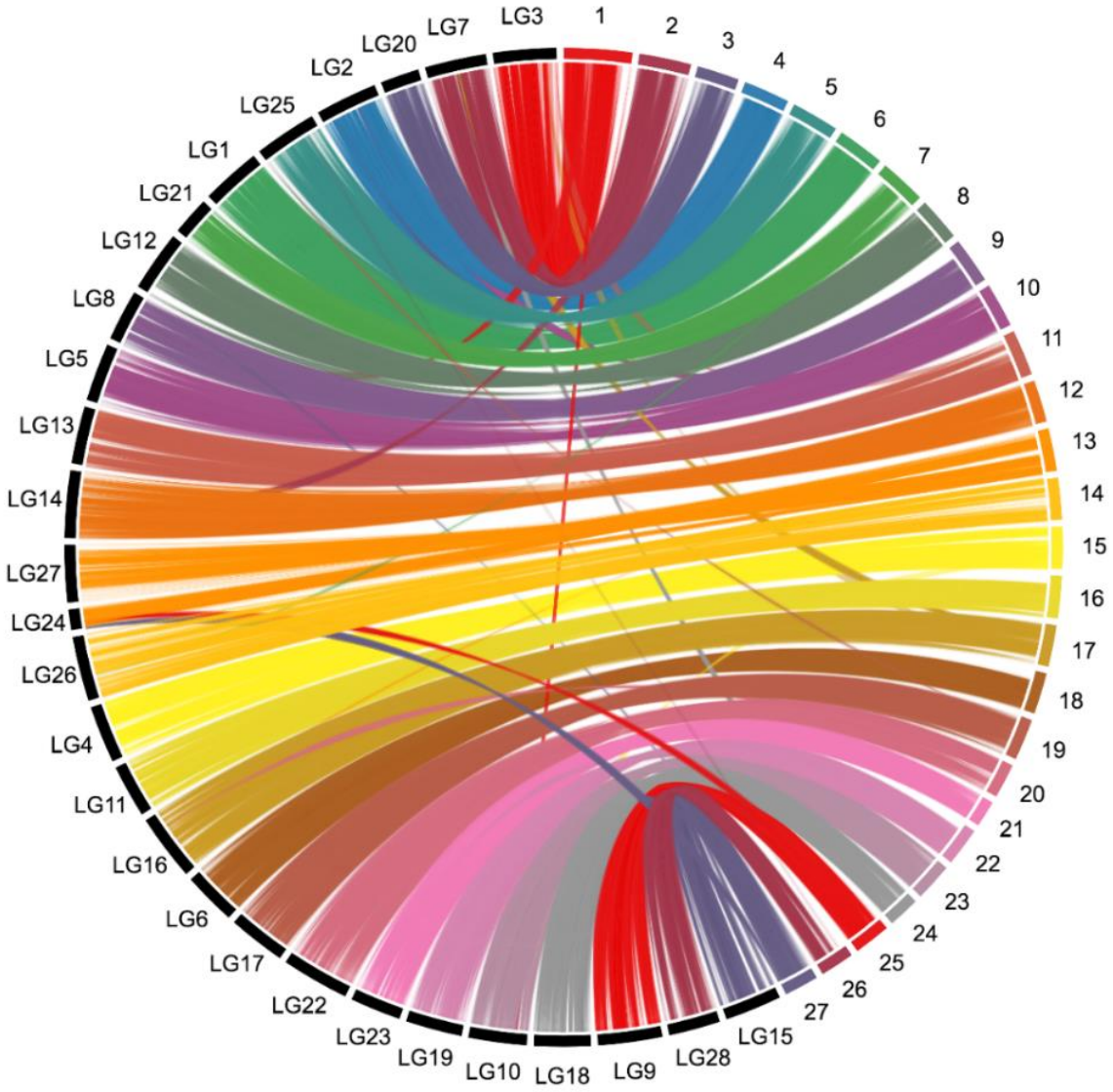


Figure 11. Genomic synteny for each pair of alignments between the linkage groups of the genetic map and the chromosomes of *Megaleporinus macrocephalus*.

