



**MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL RURAL DA AMAZÔNIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA APLICADA À
AGROPECUÁRIA**

ÁDRIA DIVA DOS SANTOS PEREIRA

**DESENVOLVIMENTO E CARACTERIZAÇÃO DE MARCADORES
MICROSSATÉLITES EM *SALMINUS BRASILIENSIS***

**BELÉM - PA
2024**

ÁDRIA DIVA DOS SANTOS PEREIRA

**DESENVOLVIMENTO E CARACTERIZAÇÃO DE MARCADORES
MICROSSATÉLITES EM *SALMINUS BRASILIENSIS***

Dissertação apresentada à Universidade Federal Rural da Amazônia, como parte das exigências do curso de mestrado em Biotecnologia Aplicada à Agropecuária: área de concentração em Biotecnologia, para a obtenção do título de Mestre.

Área de concentração: Biotecnologia Animal

Orientador: Prof. Dr^a Marília Danyelle Nunes Rodrigues.

**BELÉM
2024**

Dados Internacionais de Catalogação na Publicação (CIP)
Bibliotecas da Universidade Federal Rural da Amazônia
Gerada automaticamente mediante os dados fornecidos pelo(a) autor(a)

P436 Pereira, Ádria Diva dos Santos
DESENVOLVIMENTO E CARACTERIZAÇÃO DE MARCADORES
MICROSSATÉLITES EM
SALMINUS BRASILIENSIS / Ádria Diva dos Santos Pereira. - 2024.
43 f. : il.

Dissertação (Mestrado) - Programa de PÓS-GRADUAÇÃO em Biotecnologia Aplicada à Agropecuária (PPGBAA), Campus Universitário de Belém, Universidade Federal Rural Da Amazônia, Belém, 2024.

Orientador: Profa. Dra. Marília Danyelle Nunes Rodrigues

1. Salminus brasiliensis - Dourado. 2. marcadores moleculares. 3. repetições de sequência simples. 4. microssatélites. I. Rodrigues, Marília Danyelle Nunes, *orient.* II. Título

CDD 660.6

ÁDRIA DIVA DOS SANTOS PEREIRA

**DESENVOLVIMENTO E CARACTERIZAÇÃO DE MARCADORES
MICROSSATÉLITES EM *SALMINUS BRASILIENSIS***

Dissertação apresentada à Universidade Federal Rural da Amazônia, como parte das exigências do Curso de Mestrado em Biotecnologia Aplicada à Agropecuária, para obtenção do título de Mestre. Área de concentração: Biotecnologia Animal.

Data da aprovação: 29/02/2024

Banca Examinadora

_____ Orientador

Prof. Dra. Marília Danyelle Nunes Rodrigues (ISARH – UFRA)

_____ Membro Interno

Prof. Dra. Regianne Maciel dos Santos Correa (ISARH – UFRA)

_____ Membro Interno

Prof. Dr. Leonardo Castelo Branco Carvalho (ISARH – UFRA)

_____ Membro Externo

Prof. Dr. Oscar Alegria (ICB - UFPA)

_____ Membro Suplente

Prof. Dr. Marcos de Barros Braga (UFRA- Paragominas)

AGRADECIMENTOS

Primeiro, o meu maior agradecimento é para a espiritualidade e aos meus orixás que me mantiveram forte, e me ajudaram a chegar a finalização da pesquisa.

A queridíssima prof^a Marília Rodrigues que aceitou ser orientadora, onde me propôs um desafio e em todas as minhas inseguranças me passou segurança, durante todo processo foi uma mulher incrível fazendo eu desenvolver um carinho imenso.

Aos membros da Banca Examinadora, pelo aceite em participar da avaliação desta dissertação contribuindo com suas sugestões e críticas construtivas.

Ao Programa de Pós-Graduação de Biotecnologia Aplicada à Agropecuária da UFRA, pela ajuda e apoio.

Aos meus pais, Maria Fernanda e Raimundo Silva, que sempre me apoiam na minha caminhada, e que nunca deixaram eu desistir e me deram todo apoio que foi necessário.

Aos todos amigos de turma do mestrado, pelo convívio, compartilhamento de conhecimentos e principalmente pela ajuda nas disciplinas.

E a todos que de alguma forma, direta ou indireta, contribuíram para a realização deste trabalho.

RESUMO

Salminus brasiliensis é uma espécie endêmica do rio Paraná e Lagoa dos Patos, englobando cinco países que inclui Argentina, Bolívia, Brasil, Paraguai e Uruguai. É uma espécie de grande importância econômica e social na pesca artesanal, de subsistência e esportiva. O presente trabalho objetivou o desenvolvimento de marcadores moleculares do tipo microssatélites na espécie *Salminus brasiliensis*. Inicialmente, o genoma da espécie foi obtido por meio do sequenciamento de nova geração pela plataforma Illumina. A qualidade das reads foram analisadas pelo programa FastQc. A montagem *de novo* foi realizada pelos programas SOAPdenovo2 e SPAdes e a mineração de microssatélites realizada utilizando os programas MISA e EasySSR. Para determinação de *primers* foi utilizado o programa Batchprimers3. O total de *reads* sequenciadas, foi de 4.119.94 com comprimento de 36-100 pb e Q=39,06. Com o programa SOAPdenovo2 foi obtido um N50=731 e com SPAdes um N50=1649. Os resultados do programa MISA apontaram um total de 56.893 microssatélites, predominando os pentanucleotídeos com 63,7%, e o programa EasySSR o total de microssatélites foi de 29.098, predominando trinucleotídeos com 46,3%. Foram desenhados um total de 89 primers, dentre eles, 80 primers bem-sucedidos. Portanto, através do sequenciamento de nova geração pela plataforma Illumina foram desenvolvidos marcadores microssatélites para futuros estudos de polimorfismos genéticos em *Salminus brasiliensis*, onde também primers foram desenhados e testados *in silico* para nortear futuras pesquisas.

Palavras-Chave: repetições de sequência simples; marcadores moleculares; *Salminus brasiliensis*

ABSTRAT

Salminus brasiliensis is a species of great economic and social importance in artisanal, subsistence, and sport fishing. It is endemic to the Paraná River and Lagoa dos Patos, encompassing five countries including Argentina, Bolivia, Brazil, Paraguay, and Uruguay. The aim of this study was to develop microsatellite molecular markers for this species. Its genome sequence was initially obtained through next-generation sequencing using an Illumina platform. Read quality was analyzed using the FastQC software. *De novo* assembly was carried out using SOAPdenovo2 and SPAdes programs and microsatellite mining using the MicroSatellite Identification Tool (MISA) and EasySSR programs. Batchprimers3 was employed to design the polymerase chain reaction (PCR) primers. The total number of sequenced reads was 411,994 with a length of 36–100 base pairs and $Q = 39.06$. SOAPdenovo2 yielded an $N_{50} = 731$, whereas SPAdes yielded an $N_{50} = 1,649$. A total of 56,893 microsatellites, with pentanucleotides predominating at 63.7%, were identified using the MISA program, whereas a total of 29,098 microsatellites, with trinucleotides predominating at 46.3%, were identified using the EasySSR program. A total of 89 PCR primers were designed, of which 80 were successful. Thus, next-generation sequencing using the Illumina platform allowed the development of microsatellite markers for future genetic polymorphism studies in *S. brasiliensis*, where primers were also designed and tested in silico to guide future research.

Keywords: simple sequence repeats; molecular markers; *Salminus brasiliensis*.

LISTA DE FIGURAS

Figura 1. Espécie <i>Salminus Brasiliensis</i> em seu habitat natural, no Mato grosso do Sul.....	14
Figura 2. Equipamentos utilizados para sequenciamento de DNA. 1ª geração (A) Sanger; 2ª geração (B) 454-Roche, (C) Illumina HiSeq, (D) Illumina MiSeq, (E) Solid; 3ª geração (F) Ion Torrent, (G) PacBio, (H) Nanopore	25
Figura 3. Classificação dos microssatélites.....	29

LISTA DE TABELAS

Tabela 1. Classificação de marcadores microssatélites.....	27
Tabela 2. Parâmetros gerais de microssatélites encontrados no genoma de <i>Salminus brasiliensis</i> (MISA).....	32
Tabela 3. Parâmetros gerais de microssatélites encontrados no genoma de <i>Salminus brasiliensis</i> (EasySSR).....	33
Tabela 4. Identificação dos microssatélites em porcentagem no EasySSR e MISA.....	33
Tabela 5. Lista de 64 <i>primers</i> de SSR desenvolvidos para <i>Salminus brasiliensis</i>	36

SUMÁRIO

1 INTRODUÇÃO	12
2 OBJETIVOS	14
2.1 Objetivo geral.....	14
2.2 Objetivos específicos	14
3 REVISÃO DE LITERATURA	15
3.1 Biologia da espécie	15
3.2 Importância econômica da espécie	17
3.3 Biotecnologia aplicada a espécie	18
3.4 Sequenciamento de próxima geração (NGS)	20
3.4.1 - 1ª GERAÇÃO (Degradação química e Interrupção da cadeia-Sanger).....	20
3.4.2 - 2ª GERAÇÃO (454-Roche, Illumina Genome Analyser-HiSeq/MiSeq e SOLID)	20
3.4.2.1 - Roche 454.....	20
3.4.2.2 AB Solid.....	21
3.4.2.2 - Sistema Illumina GA e HiSeq/MiSeq.....	22
3.4.3 - 3ª GERAÇÃO (Ion Torrent e PacBio RS).....	23
3.4.3.1 - Ion PGM de Ion Torrent	23
3.4.3.2 - PacBio.....	24
3.4.4.3 - Nanopore.....	25
3.5 Marcadores moleculares	26
3.5.1 Marcador Microsatélite	28
3.6 Estratégias de melhoramento genético para <i>Salminus brasiliensis</i>	30
4 MATERIAL E MÉTODO	31
4.1 Aspectos éticos e coleta de amostras	31
4.2 Pré-processamento de dados e montagem de novo do genoma	31
4.3 Identificação de microsatélites	31
4.4 Desenho de primers e PCR in silico.....	31
5 RESULTADOS.....	32
5.1 Sequenciamento	32
5.2 Parâmetros de SSRs	32
5.3 Caracterização de microsatélites	33
6. DISCUSSÃO	36
7. CONCLUSÃO	39
REFERÊNCIAS.....	40

1 INTRODUÇÃO

Salminus brasiliensis de nomes vulgares “dourado”, “pirayú”, “saipé”, “pirajuba” e “dama”, é uma espécie de peixe de grande porte, cujas fêmeas podem atingir em torno de 26 Kg, a espécie possui características próprias para realização de migrações de longas distâncias e para predação de outros peixes. Além disso, é uma espécie apreciada pelo mercado consumidor mediante as características sensoriais de sua carne como sabor e suculência, além do peso bruto do animal conferir alto rendimento para a produção de filé (GRACIANO *et al.*, 2022).

Dessa forma, *S. brasiliensis* é objeto de grande interesse para criações em cativeiro, apesar de ser predominantemente oriunda de extrativismo para comercialização. A espécie é nativa da Bacia do Rio da Prata, ocorrendo nos rios Paraná, Paraguai e Uruguai, no Brasil, está presente em estados das regiões Sul, Sudeste e Centro-Oeste (GRACIANO *et al.*, 2022).