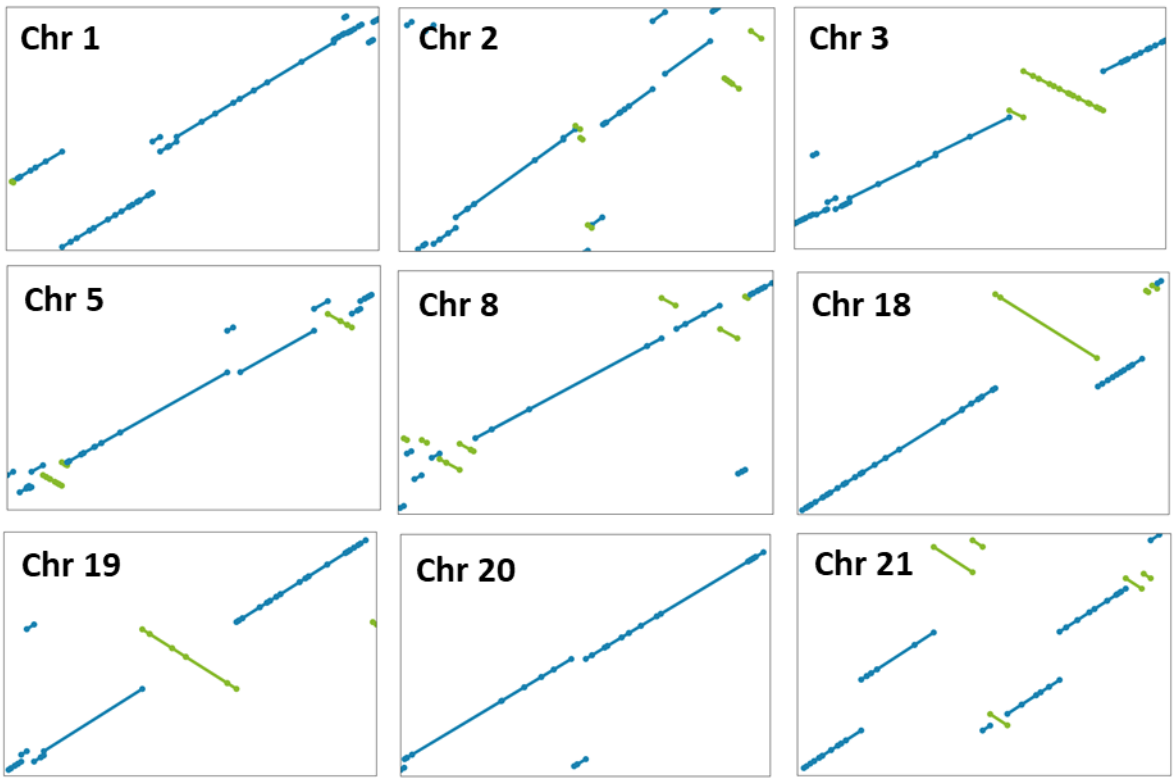


Figure 12. Dotplot synteny between chromosomes constructed with Hi-C data (*x*-axis) and scaffolds of the linkage groups (*y*-axis). In blue, forward alignments; and in green, reverse alignments (inversions). The dots represent the end of scaffolds.

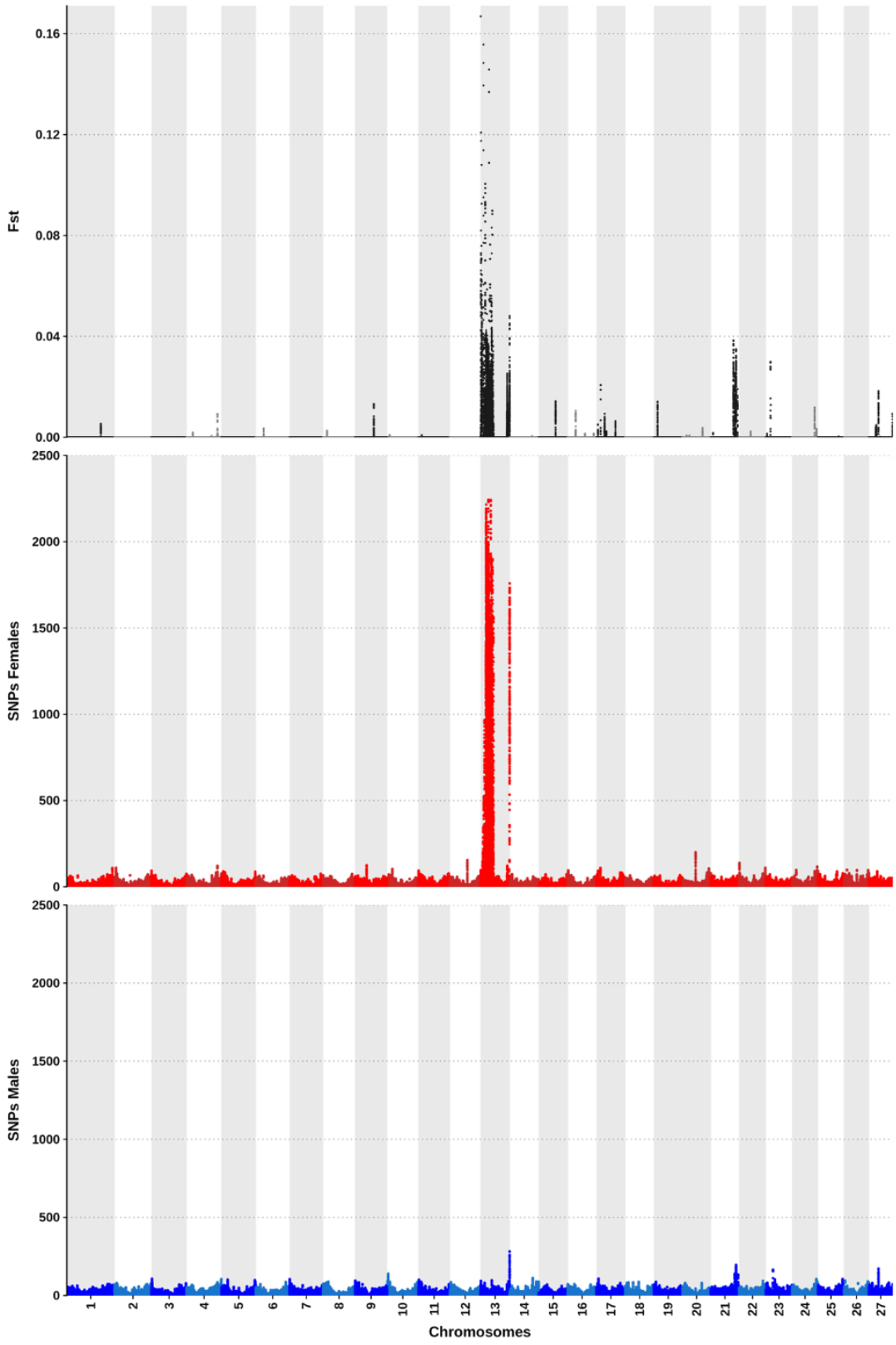


Figure 13. Manhattan plots of FST, female and male-specific SNPs, respectively, accounted through windows of 50 kb along the 27 chromosomes of *Megaleporinus. macrocephalus* genome.