S. brasiliensis, a espécie nativa ainda não possui genoma completo publicado, mas 8 loci de microssatélites polimórficos foram descritos em por Rueda *et al.* (2011). Diante disso, é possível observar que ainda há carência de dados genéticos para a espécie, o que torna limitado o estabelecimento de programas de manejo sustentável para conservação e produtividade aquícola.

Os estudos genéticos na área da piscicultura são intensos e buscam o avanço do conhecimento nas áreas de seleção reprodutiva, genética molecular e biotecnologia de organismos, principalmente em peixes, ostras e camarões, programas de melhoramento genético estão sendo desenvolvidos para espécies de importância comercial, como salmão-do-Pacífico, bacalhau, linguado, tilápia-do-Nilo e camarão (ALVES, A. L.; VARELA E. S.; HASHIMOTO D. T., [2024]). Os recursos genéticos com marcadores moleculares possibilitam a avaliação da possibilidade da variabilidade genética que pode existir entre espécies específicas, marcadores genéticos possibilitam o mapeamento sendo assim contribuem evolutivamente para incorporação de programas de melhoramento (BERED *et al.*, 1997).

O Sequenciamento de Nova Geração podem ser destinadas para fins de investigações genéticas ou medicina de precisão (SAFADY, 2019). O NGS permite maior rendimento de sequências de ácidos nucleicos de forma rápida e barata em comparação ao sequenciamento de Sanger que atualmente é uma ferramenta com pouco rendimento e que leva mais tempo. Em relação à piscicultura, pesquisas de cunho genético são fundamentais para o desenvolvimento e aperfeiçoamento da atividade, bem como para programas de conservação (SAFADY, 2019).

No que se refere aos recursos genéticos para *S. brasiliensis*, a espécie nativa ainda não possui genoma completo publicado, tornando limitado o estabelecimento de programas de manejo sustentável para conservação e produtividade aquícola, mediante a isso os marcadores

microssatélites são ferramentas fundamentais para desenvolvimento de recurso para espécie, pois a partir dele poderá desenvolver outros estudos importantes para espécie, e atualmente esse marcador tem sido utilizado na construção de mapas genéticos, associação entre instabilidade do número de repetições e doenças genéticas, estrutura genética de populações e análises de paternidade. Nossa hipótese foi que sequenciamento de próxima geração, gerou muitos marcadores microssatélites, a partir disso foram desenhados *primers* para serem utilizados como ferramentas moleculares em pesquisas futuras.

2 OBJETIVOS

2.1 Objetivo geral

Caracterizar e desenvolver um painel de marcadores microssatélites para *Salminus brasiliensis*.

2.2 Objetivos específicos

- Realizar a montagem *de novo* do genoma de *Salminus brasiliensis*;
- de *S. brasiliensis*;
- Identificar os *primers* para amplificação dos microssatélites do genoma montado de *S. brasiliensis*;
- Avaliar, mediante ferramentas de PCR in silico os *primers* desenhados para a amplificação dos microssatélites.

3 REVISÃO DE LITERATURA

3.1 Biologia da espécie

Salminus brasiliensis (figura 1) pertencente à família *Characidae*, é descrita como peixe de escamas que apresenta coloração dourada com reflexos avermelhados, cada escama apresenta um filete negro no meio formando riscas longitudinais da cabeça à cauda, do dorso até abaixo da linha lateral, podem alcançar um metro e pesar em média 25 kg, e é uma espécie carnívora, possui ato migratório para sua reprodução, ou seja, reofílico, além de possuir características sensórias favoráveis na sua carne e seu crescimento acelerado favorece a indústria alimentícia (AMBIENTAL BRASIL, 2021).

O comportamento migratório é conforme seu estágio de vida (larva, juvenil ou adulto) ocupam diferentes ambientes na planície de inundação. As espécies migratórias geralmente desovam em águas abertas no canal principal e tributários, seus ovos e larvas são levadas para áreas alagadas e lagoas marginais (áreas de alagados, alagadiços, lagos, banhados, canais ou poços naturais que recebam águas dos rios ou lagoas em caráter permanente ou temporário). Quando estão na fase jovem, exploram ambientes de lagoas naturais e rios tributários represados, que atribuem como áreas de alimentação e crescimento. (SERAFINI, 2005).

Com as características reofílica dessa espécie os fatores mais recorrentes causadores do problema na pesca são o desmatamento da vegetação ciliar, excessivo esforço de pesca, captura de indivíduos jovens, drenagem de lagoas marginais, regulação do regime hidrológico dos rios, construção de barragens, poluição das águas e introdução de espécies exóticas (ZANIBONI; FILHO *et al.*, 2000).

Figura 1. Espécie *Salminus Brasiliensis* em seu habitat natural, no Mato grosso do Sul



Fonte: SEALE, 2022

O Dourado, *Salminus brasiliensis*, pertence ao gênero *Salminus*, ordem dos Characiformes e classe *Actinopterygii*. O gênero *Salminus* é descrito em cinco espécies: *S. affinis*, *S. hilarii*, *S. brasiliensis*, *S. iquitensis* e *S. franciscanus*. A espécie *S. brasiliensis* é a que possui o maior potencial para piscicultura (FLORA *et al.*, 2010).

Sendo assim, a taxonomia de *Salminus* pode ser apresentada da seguinte forma:

Super Classe: Pisces

Classe: Osteichthyes

Subclasse: Actinopterygii

Ordem: Characiformes

Subordem: Characoidea

Família: Characidae (Incertae Sedis)

Gênero: *Salminus* (Agassiz, 1829)

Espécie: *Salminus affinis* (Steindachner, 1988)

Salminus brasiliensis (Cuvier, 1816)

Salminus hilarii (Valenciennes, 1850)

Salminus iquitensis (Nakashima, 1941)

Salminus franciscanus (nova espécie, 2007) (FREITAS, 2010).

A espécie *S. brasiliensis* ocupa o topo da cadeia alimentar, sendo considerada uma espécie piscívora ou ictiófaga por excelência. Seu hábito alimentar apresenta dificuldades para a criação porque ele necessita de alimentos altamente proteicos e de alta qualidade, tornando-se caro (FLORA *et al.*, 2010).

Na piscicultura, a quantidade de ração fornecida deve variar de acordo com o tamanho do peixe e a temperatura da água, por isso na prática recomenda-se que o peixe esteja aparentemente saciado, evitando assim o desperdício de ração, pois os custos com alimentação representam 50-60% dos custos totais de produção (FLORA *et al.*, 2010).

É um peixe piracema que faz sua reprodução nos cursos superiores dos rios durante a estação das cheias, que coincide com os meses quentes, de outubro a março (FREITAS, 2010). Essa sazonalidade reprodutiva resulta da interação de fatores endógenos e abióticos, como temperatura da água, fotoperíodo e precipitação atmosférica, quando o nível da água aumenta, esses fatores atuam na maturação das gônadas, preparando-as para a reprodução. (FREITAS, 2010). A maioria dos peixes de piracema não apresentam indícios de desova em cativeiro, a desova em cativeiro pode ser obtida através de indutores hormonais (SERAFINI, 2005).

Em seu habitat natural, essa espécie prefere ambientes de corredeiras, onde os níveis de oxigênio costumam ser mais elevados. Para um bom crescimento desta espécie, recomenda-se que o teor de oxigênio não seja inferior a 4 mg/l. (FLORA *et al.*, 2010). Os parâmetros de

qualidade da água são extremamente importantes na aquicultura, porque as características da água influenciam na sobrevivência, crescimento e reprodução dos organismos aquáticos. O entendimento dos efeitos causados por variáveis de qualidade da água é essencial para adotar medidas de manejo, avaliar os impactos ambientais e manter a qualidade dos organismos cultivados (SERAFINI, 2005).

3.2 Importância econômica da espécie

Salminus brasiliensis é uma espécie de grande importância econômica e social na pesca artesanal, de subsistência e esportiva. Valorizada por sua agressividade e voracidade, é uma das espécies mais procuradas pelos pescadores (RIBOLLI, 2014).

A pesca artesanal, onde o *S. brasiliensis* é mais consumido, tem por definição a atividade que não é em larga escala, nem utiliza tecnologias sofisticadas e envolve pesados investimentos, além disso a pesca artesanal brasileira possui numerosas e complexas especificidades e que envolvem os fatores sociais, políticos, institucionais, econômicos e ambientais típico a cada local (SILVA, 2014). Em outro aspecto, a pesca artesanal é uma atividade caracterizada pela baixa produtividade e taxa de rendimento e de subsistência (SILVA, 2014).

Assim como a pesca artesanal, a espécie é importante também na pesca esportiva que é regulamentada pela Lei 11.959, de 29 de junho de 2009 conhecida como Lei da Pesca, “define a pesca amadora e esportiva como uma atividade de pesca de natureza não comercial voltada para o lazer, turismo e o desporto de seus praticantes, onde o pescado capturado e não é fonte de subsistência e nem de renda para o pescador” (MPA, [2024]).

No que se refere pesca esportiva o objetivo é apenas físgar o peixe, ou seja não tem o intuito do consumo e nem do comércio, é somente pelo prazer da pesca. Sendo assim os peixes são devolvidos vivos, antes disso os pescadores medem, pesam, fotografam para devolvê-los para natureza, a devolução tem o objetivo de deixar o peixe crescer mais, e desovar mais vezes, para o aumento da população, no quesito pesca esportiva é possível afirmar que na água doce tem maior chance de sobrevivência do peixe solto do que na água salgada (AMBIENTEBRASIL, 2021).

Sobretudo, *Salminus brasiliensis* tem despertado o interesse dos piscicultores pelo fato de ser um peixe muito conhecido na pesca esportiva e é uma espécie apreciada pela indústria alimentícia, pois sua carne é sensorialmente atrativa pelos consumidores, além de ter inúmeras características favoráveis para sua criação como crescimento acelerado favorecendo sua produção, domínio das técnicas reprodutivas, alto valor agregado, rusticidade, fácil adaptação e boa aceitação de alimentação artificial, ou seja, favorecendo sua procura no mercado

consumidor (NETO, 2020).

3.3 Biotecnologia aplicada a espécie

A espécie *Salminus brasiliensis* possui 8 loci de microssatélites polimórficos foram descritos por Rueda *et al.* (2011). Em estudo mais recente por Graciano *et al.* (2020) gerou um conjunto massivo de leituras curtas de DNA com pontuações de qualidade, chamada de bases associadas. O conjunto de dados, gerado pelo autor, embora de baixa cobertura genômica, é considerado de alta qualidade (média de um erro a cada 6.250 pb).

Outra biotecnologia aplicada foi de um estudo intitulado “Morfometria e alometria das larvas de cinco espécies de Characiformes na Bacia Hidrográfica do Rio Paraíba do Sul” que consistiu em analisar a morfologia e alometria de larvas pertencentes a cinco espécies potamódromas, dentre elas *Salminus brasiliensis* e as demais foram *Leporinus steindachneri*, *Eigenmann*, 1907, *Prochilodus lineatus* (Valenciennes, 1837), *Prochilodus vimboides* (Kner, 1859), *Brycon insignis* e *Steindachner*, 1877 as larvas usadas foram as que estavam nos estágios de pré-flexão, pós-flexão e desenvolvimento juvenil, foi possível classificar as larvas através de análises discriminantes validadas cruzadas com base em sete variáveis morfométricas com precisão 96% para *S. Brasiliensis* sendo o maior resultado e isso indica a característica larval que pode ser utilizada para a identificação taxonômica do ictioplâncton (SOUZA G.; CARAMASCHI E.P.; MONTEIRO L.R.; 2017).

A tecnologia da criopreservação do sêmen é uma técnica que exige solução crioprotetora adequada e taxas de congelamento e de descongelamento, são necessárias soluções crioprotetoras que são compostas por um diluidor e um crioprotetor intracelular, com a função de prevenir as crioinjúrias dos espermatozoides durante os procedimentos de congelamento e de descongelamento, além de manter os espermatozoides imóveis durante o armazenamento. Em um estudo com o Dourado (*Salminus brasiliensis*) foram realizados três experimentos, para avaliar sensibilidade dos espermatozoides com diferentes soluções crioprotetoras. No experimento 1, o sêmen foi diluído, 1:10, em 12 soluções. Metade de cada amostra foi resfriada por uma hora e a outra, criopreservada. A motilidade espermática foi avaliada imediatamente após a diluição e após o resfriamento em todas as amostras e, após o descongelamento, apenas nas amostras criopreservadas em DMSO. No experimento 2, o sêmen foi diluído, 1:5, em cinco soluções contendo DMSO e resfriado, criopreservado e avaliado como no experimento 1. No experimento 3, o sêmen foi diluído, 1:5, em quatro soluções contendo DMSO e criopreservado e avaliado quanto à motilidade e à fertilidade. Quando o sêmen foi diluído 1:10, observou-se motilidade acima de 58% em todas as amostras resfriadas em DMSO e em NaCl-tris-

metilglicol. Baixa motilidade foi observada nas amostras resfriadas nas outras combinações com metilglicol (5-32%) ou glicerol (0-8%) e naquelas criopreservadas (16-20%). Todas as amostras diluídas 1:5 apresentaram motilidade de 65-72% após o resfriamento e de 45-66% após o descongelamento (experimentos 2 e 3). As taxas de eclosão produzidas com sêmen criopreservado, entretanto, foram baixas (17-23%) em relação ao sêmen fresco (60%) (VIVEIROS *et. al.*, 2009).

Citogenética é um ramo da genética que estuda a estrutura e função da célula, especialmente os cromossomos. O estudo citogenético da espécie *Salminus brasiliensis* pelo autor Caramelo *et. al.* (2014) revelou a presença de um número diplóide de 50 cromossomos. A fórmula cariotípica consistiu em (10 M + 20 SM + 10 ST-A) e NF= 80. Nenhum heteromorfismo relacionado ao sexo foi detectado. As regiões organizadoras do nucléolo evidenciadas pela coloração com nitrato de prata estavam localizadas na região telomérica do braço longo do sexto par de cromossomos (SM). Através do bandamento C, foi observada a presença de blocos de heterocromatina constitutiva nas regiões centromérica e pericentromérica da maioria dos cromossomos do complemento. Os primeiros pares de cromossomos metacêntricos e submetacêntricos do gênero *Salminus* apresentam arranjo típico da heterocromatina constitutiva em nível centromérico e pericentromérico. Concluindo, o cariótipo de *Salminus brasiliensis* do Médio Paraná, Argentina, apresenta semelhança em relação a indivíduos da mesma espécie pertencentes às bacias do Alto Paraná.

Foi relatado o genoma mitocondrial completo do peixe *Salminus brasiliensis* que é uma molécula de DNA circular de 17.721 pb de comprimento, contendo 13 genes codificadores de proteínas, 2 genes de rRNA, 22 genes de tRNA e uma região de controle não codificante de 2.128 pb. Todos os genes codificadores de proteínas estão na fita pesada, exceto *Nd6*, e todos têm ATG como códon de início, exceto o gene *Cox1* que tem um códon de início GTG. Sete genes codificadores de proteínas têm códons de parada incompletos (*Nd2*, *Cox2*, *Atp6*, *Nd4* e *Cytb* têm T–, e *Cox3* e *Nd3* têm TA–). TAG é o códon de parada para *Nd6* e AGG é o códon de parada para *Cox1* (DIAS; CARMO; MARTINS *et. al.*, 2014).

Foi possível identificar um marcador molecular específico para o sexo em *Salminus brasiliensis* baseado no marcador SCAR. O marcador AFLP específicos do sexo masculino no estudo, foi isolado (Les1) e convertido em um marcador SCAR. O marcador Les1 foi confirmado como específico do sexo em todas as amostras testadas. Um marcador de DNA específico do sexo aplicável aos peixes seria muito útil para elucidar o mecanismo de determinação do sexo nos peixes (ROSA; CRUZ; SERRANO; *et al.*, 2017).

3.4 Sequenciamento de próxima geração (NGS)

Na década de 80 até a década de 2000 o Sequenciamento pelo método de Sanger era o método utilizado, a partir da segunda metade da década de 2000, a técnica de Sequenciamento de Nova Geração (Next Generation Sequencing, NGS) surgiu e aos poucos substituiu o método de Sanger (FARIA, 2021).

O método Sanger sequencia os fragmentos de DNA individualmente, enquanto o NGS é uma técnica de sequenciamento em larga escala, onde até bilhões de fragmentos são sequenciados simultaneamente em uma única corrida (FARIA, 2021).

O NGS promove sequenciamento de DNA em plataformas capazes de gerar informação sobre milhões de pares de bases em uma única corrida. Atualmente, os sequenciadores são da 1ª à 3ª geração (HORODESKY, 2014).

3.4.1 - 1ª GERAÇÃO (Degradação química e Interrupção da cadeia-Sanger)

Quando se trata de sequenciamento, essa trajetória começou em 1977, quando Frederick Sanger desenvolve a tecnologia de sequenciamento de DNA com o método de terminação de cadeia, hoje conhecido como sequenciamento Sanger, e Walter Gilbert criou outra tecnologia de sequenciamento, mas essa é baseada na modificação química do DNA e posteriormente a clivagem em bases específicas. Por motivos de sucesso e a baixa radioatividade, o sequenciamento Sanger foi denominado “primeira geração” com aplicações para sequenciamento laboratorial e comercial (LIN LIU *et. al.*, 2012).

3.4.2 - 2ª GERAÇÃO (454-Roche, Illumina Genome Analyser-HiSeq/MiSeq e SOLID)

3.4.2.1 - Roche 454

Roche 454 é um sequenciador usa tecnologia de pirosequenciamento, essa tecnologia depende da detecção de pirofosfato liberado durante a incorporação de nucleotídeos. Os DNAs da biblioteca com adaptadores específicos para 454 são desnaturados em fita simples e capturados por esferas de amplificação seguidas por PCR em emulsão. Em seguida, em uma placa de picotitulação, um dos dNTP complementar as bases da fita modelo com o auxílio de *ATP sulfurylase*, *luciferase*, luciferina, DNA *polimerase* e adenosina 5' fosfossulfato (APS) e isso vai liberar pirofosfato (PPi) igual à quantidade de nucleotídeo incorporado. O ATP que se transformou no PPi também vai transforma a luciferina em oxiluciferina o que gera luz visível. Ao mesmo tempo, as bases não compatíveis são degradadas pela *apirase*. Em seguida, outro

dNTP é adicionado ao sistema de reação e a reação de pirosequenciamento se repete (LIN LIU *et. al.*, 2012).

Em 2005 o comprimento de leitura do Roche 454 era de 100-150 pb, mais de 200.000 leituras e podia produzir 20 Mb por execução. Já em 2008 foi lançado o sistema 454 GS FLX Titanium; por meio da atualização, o comprimento de leitura conseguiu atingir 700 pb com precisão de 99,9% após filtrar e gerar 0,7 G de dados por execução em um dia. Em 2009 a Roche combinou o GS Junior com um sistema de bancada no sistema de sequenciamento 454, que facilitou a preparação da biblioteca e o processo de dados, e a produção foi atualizada para 14 G por execução. A maior vantagem da Roche é a sua velocidade pois leva apenas 10 horas para a conclusão do sequenciamento, seu comprimento de leitura também é uma característica diferenciada em comparação com outros sistemas NGS. A construção de sua biblioteca pode ser automatizada, e a PCR em emulsão pode ser semi-automatizada, o que pode reduzir em grande parte a mão de obra. No entanto o elevado custo dos reagentes continua a ser um desafio para a Roche 454, além disso uma das deficiências é que possui uma taxa de erro relativamente alta em termos de polibases maiores que 6 pb (LIN LIU *et. al.*, 2012).

Um exemplo que aplicou a técnica foi com tubarão-branco (*Carcharodon carcharias*) que é uma espécie interessante porque é representante da mais antiga linhagem de vertebrados com mandíbula existente, os tubarões-brancos também são motivo de preocupação para a conservação devido ao seu pequeno tamanho populacional e à ameaça da pesca excessiva. Por esse motivo pouco se sabe da biologia desse peixe e dos seus recursos que não estão disponíveis. A fim de resolver esse problema, os pesquisadores combinaram tecnologias de sequenciamento Roche-454 e Illumina para caracterizar o primeiro transcriptoma de qualquer tecido para esta espécie. Resumindo, com cDNA do coração de tubarão-branco, foi possível gerar 665.399 leituras Roche 454 (comprimento médio de 387 pb) que foram montadas em 141.626 contigs (comprimento médio de 503 pb) e com Illumina foi gerado 78.566.588 leituras Illumina, onde alinhou 454 contigs produzindo 105.014 sequências de consenso 454/Illumina (RICHARDS V.P.; SUZUKI H.; STANHOPE M.J.; SHIVJI M.S., 2013).

3.4.2.2 AB Solid (Sequenciamento por Oligo Ligation Detection)

SOLiD (figura 2) foi adquirido pela Applied Biosystems em 2006, esse sequenciador adota a tecnologia de sequenciamento de duas bases de acordo com sequenciamento de ligação. Na célula de fluxo SOLiD, as bibliotecas podem ser sequenciadas por ligação de sonda de 8 bases que contém local de ligação isso na primeira base, local de clivagem na quinta base e 4

corantes fluorescentes diferentes que são ligados na última base. O sinal fluorescente será acometido durante as sondas complementares à fita modelo e desaparecerá pela clivagem das últimas 3 bases das sondas. E a sequência do fragmento pode ser deduzida após 5 rodadas de sequenciamento usando conjuntos de primers ladder (LIN LIU *et. al.*, 2012).

O comprimento de leitura do SOLiD era de 35 pb e a saída foi de 3G de dados por execução. Como é um método de sequenciamento de duas bases, o SOLiD pode atingir uma alta exatidão de 99,85% após a filtragem. Em 2007, a ABI lançou o primeiro sistema SOLiD, posteriormente no final de 2010, o sistema de sequenciamento SOLiD 5500xl foi lançado, em seguida o SOLiD 5500xl teve cinco atualizações que foram lançadas pela ABI em apenas três anos. O SOLiD 5500xl obteve comprimento de leitura, precisão e saída de dados aprimorados de 85 pb, tendo a precisão de 99,99% e 30 G por execução, o processo de execução completo pode ser finalizado em 7 dias, o custo de sequenciamento é estabelecido por base a partir do uso de reagentes apenas pelos usuários do BGI. Porém o curto comprimento de leitura e o ressequenciamento apenas em aplicações ainda é sua principal deficiência. A aplicação do SOLiD inclui resequenciamento do genoma completo, resequenciamento direcionado, pesquisa do transcriptoma (que inclui perfil de expressão gênica, análise de RNA pequeno e análise do transcriptoma completo) e epigenoma (como ChIP-Seq e metilação) (LIN LIU *et. al.*, 2012).