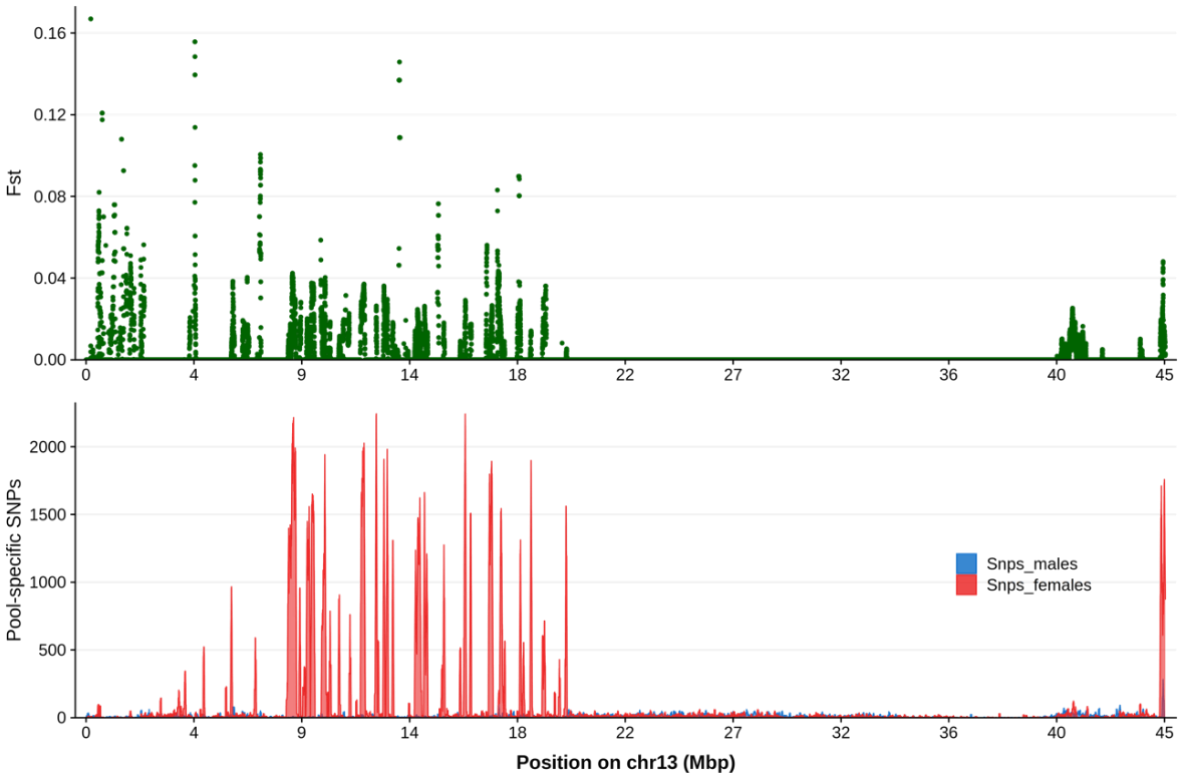


Figure 14. FST and pool-specific SNPs along the sex chromosome of *Megaleporinus macrocephalus.*