Como outros sistemas NGS, a infraestrutura computacional do SOLiD é cara e não é habitual de usar; pois requer um data center com ar condicionado, cluster de computação, pessoal qualificado em computação, cluster de memória distribuída, redes rápidas e sistema de fila em lote, o seu sistema operacional utilizado pela maioria dos pesquisadores é GNU/LINUX, a execução do sequenciador sólido leva 7 dias e gera cerca de 4 TB de dados brutos, porém mais dados serão gerados após análise de bioinformática (LIN LIU *et. al.*, 2012).

3.4.2.2 - Sistema Illumina GA e HiSeq/MiSeq

Em 2006, a Solexa lançou o Genome Analyzer (GA), e em 2007 a empresa foi adquirida pela Illumina. O sequenciador Illumina adota a tecnologia de sequenciamento por síntese (SBS) sua biblioteca possui adaptadores fixos que é desnaturada em cadeias simples e enxertada na célula de fluxo, em seguida tem a amplificação em ponte para formar agrupamentos que contêm fragmentos de DNA clonal. Antecedendo o sequenciamento, a biblioteca é dividida em fitas simples com a ajuda da enzima de linearização e, posteriormente, quatro tipos de nucleotídeos ddATP, ddGTP, ddCTP, ddTTP que possuem diferentes corantes fluorescentes cliváveis e um grupo de bloqueio removível complementarizam o modelo um base de cada vez, e o sinal pode

ser capturado por um dispositivo de carga acoplada (LIN LIU *et. al.*, 2012).

No início, a saída do Solexa GA era de 1G por execução, mas através das melhorias na polimerase, buffer, célula de fluxo e software, em 2009 a produção de GA subiu para 20 G por execução em agosto, 30G por execução em outubro e 50 G por execução em dezembro, e a mais recente série GAIIx pode alcançar 85G por execução. Em 2010, a Illumina lançou o HiSeq 2000, que adota a mesma estratégia de sequenciamento do GA, e a BGI foi uma das primeiras a nível mundial a adotar o sistema HiSeq, sua produção foi inicialmente de 200 G por execução, melhorada para 600 G por execução atualmente, que poderia ser concluída em 8 dias. Comparado com 454 e SOLiD, o HiSeq 2000 é o mais barato em sequenciamento com US\$ 0,02/milhão de bases (reagente contado apenas pelo BGI) (LIN LIU *et. al.*, 2012).

O MiSeq engloba as funções de geração de cluster, SBS e análise de dados em um único instrumento e pode ir da amostra à resposta em um único dia (até 8 horas). O sequenciamento é baseado em terminador reversível Nextera, TruSeq e Illumina por química de síntese. Com esse sequenciamento os dados possuem maior integridade e uma gama mais ampla de aplicações, incluindo sequenciamento de amplicon, verificação de clones, ChIP-Seq e sequenciamento de genoma pequeno, são as partes marcantes do MiSeq. Também é flexível para realizar leituras únicas de 36 bp (saída de 120 MB) até 2×150 leituras emparelhadas. Devido à sua melhoria significativa no comprimento de leitura, os dados resultantes têm melhor desempenho na montagem contig comparado com o HiSeq (LIN LIU *et. al.*, 2012),

3.4.3 - 3ª GERAÇÃO (Ion Torrent e PacBio RS)

3.4.3.1 - Ion PGM de Ion Torrent

Ion PGM (figura 2) foi lançado pela Ion Torrent no ano de 2010, PGM usa tecnologia de sequenciamento de semicondutores, ou seja, quando um nucleotídeo é incorporado nas moléculas de DNA pela polimerase, um próton é liberado e ao detectar a mudança no pH, o PGM reconhece se o nucleotídeo foi adicionado ou não. Cada vez que o chip for inundado com um nucleotídeo após o outro, se não for o nucleotídeo correto, nenhuma voltagem será encontrada; se houver 2 nucleotídeos adicionados, será detectada voltagem dupla. O Ion PGM é a primeira máquina de sequenciamento comercial que não tem fluorescência e varredura de câmera, por esse motivo resulta em maior velocidade, menor custo e menor tamanho de instrumento, permitindo leituras de 200 pb em 2 horas e o tempo de preparação da amostra é inferior a 6 horas para 8 amostras (LIN LIU *et. al.*, 2012).

Uma aplicação exemplar do sequenciador Ion Torrent PGM é a identificação de patógenos microbianos, porque em 2011, houve um surto contínuo de *Escherichia coli* O104:H4 produtora de toxina Shiga virulenta na Alemanha. O sequenciamento completo do genoma no sequenciador Ion Torrent PGM ajudou os cientistas a identificar o tipo de *E. coli* que aplicaria diretamente a pista para encontrar a resistência aos antibióticos. A cepa parecia ser um híbrido de duas cepas *E. coli* entero agregativa e *E. coli* entero hemorrágica, a partir do resultado do sequenciamento de *E. coli* TY2482, o PGM mostra o potencial de ter um sequenciador rápido, mas de rendimento limitado, quando há um surto de uma nova doença (LIN LIU *et. al.*, 2012).

Um exemplo para essa plataforma foi o sequenciamento do genoma mitocondrial completo da carpa koi (*Cyprinus carpio* , Cyprinidae) pela primeira vez, a sequência do genoma mitocondrial tem 16.581 pb de tamanho e consiste em 13 genes codificadores de proteínas, 22 genes de tRNA, dois genes de rRNA e uma região de controle. Com o sequenciamento realizado foi construído a árvore filogenética usando um modelo de máxima verossimilhança mostrou uma relação irmã da carpa koi com outros peixes Cyprinidae (LI, X.; ZENG J.; ZHANG H.; WENG S.; HE J.; DONG C., 2016)

3.4.3.2 - PacBio

O sequenciamento do PacBio utiliza enzima modificada e observação direta da reação enzimática em tempo real, isso ocorre da seguinte forma a célula SMRT contém milhões de guias de onda de modo zero, incorporados com apenas um conjunto de enzimas e modelo de DNA que podem ser detectados durante todo o processo, ao decorrer da reação, a enzima incorporará o nucleotídeo na fita complementar e clivará o corante fluorescente previamente ligado ao nucleotídeo, em seguida, a câmera dentro da máquina captura o sinal em formato de filme em observação em tempo real, isto fornecerá o sinal fluorescente e também a diferença de sinal ao longo do tempo, o que pode ser útil para a previsão da variação estrutural na sequência, especialmente útil em estudos epigenéticos, como a metilação do DNA (LIN LIU *et. al.*, 2012).

O PacBio RS (figura 2) apresenta diversas vantagens: Primeiro, a preparação da amostra é muito rápida; leva de 4 a 6 horas em vez de dias, além disso, não necessita da etapa de PCR na etapa de preparação, o que reduz viés e erros causados pela PCR; Segundo, a taxa de rotatividade é bastante rápida, as corridas são concluídas em um dia; Terceiro, o comprimento médio de leitura é de 1300 pb que é maior do que qualquer tecnologia de sequenciamento de segunda geração. Embora o rendimento do PacBioRS seja inferior ao do sequenciador de

segunda geração, esta tecnologia é bastante útil para laboratórios clínicos, especialmente para pesquisas em microbiologia (LIN LIU *et. al.*, 2012).

A espécie que usou essa plataforma foi a dourada (*Oplegnathus fasciatus*) que é um peixe de recife rochoso economicamente importante do noroeste do Oceano Pacífico. Apesar da sua importância econômica, os recursos genômicos da dourada são em grande parte desconhecidas. Os autores utilizaram um genoma de referência em nível de cromossomo da fêmea de *O. fasciatus* usando a técnica de sequenciamento de molécula única PacBio (SMRT) e tecnologias de captura de conformação cromossômica de alta qualidade (Hi-C). O genoma foi montado em 120 contigs com comprimento total de 732,95 Mb e comprimento contig N50 de 27,33 Mb. *O. fasciatus* fornecerá um recurso importante não apenas para estudos ecológicos básicos e de genética populacional, mas também para dissecar mecanismos de seleção artificial na aquicultura marinha (BAI Y.; GONG J.; ZHOU Z.; LI B. *et. al.*, 2021)

3.4.4.3 - Nanopore

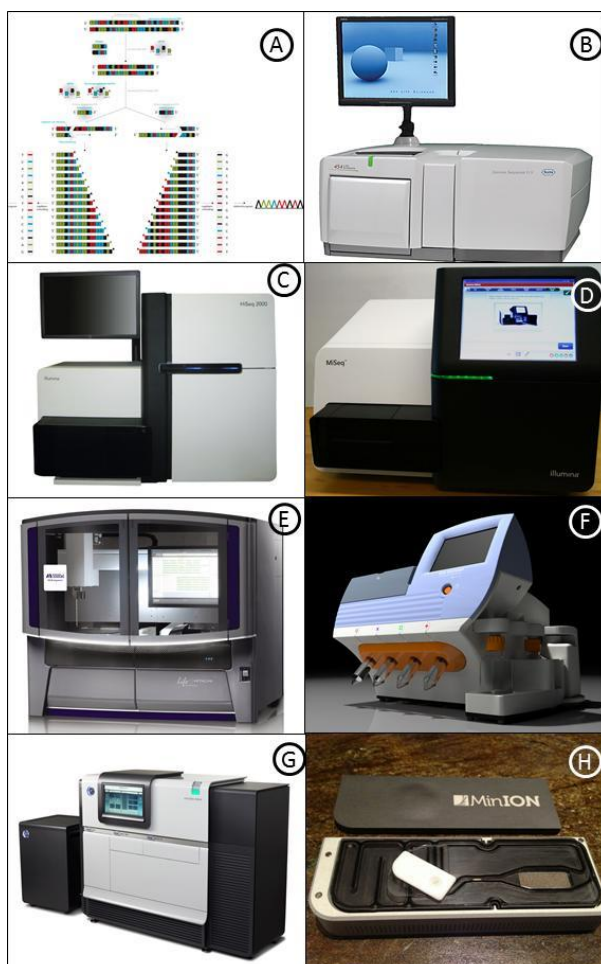
O Nanopore é um minúsculo bioporo com diâmetro em nanoescala, que pode ser encontrado em canais de proteínas embutidos na bicamada lipídica que facilita a troca iônica. No sequenciamento de nanoporos, um fluxo iônico é aplicado continuamente; A interrupção da corrente é simplesmente detectada pela técnica eletrofisiológica padrão; A leitura baseia-se na diferença de tamanho entre todos os desoxirribonucleósidos monofosfato (dNMP); Para determinado dNMP, a modulação de corrente característica é mostrada para discriminação e a corrente iônica é retomada após o nucleotídeo preso ser totalmente expelido (LIN LIU *et. al.*, 2012).

O sequenciamento Nanopore (figura 2) possui uma série de vantagens sobre as tecnologias de sequenciamento de próxima geração existentes, como atingir potencialmente um comprimento de leitura longo; a detecção de bases é livre de marcadores fluorescentes; o envolvimento da enzima é notavelmente evitado no sequenciamento de nanoporos, exceto o uso de exonuclease para manter o sDNA e a clivagem de nucleotídeos; em vez de sequenciar o DNA durante a polimerização, as cadeias simples de DNA são sequenciadas através de nanoporos por meio da despolimerização da cadeia de DNA, reduzindo significativamente o tempo prático (LIN LIU *et. al.*, 2012).

Autores em um estudo demonstraram o sequenciamento rápido e preciso de dois vírus

causadores de doenças que afetam a aquicultura global de salmonídeos, o alfavírus salmonídeo (SAV) e o vírus da anemia infecciosa do salmão (ISAV), usando sequenciamento de nanoporos na plataforma MinION (Oxford Nanopore Technologies), eles usaram a abordagem complementa a PCR de material infectado com sequenciamento MinION para recuperar informações genômicas, este método foi usado para apresentar o primeiro genoma do subtipo 6 do SAV, que se ramifica como irmão de todas as outras linhagens do SAV em uma reconstrução filogenética de todo o genoma (GALLAGHER M.D.; MATEJUSOVA I.; NGUYEN L. *et. al.*, 2018).

Figura 2. Equipamentos utilizados para sequenciamento de DNA. 1ª geração (A) Sanger; 2ª geração (B) 454-Roche, (C) Illumina HiSeq, (D) Illumina MiSeq, (E) Solid; 3ª geração (F) Ion Torrent, (G) PacBio, (H) Nanopore



Fonte: (HORODESKY, 2014)

3.5 Marcadores moleculares

Em síntese, marcador molecular é a diferença hereditária na sequência de DNA na posição correspondente do cromossomo homólogo de dois indivíduos diferentes, que seguem um padrão de herança mendeliana simples, além disso é um locus polimórfico que informa

sobre o genótipo do indivíduo e sua análise está orientada a todo genoma, independente das condições ambientais e do estágio fisiológico (ELEVAGRO, 2015).

Os marcadores moleculares fornecem um número de polimorfismos distribuídos aleatoriamente em torno de todo o genoma, permitindo a identificação precisa dos genótipos no início do desenvolvimento. Os avanços das técnicas moleculares possuem grande desenvolvimento nas áreas da bioinformática e da genética quantitativa. A aplicação dessas técnicas acelera e monitora os programas de melhoramento genético gerando grandes avanços no desenvolvimento de variedades melhoradas (LANZA, M. A.; GUIMARAES, C. T.; SCHUSTER, I., 2000).

Os principais tipos de marcadores moleculares, podem ser classificados em dois grupos, marcadores que usam hibridização e que usam a amplificação do DNA, além disso muitos marcadores baseados em PCR foram desenvolvidos nos últimos anos (GOMES, 2010).

O uso combinado de marcadores associado às tecnologias avançadas tem-se mostrado bastante eficientes para identificação de espécies, discriminação de subespécies, discriminação de subpopulações ou raças, inferência de distância genética entre populações, inferência de parentesco entre indivíduos em cativeiro ou na natureza e em estudos biogeográficos e evolutivos de padrões de migração (ROSA; PAIVA, 2009).

A maioria dos marcadores moleculares podem estar localizados em qualquer lugar sem estarem associados a um gene. Quando o marcador faz parte ou está próximo e ligado a um gene, determina uma característica fenotípica conhecida, esse marcador é então chamado de Marcador Funcional ou “Perfeito”. Então os marcadores moleculares mais utilizados em programas de melhoramento são: (CROPLIFE, 2020).

- SSR (*Sequências Simples Repetidas – Simple Sequence Repeats*): Os microssatélites, são marcadores com alto nível de polimorfismo e são formados por curtas sequências de nucleotídeos, normalmente de 1 a 6, sendo nomeadas de acordo com o número de nucleotídeos que as compõem, como mono, di, tri, tetra, penta e hexanucleotídeos, respectivamente repetidos em série em número variável como mostra a tabela 1.

Tabela 1. Classificação de marcadores microssatélites.

CLASSIFICAÇÃO DE MICROSSATÉLITES	MOTIF
Mononucleotídeos	...AAAAAAAAAAAA...
Dinucleotídeos	...ATATATATATAT...
Trinucleotídeos	...ATGATGATGATG...
Tetranucleotídeos	...ATGCATGCATGC...
Pentanucleotídeos	...AACCTAACCTAACCT...

- RFLPs (*Polimorfismo no Comprimento dos Fragmentos de Restrição – Restriction Fragment Length Polymorphism*). A técnica é que o DNA é “quebrado” por enzimas de restrição que cortam partes da sequência do genoma e gera um grande número de fragmentos. Esse foi o primeiro marcador de DNA a ser utilizado na construção de mapas genéticos humanos.
- RAPD (*Polimorfismo do DNA Amplificado ao Acaso – Random Amplified Polymorphism DNA*). São fragmentos de DNA amplificados ao acaso pela PCR utilizando iniciadores curtos de sequência aleatória, sendo assim a técnica é capaz de identificar o polimorfismo de maneira mais rápida e prática.
- SNPs (*Polimorfismo de nucleotídeo único – Single Nucleotide Polymorphisms*): Esse marcador também é altamente polimórfico e abundante no genoma, ele é baseado na identificação de alterações de uma única base de nucleotídeo em uma sequência de DNA, normalmente um gene.

De acordo com Faleiro (2007), os resultados da análise das sequências dos marcadores moleculares para o estudo filogenético podem gerar matrizes de distâncias e suas variações estatísticas, como também favorecer a realização de algoritmos de análise filogenética baseados no princípio de agrupamento ou em análises de parcimônia.

3.5.1 Marcador Microssatélite

Sequências repetitivas em tandem, nomeadas microssatélites, *Simple sequence repeats* (SSRs) ou *Short Tandem Repeats* (STRs), são repetições simples de mononucleotídeos a hexanucleotídeos. Algumas classes são preferíveis para fins de discriminação e genotipagem de indivíduos e populações. (BENNETT, 2000).

As regiões dos marcadores microssatélites são divididas em grupos com algumas características como, o tamanho do motivo de repetição, local no genoma e arranjo dos nucleotídeos nos motivos de repetição. O tamanho da repetição nas regiões microssatélites são

subdivididas em -mono, -di, -tri, -tetra, -penta e hexanucleotídicas. Enquanto a localização no genoma, estão subdivididos em microssatélites nucleares (nuSSR); microssatélites cloroplastidiais (cpSSR); e microssatélites mitocondriais (mtSSR). No arranjo dos nucleotídeos no motivo de repetição são classificados em perfeitos aqueles possuem apenas os nucleotídeos pertencentes ao motivo da sequência; ou imperfeitos que apresentam bases que não corresponde ao motivo; Os interrompidos possuem em sua constituição uma sequência curta que não correspondem ao seu motivo original; ou compostos que são constituídos de dois motivos de repetição arranjados em tandem como é possível verificar na figura 3 (THELLES, 2018).

Figura 3. Classificação dos microssatélites

CLASSIFICAÇÃO DE MICROSSATÉLITES	
Número de nucleotídeos por repetição:	Arranjo dos nucleotídeos nos motivos de repetição:
1. Mononucleotídeos AAAAAAAAAAAA	1. Perfeitos CTCTCTCTCTCT
2. Dinucleotídeos TATATATATATA	2. Imperfeitos CTCTCTATCTCT
3. Trinucleotídeos TAGTAGTAGTAG	3. Interrompidos CTCTAAACTCT
4. Tetranucleotídeos TAGCTAGCTAGC	4. Compostos CTCTCTAGAGAG

Fonte: Farias, 2021

Os microssatélites estão distribuídos por todos os genomas eucariotos, cujo grau de polimorfismo diminui à medida que o tamanho da unidade de repetição aumenta. Acredita-se que repetições de mononucleotídeos estejam presentes em cerca de meio milhão de loci, mas as repetições de pentanucleotídeos estão em apenas alguns milhares (BENNETT, 2000).

Segundo Wenne (2023) os marcadores microssatélites têm uma variedade de aplicações em pesquisas genéticas relacionadas a populações naturais e estoques cultivados nos últimos 30 anos. Os microssatélites são conhecidos como marcadores moleculares neutros hipervariáveis com o maior poder de resolução em comparação com quaisquer outros marcadores. O maior número de publicações dizia respeito à detecção da estrutura genética de populações intocadas em condições naturais, identificação de espécies invasoras, detecção de unidades de gestão evolutivas para genética de conservação, identificação de unidades de gestão pesqueira e análise de stocks mistos, efeitos de lotação ou fugas de peixes de criação peixes e

suas interações com populações naturais, análise de estoques de incubação, incluindo análise de parentesco e filiação, avaliação da ginogênese, seleção de características de escolha e herdabilidade na natureza e identificação de loci de características quantitativas (QTL). Comparado a outros marcadores, os microssatélites podem ser considerados marcadores superiores em tópicos como análise de parentesco e filiação em populações cultivadas e naturais, avaliação de ginogênese, androgênese e ploidização. Em outra situação onde microssatélites acoplados a SNPs são em potencial utilizados para mapeamento de QTL o que permanecerão viáveis. Os microssatélites continuarão a ser utilizados em estudos de diversidade genética em estoques cultivados, e também na pesquisa de populações naturais como técnica de pesquisa economicamente vantajosa.

3.6 Estratégias de melhoramento genético para *Salminus brasiliensis*

Ainda há pouco material publicado referente a *S. brasiliensis*, mas o pouco que tem faz-se necessário para conhecer mais sobre a espécie, então sabe-se que a hibridização e a criopreservação são umas dessas técnicas aplicadas para o Dourado.

Para o melhoramento genético, a hibridação na piscicultura tem sido atualmente uma técnica já bastante utilizada entre piscicultores em todo Brasil, onde possui seus primeiros registros em 1828 na Inglaterra. Segundo Fernandes (2007), o objetivo para o desenvolvimento de híbridos na piscicultura é descrito da seguinte forma: “diminuir o tempo de engorda, obter populações monossexo sem hormônios, obter indivíduos dóceis e aptos ao manuseio comum, resistência à patógenos e as condições ambientais como salinidade, temperaturas altas ou baixas, baixos teores de oxigênio dissolvido, etc.”

Já a criopreservação de sêmen de peixes é um método que promove o melhoramento genético devido ao fácil aproveitamento do sêmen e contribui para o desenvolvimento de novas linhagens. Além disso, um dos 19 fatores mais importantes dessa tecnologia é a possibilidade de armazenar germoplasma, o que garante a conservação da variação genética em populações ameaçadas. (WEINGARTNER, 2010).

4 MATERIAL E MÉTODO

4.1 Aspectos éticos e coleta de amostras

As amostras de nadadeira caudal de *Salminus brasiliensis* oriundas da Lagoa dos Patos no Rio Grande do Sul, Brasil, foram coletadas conforme protocolo CEUA-UFRA nº 4663010323. Para extração de DNA, utilizou-se o kit comercial Wizard® Promega (Promega Bio Sciences, LLC. San Luis Obispo, CA, USA), seguindo o protocolo do fabricante. Em espectrofotômetro NanoDrop™ ND-1000 (Thermo Scientific), a concentração de DNA das amostras foi calculada pelo índice de absorvância das bases a 260 nm e padronizadas para concentração de 100 ng/μl.