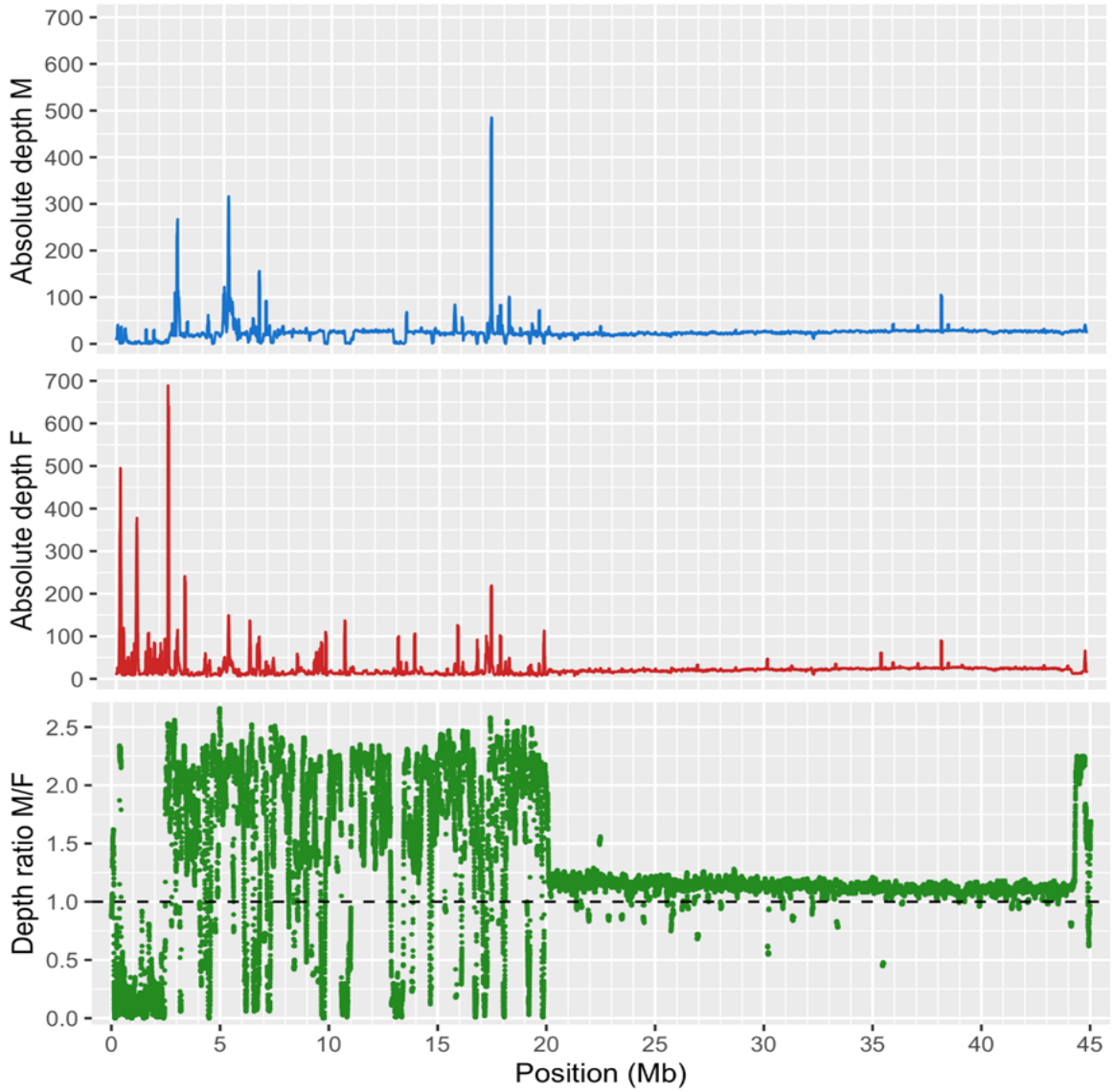


Figure 15. Absolute depth (read coverage/ 50 kb) of males and females pools and depth ratio (absolute depth males/ absolute depth of females) along the sex chromosome of *Megaleporinus macrocephalus*. Depth ratio of = 1 indicates equal read coverage in males and females (dashed line), while depth ratio > 1 and < 1 mean superior coverage in males ZZ (duplicated chromosome) and females ZW (single copy of each chromosome), respectively.

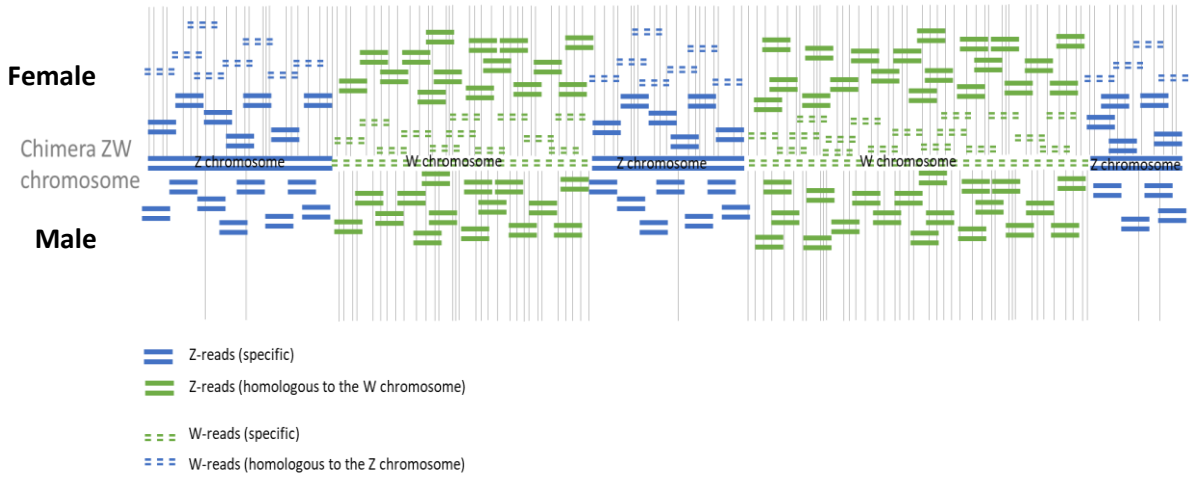


Figure 16. Chimeric region of the sex chromosome of *Megaleporinus macrocephalus*. In the areas where the reads are not properly aligned to the genome reference sequence (*e.g.* Z-reads aligned to a fragment of the W chromosome or W-reads reads aligned to a fragment of the Z chromosome), we note a high concentration of SNPs due to the sequence divergence. Grey bars mean SNPs.