4.2 Pré-processamento de dados e montagem *de novo* do genoma

As amostras foram submetidas ao sequenciamento pela plataforma Illumina, utilizando o sistema MiSeq™. O Pré-processamento foi realizado pelo software FastQc (ANDREWS, 2010), onde foi analisado a qualidade das reads por meio dos parâmetros de qualidade média total por base, qualidade média por *read* e conteúdo GC das sequências. A filtragem foi realizada para controle de qualidade de bases e remoção de adaptadores, baseada no score de Phred mínimo de 20. Posteriormente, as reads foram submetidas à montagem *de novo*. Para isso, foram selecionados valores de *k* dos dados obtidos pelo programa KmerGenie. A escolha do montador a ser utilizado, foi testada por meio da montagem *de novo* do genoma pelos softwares SOAPdenovo2 (LUO *et al.*, 2012) e SPAdes (BANKEVICH *et al.*, 2012). A comparação de qualidade da montagem foi mediada pelo software QUAST (GUREVICH *et al.*, 2013), considerando a distribuição de *contigs* e os valores estatísticos de N50 e L50. Todo esse processamento foi realizado pelo Laboratório de Bioinformática e Genômica de Microrganismo da Universidade Federal do Pará (LBGM/UFPA).

4.3 Identificação de microssatélites

Para a busca dos marcadores microssatélites da espécie, utilizou-se os programas EasySSR com padrão do programa (ALVES *et al.*, 2023) e o MISA (BEIER *et al.*, 2017) com parâmetros de seleção de 1 a 6 motifs de repetição, mono a hexanucleotídeos, com no máximo 10 bases interrompendo 2 SSRs em um microssatélite composto. O número mínimo de repetições para cada motivo de repetição foi de (1/10) (2/5) (3/4) (4/3) (5/2) (6/2).

4.4 Desenho de primers e PCR *in silico*

Para o desenvolvimento de *primers* utilizou-se o software BatchPrimer3 (You *et al.*,

2008). Os parâmetros utilizados para desenho dos *primers* foram: tamanho do produto de 100 a 300 bases, tamanho do *primer* de 15 a 30 bases (tamanho ótimo de 21 bases), T_m de 50 a 70°C (ótimo a 55°C e diferença máxima de 3°C), ausência de *hairpins* e dímeros, e um conteúdo GC entre 40 e 50%. A presença de dímeros, *hairpins*, e foi verificado a temperatura de otimização pelo OligoAnalyser como uma alternativa para tornar os *primers* ainda mais robustos (Integrated DNA Technologies).

5 RESULTADOS

5.1 Sequenciamento

O total de *reads* sequenciadas visualizadas no programa *FastQc*, foi de 4.119.940 com comprimento de 36-100 pb. O parâmetro de qualidade $Q \geq 20$ por *read* obtida, foi de $Q = 39,06$. Utilizou-se dois montadores, o SPAdes que atingiu o $N_{50} = 1649$ e o SOAPdenovo2 com o $N_{50} = 731$, consideramos assim que a melhor montagem foi do SPAdes. O conteúdo GC foi de 39,64%, um valor relativamente próximo da distribuição normal pelo modelo do programa *FastQc*, que é de 42,79%.

5.2 Parâmetros de SSRs

Através do MISA foram identificadas 56.893 regiões repetitivas. A Tabela 1 está apresentando os parâmetros das regiões microssatélites analisadas para a espécie *Salminus brasiliensis* no programa MISA.

Tabela 2. Parâmetros gerais de microssatélites encontrados no genoma de *Salminus brasiliensis* (MISA).

PARÂMETROS DE SSRs	Nº
Total de sequências examinadas	80.576
Comprimento total do SSR (pb)	300.908
Total de SSR identificados	56.893
Sequências contendo SSR	34.182
Sequências contendo mais de 1 SSR	12.065
SSRs composto	1.591
Abundância relativa (SSR/Mb)	1.461.55
Densidade relativa (pb/Mb)	7.730.17

A saída do EasySSR mostra parâmetros e resultados diferentes do MISA, como o total de SSR que foi de 29.098, enquanto o MISA foi de 56.893 (Tabela 2).

Tabela 3. Parâmetros gerais de microssatélites encontrados no genoma de *Salminus brasiliensis* (EasySSR).

PARAMETROS DE SSRs	N°
N° total de SSRs	29.098
SSR por pb (%)	0,07
N° de SSRs em regiões codificantes	0
N° de SSRs em regiões não codificantes	29.098
N° Total de SSRs perfeitos	840
N° Total de SSRs imperfeitos	28.258

5.3 Caracterização de microssatélites

Nossos resultados apresentam divergências na identificação dos motivos de repetição nos diferentes softwares utilizados. Para o MISA, a maior ocorrência foi para pentanucleotídeos, com 63,7% e no EasySSR, a ocorrência maior foi para trinucleotídeos, com 46,3% (Tabelas 3).

Tabela 4. Identificação dos microssatélites em porcentagem no EasySSR e MISA.

SSR total	Mono	Di	Tri	Tetra	Penta	Hexa
EasySSR 29.098	47 (0,2%)	1.018 (3,5%)	13.476 (46,3%)	12.121 (41,6%)	1.537 (5,3%)	899 (3,1%)
MISA 56.893	102 (0,2%)	205 (0,4%)	561 (1,0%)	599 (1,1%)	36.239 (63,7%)	19.187 (33,7%)

Os *motifs* mais frequentes apresentados pelo MISA, foram os pentanucleotídeos e os hexanucleotídeos, que tiveram a maior ocorrência, 36.239 e 19.187, respectivamente. Os *motifs* mais frequentes mostrados pelo programa EasySSR, foram os trinucleotídeos e tetranucleotídeos, 13.476 e 12.121, respectivamente.

Os marcadores microssatélites foram obtidos através dos softwares MISA e EasySSR, porém o maior número de SSRs foi encontrado pelo MISA, com 56.893, sendo os pentanucleotídeos os mais frequentes; e o dados do EasySSR apresentaram 29.098 SSRs, predominantemente trinucleotídeos, divergindo do MISA e próximo ao easySSR.

5.4 Design de primers flanqueadores de SSR e PCR *in silico*

Para selecionar os marcadores potencialmente amplificáveis e polimórficos, selecionamos e desenhamos *primers* para regiões de microssatélites caracterizados pelo programa MISA do tipo tetranucleotídeos e trinucleotídeos. Do total de 501 sequências, foram detectados 93 SSRs, onde o total de *primers* escolhidos foram 89 e conseqüentemente *primers* bem-sucedidos, foram 80. Esses números foram obtidos através do programa bachprimers3 que

define o total de *primers* escolhidos e *primers* bem-sucedidos ([Dados suplementares](#)). Além disso, foi verificado a temperatura de otimização e com isso foi possível selecionar os *primers* ainda mais robustos, resultando em 64 *primers* esse valor foi obtido manualmente pelo programa que avalia a presença de *harpins* e *dimeros* de cada *primers* (Tabela 4).

Tabela 5. Lista de 64 *primers* de SSR desenvolvidos para *Salminus brasiliensis*.

EST-SSR	Motif	Primer (5'-3')	Tm (°C)	Length	Product Size
SbR1	(AAAC)3	F: GCTACTCCAGAAGACACAAAA R: TTCCTTTCTCTGAATCTACCC	54.76 55.12	21 21	154
SbR2	(TTTC)3	F: AACGTTAAAAGAGCCTTTCC R: CACTTTTATAAGCACCTGAGC	55.37 54.52	20 21	148
SbR3	(TCA)4	F: AGCTCAACTGCTACTTTACGA R: CGGGTAAATCAGATCAAGAG	54.73 54.89	21 20	149
SbR4	(ACTA)3	F: GCAAGAAGTCAAAACCAAGA R: CTCATACCTAATCCCAAACCT	55.58 54.89	20 21	170
SbR5	(ACC)4	F: CTAGGGAAGTGGGATTACCT R: CATTACTTGGGTAACCTCCTGA	54.86 54.43	20 21	161
SbR6	(ATT)4	F: CGTCTAGCTCGTTTTAATGTG R: CTGCACTAATCGCTTGAATAC	55.43 55.34	21 21	166
SbR7	(CTG)5	F: GTTGCGCAAGTAACTGAACTA R: GCATTTGGACCTGTAAAGAA	55.90 54.38	21 20	140
SbR8	(AACT)3	F: AGAATAATAGGCCATGACTGG R: ATTGCTTCATAACACCTTGC	55.86 55.40	21 20	155
SbR9	(GTCT)3	F: TCCAAGTACATTACCTGCAAC R: ACGCAGACTGACACATTTAAC	55.34 55.41	21 21	180
SbR10	(AATT)3	F: CTGCATCCAATACAACAGTG R: GTACATCTGTAATTCGGGTCA	55.06 55.10	20 21	153
SbR11	(TTAA)3	F: GATGGGCATCACAGATATTA R: TCTCTCGCGGTAAATAACTT	53.40 54.42	20 20	182
SbR12	(TAAA)3	F: CTGGTAAAATCAGCAGTCCTA R: AACTTTGCAGTCAGACAGGT	54.75 55.90	21 20	151
SbR13	(AAG)4	F: GCTTGCAATTATCAGCTATG R: AAAAACGCCAGTTGACTTAG	53.82 54.81	20 20	150
SbR14	(TGT)4	F: GTAGTGTTCTGATCTGGTGA R: ATAGGCCCCATAAATAAAGG	55.11 54.95	21 20	152
SbR15	(GAAA)3	F: GAAGATGAATATCAGCAGCAC R: AATCTTTCCTGCCAGACTTT	55.04 55.61	21 20	152
SbR16	(CAG)4	F: CAGACTCTGACAACCTTCAAA R: GCACAGGCAAGACTTTTACT	55.52 54.77	21 20	176
SbR17	(GCTT)3	F: TTACAGAAAGACATGCCAGA R: CTTAGGCTTCACTTCTTCTCC	54.37 55.11	20 21	149
SbR18	(TGT)4	F: AAGGACTTAGCGCTAGATTGA R: AGTTGTTGCACATTCAGTTG	56.11 54.64	21 20	149
SbR19	(AGC)5	F: CACAGGCTCTCTTTTGTACTA R: TGCATCTGAAGTTGCTGTAG	54.94 55.10	21 20	157
SbR20	(AGC)4	F: GAAATGAAGAGGAAGCTGAAC R: GTTCGGTCTTTTCATACAGC	55.74 54.95	21 20	157
SbR21	(TCAA)3	F: ATGAGCAGTTATGAGCAACA	54.39	20	143

		R: AGCGCATCATATCTTGTACTG	55.62	21	
SbR22	(AATT)3	F: GAGACAAAAGAGCACTTCTCA R: AACACGCTGGTTTCTGTTAT	54.87 54.90	21 20	175
SbR23	(TTGA)3	F: CTGAATCACTCTGCTGATTTT R: ATACAGCTGAGAACGGTCTTT	55.08 55.73	21 21	150
SbR24	(ACAA)3	F: TATTTAACAGCAGCTTGTGG R: GAGCGTTAAGCTACAAGTCAA	54.23 55.14	20 21	154
SbR25	(GAAA)3	F: GTGGCGATTAAGAGAGATAGA R: AGTGCATGTGTTTCAAACCTG	53.90 54.64	21 20	164
SbR26	(TGGT)3	F: TAACATTTCTGTGGTCATGG R: GTATGTTCTCTCGAATGATGC	55.24 54.87	19 21	155
SbR27	(TCCA)3	F: GATAACAACACAGGCACTTTC R: GTTCATTGGCTTGGAATAG	54.88 54.90	21 20	170
SbR28	(GCA)4	F:ACTATTACCGCTACTGAAAACC R: GCATCATTCAAGCAAGTTT	54.55 55.36	22 20	149
SbR29	(AAT)4	F: GTTGCTCATCACCTTTCATT R: CTGACTCGTCCAGATGAAATA	55.19 55.40	20 21	175
SbR30	(AAGA)3	F: GCTTTCTGTGGGACAGATT R: CAGGCTCATTGAGTTGTTCTA	55.16 55.64	19 20	149
SbR31	(TAT)4	F: CCACTGTTGTTTGTCTAAACG R: ATACTGATCACCCATTGCTC	55.92 54.97	21 20	155
SbR32	(AAAG)3	F: TCAGGAAACATGAGCTTAGTC R: TTACATCTCACTGACCCAGTT	54.67 54.69	21 21	159
SbR33	(TCA)4	F: GTCGCTATGTTGTAACCATGT R: GCATGACCTTTCGATCTTTAC	55.21 56.45	21 21	140
SbR34	(AATT)3	F:GAGAATCTCCATAACTGAGAC AR:CAACATGGTACCGTAGTACA A	53.56 55.33	22 22	166
SbR35	(TCA)4	F: TGTCCTCAGTTACTCATTACGA R: AACTATTGCCGCTAAAGAAG	55.08 54.22	22 20	151
SbR36	(GAA)4	F:ACGGAGACTTCAAAGAAGAAG R: ATACGCATCAATCCAATGTC	55.42 55.88	21 20	147
SbR37	(ACCA)3	F: TTTGTGGCTCAATACTGTCA R: CGATATACGCCCTTCTCTAT	55.20 55.32	20 21	143
SbR38	(CATT)3	F: CCTTGTAGTTTTGGGAGAGAC R: ATAAGCCGTGCAATACTGAG	55.55 55.61	21 20	149
SbR39	(GAAA)3	F: GTACCGTGAAAGTATTGGATG R: GGCAATAAGGTCGTTTACAG	54.70 55.01	21 20	182
SbR40	(AAT)4	F: CTGCAGCTATCAAAGAGCTA R: GATGCCTAACTTCTGAGGTCT	55.39 55.21	21 21	168
SbR41	(AAG)4	F: GCTCTTGCAGAAACACTAGAA R: GAACCAAACAATTTGTCACC	55.09 55.37	21 20	140
SbR42	(GCA)4	F: AATACATCCTGCAAGTTTCC R: GAAGTTGCCAATCTGGATAA	54.30 55.30	20 20	149
SbR43	(TGTT)3	F: TTACTIONCGTAACTTCTGCTG R: GGTAATGTTCAAGACCATGC	54.99 55.49	21 20	148
SbR44	(CTG)4	F:GGAAGTAACGGGATAAATACG R: AGGTACTGTAGACGAGGCATA	55.47 54.19	21 21	153
SbR45	(AAAT)3	F: GGTAAGCTCGATTTAGGAGTT R: TTTCATGCTGCGAAGATAC	54.54 54.92	21 19	232
SbR46	(TGCT)3	F: GCTCAGGTATGACCTTATTGA	54.49	21	134

		R: GGTGTTGTTTAAGCTACATGG	54.95	21	
SbR47	(ATT)4	F: CAGCATTAAATGACCTTAGTCG R: GGAAAGGTGAAGTTACCGTAT	55.20 54.69	21 21	189
SbR48	(AGC)4	F: CTGGATGCATTAATTCAGG R: GTGCTGAGAAATGAGAACATC	55.76 54.90	20 21	156
SbR49	(GCA)4	F: TCTTCTACAGTTGCACCATCT R: GAAGTAAACTGCCAAACTGAC	55.04 54.19	21 21	149
SbR50	(GCT)4	F: GTATCCAACAACTCAAGCTG R: GATGCTGCTCTTACTTCAGAT	55.05 53.92	21 21	163
SbR51	(TCA)4	F: CAATAAACCTAGCCTGGACTT R: GGTAATTTCTCTGGATCTGGT	55.27 54.82	21 21	133
SbR52	(AAAT)3	F: GCAAAGGCAAGCATGTAT R: AGACGGATTTAGGATTGACC	54.75 55.65	18 20	287
SbR53	(GCA)4	F: ATGTCTGCATATGACCATCC R: CTTCTTCGTTAAGACCTTCC	55.71 54.97	20 21	170
SbR54	(GTTT)3	F: GCTACTGAACTGCTGACATT R: GGCATACGATTTTGCTGA	54.78 55.66	21 18	173
SbR55	(CAAC)3	F: ATTTCATGTCCTACCCCTAGC R: TGGTCGTGGTCTACAAGTATC	56.03 55.17	21 21	144
SbR56	(AACAG A)5	F: GGCGACCCGATATAAGTAGTA R: CTGCGATCTAATACCTTTCCT	56.03 55.35	21 21	146
SbR57	(GTTC)3	F: GTTGCACGTTTTGTACCTTT R: GTAGATTACCTGCCCCATAAC	55.40 55.31	20 21	154
SbR48	(ATT)4	F: TGGTCAACCTATTTTGGAAC R: CTTACCTGCATGAAGACTTA	55.05 55.64	20 21	129
SbR59	(ATA)4	F: GTCAATGAAGGTAAGATGCTG R: CCATTTACAACATCGTAGAGC	54.97 55.02	21 21	158
SbR60	(TTAA)4	F: GAACAAGGCTGCTCATAATC R: CAAGCTACTAGGGTTAAGCTG	55.03 53.93	20 21	225
SbR61	(TAAG)3	F: ACCATTCAGACCTGCTAAAGT R: GTTAAATGCTCAAGGTTTGG	55.57 54.98	21 20	148
SbR62	(CTTT)3	F: TTGTATCCACCAGTACTTTGC R: TGGTAAAGCACTTACCTGAAG	55.34 54.84	21 21	125
SbR63	(CAT)4	F: GCATCTACTGCTTTAGACACC R: CCTATTCGTGTCGCAGATATT	54.31 56.91	21 21	149
SbR64	(CAT)4	F: AAATGGTCGATCACTTAACG R: GCAACGTATTCCAAATAGCTC	55.25 56.61	20 21	223

6. DISCUSSÃO

As aplicações de informações genômicas têm sido promissoras para a conservação de peixes (BERNOS; JEFFRIES; MANDRAK, 2020). Recursos genéticos e genômicos têm sido aplicados para a resolução de desafios taxonômicos e de conservação de espécies de peixes de importância socioambiental, além de serem utilizados como ferramentas para uso no melhoramento genético (GRACIANO *et al.*, 2022; RODRIGUES *et al.*, 2015).

O resultado do sequenciamento utilizando o montador SPAdes, apresentou valor médio

de qualidade por *read* (Q = 39,06, conteúdo GC 39,64% e N50 = 1649, semelhante ao estudo de Graciano *et al.* (2022), que sequenciou a mesma espécie (Q = 38, conteúdo CG de 41,05% e N50= 10.889), demonstrando assim que nossos dados estão dentro de um limiar de qualidade diante de outros estudos que vêm sendo realizado. Porém divergências de valores podem estar associados não somente a plataforma de sequenciamento, mas também ao montador que cada estudo utilizado.

Os programas SPAdes e SOAPdenovo2 já foram utilizados, conjuntamente, para evitar erros no processo de splicing de D-Loop, estratégia que já vem sendo utilizada em montagem de genoma mitocondrial de peixes (SUN *et al.*, 2022). Porém na presente metodologia os programas foram utilizados com o objetivo de comparação e definição do montador mais adequado, onde o montador que apresentou melhores resultados foi o SPAdes.

Genomas de peixe variam em microssatélites com aproximadamente 342 Mb a 2.967 Mb, onde as repetições em tandem podem estar distribuídas ao longo de todo o genoma, contribuindo principalmente para o processo evolutivo da espécie (YUAN *et al.*, 2018; CHALOPIN *et al.*, 2015). Estima-se que o genoma parcial de *S. brasiliensis*, representa uma pequena parte do genoma completo. No entanto, ele apresenta diversas informações significativas sobre a espécie, com amplas aplicações, onde os marcadores detectados podem ser úteis para conservação de espécies que estejam ameaçadas de extinção, estudos populacionais, uso no melhoramento genético, entre outros (TAVARES *et al.*, 2014).

A enguia do pântano (*Monopterus albus*) é uma espécie nativa do Brasil cada vez mais emergente que vem sendo utilizada como modelo, apresentando 99.293 microssatélites que foram identificados baseado no genoma completo, que utilizou o programa MISA (LI *et al.*, 2017). Em nosso estudo foi possível identificar 56.893 SSRs para o genoma de *S. brasiliensis* utilizando o mesmo programa, o MISA.

A distribuição de microssatélites varia em diferentes regiões do genoma para diferentes espécies (SRIVASTAVA *et al.*, 2019). Em regiões não codificantes é mais abundante, quando comparada a regiões codificantes (LI *et al.*, 2017). Os microssatélites com número de repetições entre 15 e 20 podem ser fortemente polimórficos, essas características são indicadas para estudos de genética de populações (FAZZI-GOMES *et al.*, 2021). Entre os marcadores microssatélites desenvolvidos para *S. brasiliensis*, as classes de mononucleotídeos, dinucleotídeos e trinucleotídeos apresentaram motivos com repetições dentro desta faixa.

De acordo com Silva (2013), as leituras de single-end da Illumina quando passam pela montagem *de novo*, com elevados parâmetros de rigor, tornaram-se ainda mais adequados para identificação de regiões de microssatélites, assim, os autores afirmam que, o processo de montagem adicional adquire um potencial de qualidade de sequências que aumenta a facilidade

de encontrar regiões flaqueadoras de qualidade para projeção de pares de *primers*.

O programa EasySSR utiliza a linha de comando do IMEx com configurações de entrada e saída de dados novos. Isso é vantajoso, pois o IMEx usa um algoritmo simples, podendo gerar dados como *strings* com a abordagem de janela deslizante para a busca dos microssatélites, relata os motivos, número de cópias, localização genômica, genes próximos, e muitos outros recursos pertinentes. Diante disso, o IMEx torna-se mais sensível, eficaz e potente, quando comparado ao MISA que foi o primeiro programa utilizado no estudo e que é amplamente utilizado (MUDUNURI e NAGARAJARAM, 2007).

No presente estudo, os marcadores mais frequentes identificados no genoma de *S. brasiliensis*, foram dinucleotídeos e trinucleotídeos. Os trinucleotídeos quando comparados a dinucleotídeos, são marcadores mais eficientes para genotipagem de indivíduos e grupos de uma mesma espécie, principalmente em função do maior comprimento, diminuindo a ocorrência de erro genômico (FALCÃO *et. al.*, 2004). Diante disso, desenhamos e selecionamos *primers* que flaqueassem apenas marcadores do tipo tri e tetranucleotídeos, pois diferentes estudos apontam como os mais utilizados em análises de polimorfismo devido à alta taxa de mutação, menores possibilidades de erros de genotipagem e devido ao seu tamanho (motivos longos e perfeitos) podem estar associados à exibição de maior variabilidade alélica (FALCÃO *et. al.*, 2004).