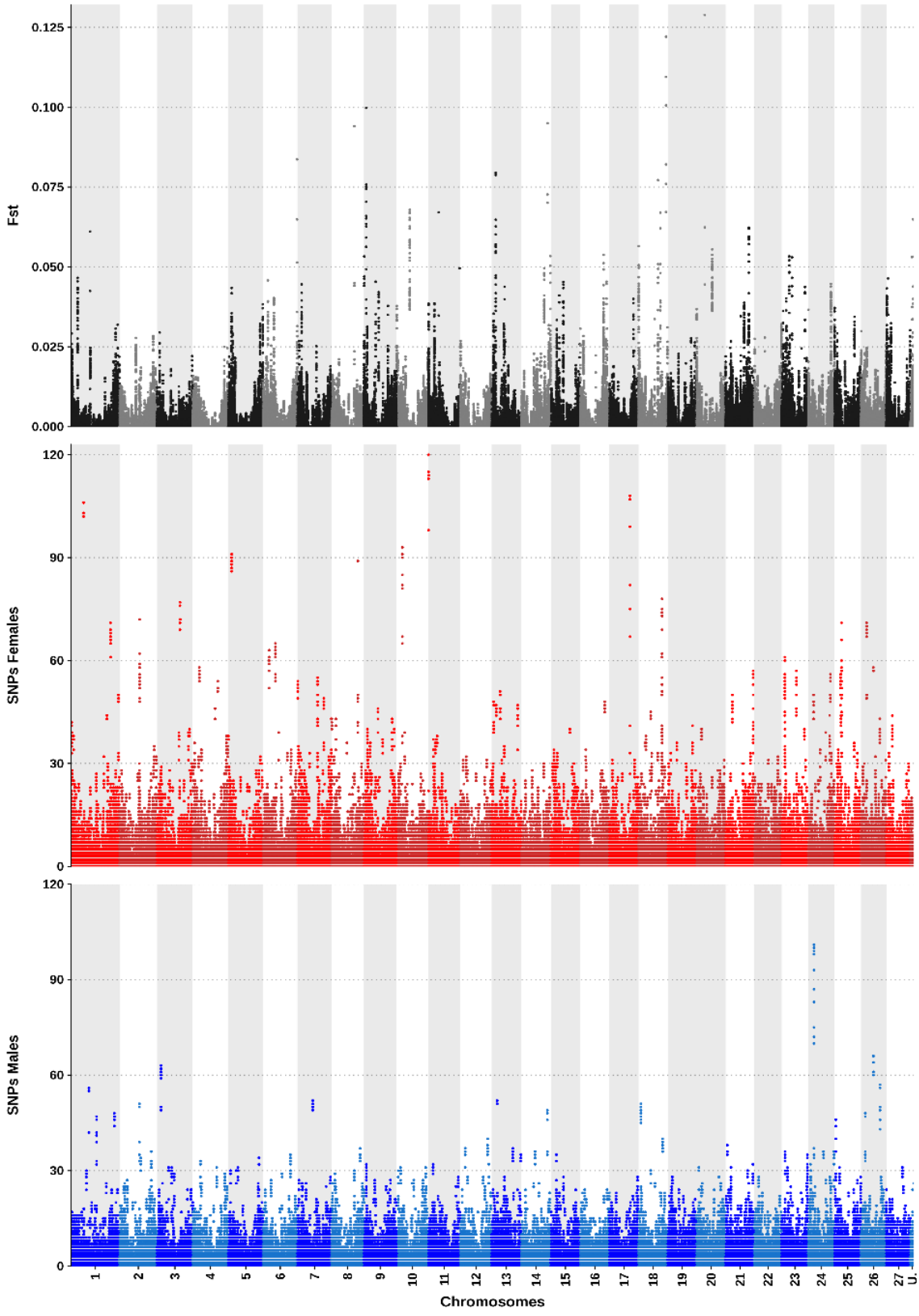


Figure 17. Manhattan plots of FST, female and male-specific SNPs, respectively, of *Leporinus friderici* reads aligned to *Megaleporinus macrocephalus* genome.

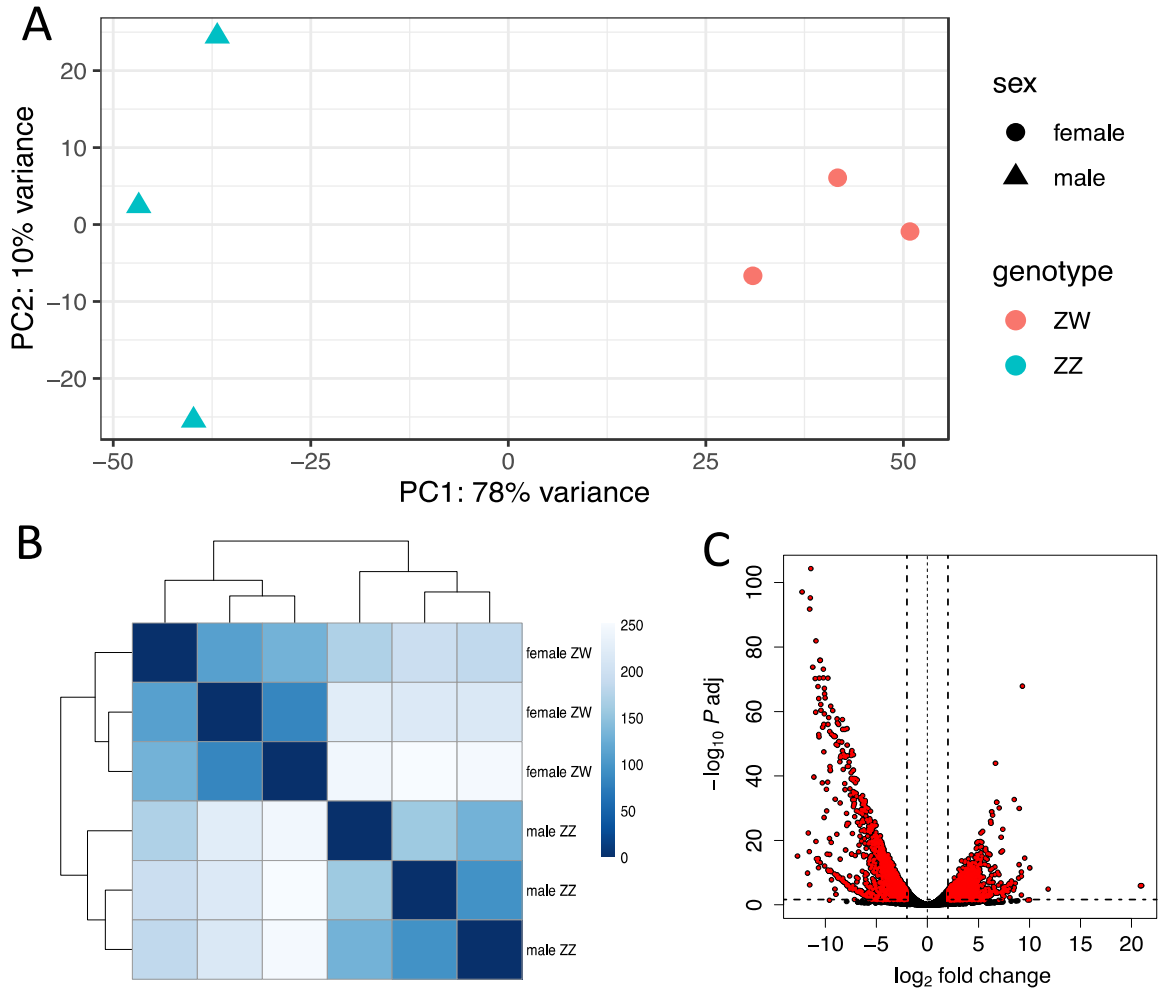


Figure 18. RNA-Seq samples clustered according to their transcript expression. Principal components analysis (PCA) (A) and Euclidean distance matrix (B). Volcano plot showing down and up-regulated transcripts in red on the left and right, respectively. Not significantly expressed transcripts (*p*adj > 0.05) in black (C).

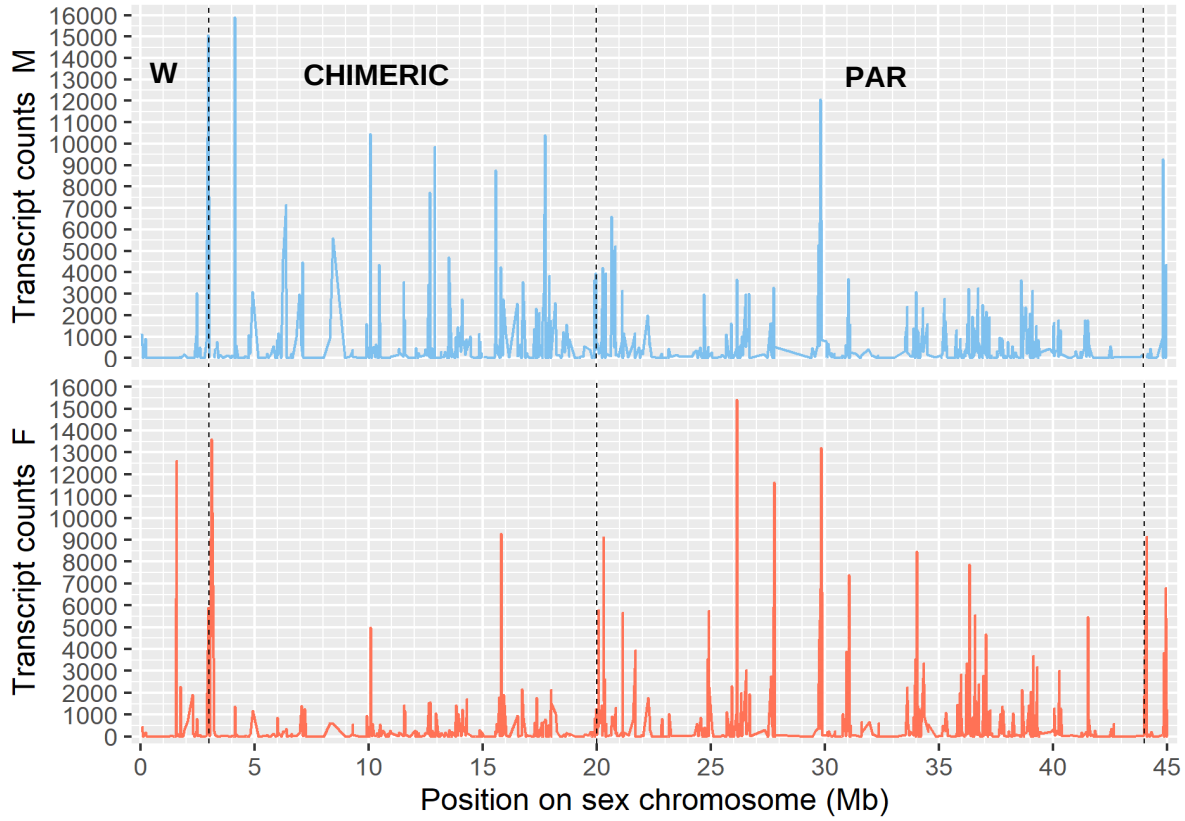


Figure 19. Female ZW and male ZZ transcript counts in the distinct regions of the sex chromosome: putative W-specific region (W), chimeric region of Z and W (chimeric) and pseudo-autosomal region (PAR).

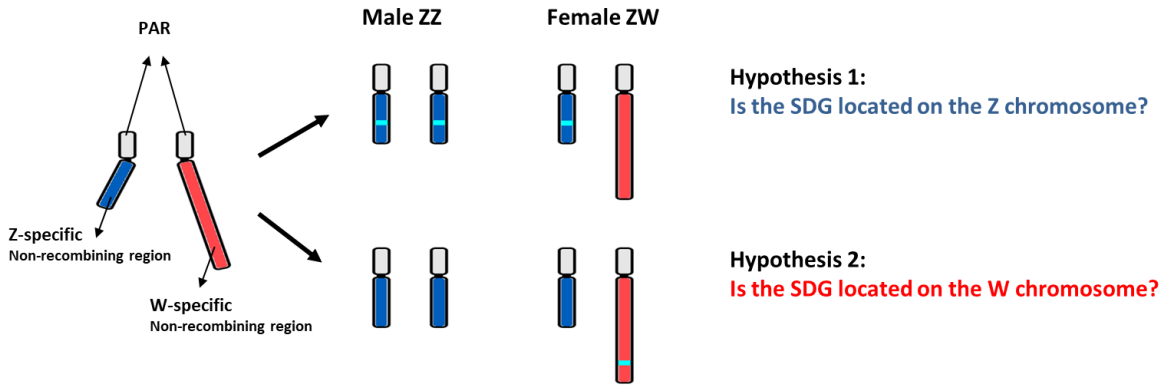


Figure 21. Hypothesis regarding the localization of the sex-determination gene in *Megaleporinus macrocephalus*.

9. Consideraçãoes Finais

O presente trabalho forneceu um genoma de alta qualidade e em nível cromossômico de uma espécie de peixe neotropical. Devido à escassez deste tipo de recurso nas bases de dados atuais, ele constitui uma ferramenta importante para estudos genéticos e genômicos neste grupo.

Fizemos a montagem do primeiro cromossomo sexual de uma espécie neotropical, que permitiu concluir alguns aspectos de sua evolução e constitui um avanço no conhecimento disponível sobre a evolução de cromossomos sexuais em vertebrados.