7. CONCLUSÃO

Mediante o sequenciamento de nova geração pela plataforma Illumina foram desenvolvidos uma quantidade considerável de marcadores microssatélites eficientes para futuros estudos de polimorfismos genéticos em *Salminus brasiliensis*, onde também *primers* foram desenhados e testados *in silico* para nortear futuras pesquisas.

REFERÊNCIAS

- ALVES, A. L. ; VARELA, E. S.; HASHIMOTO, D. Teruo. **Genética aplicada à piscicultura**. 2024. Disponível em: <chrome-extension://efaidnbnmnibpcajpcglefindmkaj/https://ainfo.cnptia.embrapa.br/digital/bitstream/doc/1083554/1/cap.8.pdf>. Acesso em: 08 fev. 2024.
- ÁRVORE filogenética: representação da evolução. Varsomics. 2021. Disponível em: <https://blog.varsomics.com/arvore-filogenetica>. Acesso em: 23 jul. 2023.
- ANDREWS, S. FastQC: a quality control tool for high throughput sequence data. 2010.
- BENNETT, L. Demystified... Microsatellites. **BMJ Journal** V. 53 p.177-183. Jul, 2022.
- BANKEVICH, A. SPAdes: A New Genome Assembly Algorithm and its Applications to Single-Cell Sequencing. **Jornal of Computational Biology**, v. 19, n. 5, p. 455–477, 2012.
- BAI, Y., *et al.* Chromosome-Level Assembly of the Southern Rock Bream (*Oplegnathus fasciatus*) Genome Using PacBio and Hi-C Technologies. **Front Genet.** 2021 Dec 21;12:811798. doi: 10.3389/fgene.2021.811798.
- BRANDÃO-DIAS, P. F. P., CARMO, A. O., MARTINS, A. P. V., PIMENTA, R. J. G., ALVES, C. B. M. e KALAPOTHAKIS, E. (2016) Genoma mitocondrial completo de *Salminus brasiliensis* (Characiformes, Characidae), **DNA mitocondrial Parte A**, 27 :3, 1577-1578, doi:10.3109/19401736.2014.958676
- BEIER, S., THIEL, T., MÜNCH, T., SCHOLZ, U., MASCHER, M. MISA-web: a web server for microsatellite prediction. **Bioinformatics**. v. 33, p. 2583 – 2585, 2017.
- BENNETT, P. Demystied... Microsatellites. **Molecular Pathology: Edt./S. I/**, 2000. ISBN 177-183.
- BERNOS,T.A., JEFFRIES K.M., MANDRAK N.E. Linking genomics and fish conservation decision making: a review. **Reviews in Fish Biology and Fisheries**. 2020 Dec; v.30 p.587-604.
- BERED, F., BARBOSA NETO, J.F; e CARVALHO, F.I.F. (1997). Marcadores moleculares e sua aplicação no melhoramento genético de plantas. **Ciência Rural**, p.513–520. <https://doi.org/10.1590/S0103-8478199700030002>
- CHALOPIN, D.,NAVILLE M, PLARD F., GALIANA D., VOLFF J.N. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. **Genome Biology Evolution**. v. 7, p. 567–80, 2015.
- DOURADO: *Salminus maxillosus*, 2021. Disponível em: https://ambientes.ambientebrasil.com.br/agua/pesca_esportiva_em_agua_doce/dourado_-_salminus_maxillosus.html. Acesso: 16. Jul. 2022.
- FARIA, A. Sequenciamento de Nova Geração (NGS) – a (re)evolução do sequenciamento de DNA. Entenda o Sequenciamento de Nova Geração (NGS), 2021. Disponível em: <https://blog.mendelics.com.br/ngs/>. Acesso em: 22 jul. 2023.

FALEIRO, F. Marcadores genético moleculares: aplicados a programas de conservação e uso de recursos genéticos. 1. ed. Planaltina: **Embrapa**, 2007. 102 p.

FALCÃO, C., *et al.* Desenvolvimento e mapeamento de microssatélites derivados de ESTs em Eucalyptus. 1. ed. Brasília: Maria das Graças Simoes Pires Negrão, 2004. 8 p. v. 150

FAZZI-GOMES, P., AGUIAR, J., CABRAL, G.F., MARQUES, D., PALHETA, H., MOREIRA, F., *et al.* (2021) Genomic approach for conservation and the sustainable management of endangered species of the Amazon. **Plos One**, 16, p.1-14, fevereiro, 2021. doi.org/10.1371/journal.pone.0240002

FERNANDES, J. Produção de Híbridos na Piscicultura: Tecnologias como a indução hormonal e reprodução artificial, tornam a produção de peixes híbridos uma prática relativamente simples. **Jornal dia de campo**, 2010. Disponível em: <http://www.diadecampo.com.br/zpublisher/materias/Newsletter.asp?id=21564&secao=Colu>. Acesso em: 24 jul. 2023.

FLORA, M., *et al.* Biologia e cultivo do dourado (*Salminus brasiliensis*). 2010. UFSM, Santa Maria, RS, 2010. **Acta veterinaria brasílica** p. 7-14 V. 4.

FREITAS, L. **Diversidade genética em Dourado (*Salminus brasiliensis*, Curvier, 1816) uma espécie de grande interesse comercial no Pantanal Mato-grossense**. Orientador: Dr. Pedro Manoel Galetti Junior. 2010. Tese (doutorado - Curso de Ecologia e Recursos Naturais do Centro de Ciências Biológicas e da Saúde) UFSCar, São Carlos, SP. 2021. Disponível em: <chrome-extension://efaidnbnmnnibpcajpegclclefindmkaj/https://repositorio.ufscar.br/bitstream/handle/ufscar/1690/3186.pdf?sequence=1&isAllowed=y>. Acesso em: 20 de Jan. 24

GALLAGHER, M.D., MATEJUSOVA, I., NGUYEN, L., RUANE, N.M., FALK, K., MACQUEEN, D.J., Nanopore sequencing for rapid diagnostics of salmonid RNA viruses. **Sci Rep**. 2018, Nov 5. doi: 10.1038/s41598-018-34464-x.

GUREVICH, A., SAVELIEV, V., VYAHHI, N., *et al.* QUAST: quality assessment tool for genome assemblies. **Bioinformatics**. v. 29 p. 1072–1075, 2013.

GRACIANO, R., *et al.* Recursos genômicos para *Salminus brasiliensis*. **Frontier** São Carlos: Tony Silveira, 2022. 6 p. v. 13. <https://doi.org/10.3389/fgene.2022.855718>

GOMES, P. **Diversidade genética de salminus brasiliensis Utilizados em programas De repovoamento**. Orientador: Dr. Ricardo Pereira Ribeiro. 2010. Tese (Doutorado - Curso de Zootecnia) Universidade Estadual De Maringá. Maringa, 2010.

HORODESKY, A. Sequenciamento de nova geração para análises metagenômicas: enfoque ao uso do sequenciador illumina. 2014. Disponível em: <https://gia.org.br/portal/sequenciamento-de-nova-geracao/>. Acesso em: 22 jul. 2023.

LANZA, M. A., GUIMARAES, C. T., SCHUSTER, I. Aplicação de marcadores moleculares no melhoramento genético. ed. Belo Horizonte: **Embrapa**, 2000. 97-108 p. v. 21. <http://www.alice.cnptia.embrapa.br/alice/handle/doc/484109>

LUO, R., LIU, B., XIE, Y., LI, Z. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. **Gigascience**, v.1, p. 18, 2012.

LI, Z; CHEN, F; HUANG, C. *et al.* Genome-wide mapping and characterization of microsatellites in the swamp eel genome. **Scientific Reports**, Wuhan University, Wuhan, P. R. China, p. 1-9, 09, jun. 2017. doi.org/10.1038/s41598-017-03330-7-7

LI, X., ZENG, J., ZHANG, H., WENG, S., HE, J., DONG, C., Ion Torrent next-generation sequencing reveals the complete mitochondrial genome of koi carp (*Cyprinus carpio*, Cyprinidae). **Mitochondrial DNA B Resour.** 2016 Feb 5;1(1):34-36. doi: 10.1080/23802359.2015.1137807.

LIN, L., YINHU, L., SILIANG, L., NI, H., YIMIN, H., RAY, P., DANNI L., LIHUA L., MAGGIE L. Comparison of Next-Generation Sequencing Systems. **BioMed Research International**, 2012, p. 11, 2012. <https://doi.org/10.1155/2012/251364>

MARCADORES moleculares: conceitos basicos. **Elevadro**, 2015. Disponível em: <https://elevagro.com/blog/marcadores-moleculares-conceitos-basicos/>. Acesso em: 04 fev. 2024.

MARCADORES moleculares, identificando a origem da diversidade entre os seres vivos. Croplife. 2020. Disponível em: <https://croplifebrasil.org/noticias/marcadores-moleculares-identificando-a-origem-da-diversidade-entre-os-seres-vivos/>. Acesso em: 23 jul. 2023.

MUDUNURI, S. B., NAGARAJARAM, H. A. IMEx: imperfect microsatellite extractor. **Bioinformatics**, v. 23, n. 10, p. 1181-1187, 2007.

NETO, A. Biometria de dourados (*salminus brasiliensis*) criados em tanque escavado. Anais do 18º Encontro Científico Cultural Interinstitucional, 2020. 5 p. ISSN 1980-7406.

PESCA esportiva. 2021. Disponível em: <https://www.ambientebrasil.com.br/sobre-nos/>. Acesso em: 23 jan. 2024.

RUEDA, E.C, AMAVET, P., BRANCOLINI, F., SOMMER, J., ORTÍ, G. 2011 Isolation and characterization of eight polymorphic microsatellite markers for the migratory characiform fish, *Salminus brasiliensis*. **J. Fish Biol**, p.370-1375. doi:10.1111/j.1095-8649.2011.03109.x

RICHARDS, V.P., SUZUKI, H., STANHOPE, M.J., SHIVJI, M.S. Characterization of the heart transcriptome of the white shark (*Carcharodon carcharias*). **BMC Genomics**. 2013 Oct 11;14:697. doi: 10.1186/1471-2164-14-697.

ROSA, A., PAIVA, S., Marcadores Moleculares e suas Aplicações em Estudos Populacionais de Espécies de Interesse Zootécnico. ed. Embrapa Cerrado, 2009. 35 p. v.

RIBOLLI, J. **Caracterização genética populacional do dourado, *Salminus brasiliensis* (Pisces, Characidae, Salmininae) na Bacia do alto e médio rio Uruguai. Orientador: Pedro Manoel Galetti Junior.** 2014. Tese (Doutorado - Curso de Ecologia e Recursos Naturais), Universidade Federal de São Carlos, 2014

SAFADY, N.. Medicina genômica: NGS: O que é Sequenciamento de Nova Geração. Varsomic. 2019. Disponível em: <https://blog.varsomics.com/ngs/>. Acesso: 22. Julho, 2022.

SOUZA, G., CARAMASCHI, EP., MONTEIRO, L.R. Morfometria e alometria das larvas de cinco espécies de Characiformes na Bacia Hidrográfica do Rio Paraíba do Sul. **Zigoto** . 2017, p.507-518. doi:10.1017/S096719941700034X

SERAFANI, R. L. **Efeito do oxigênio dissolvido e da amônia na sobrevivência e crescimento de juvenis de dourado, *Salminus brasiliensis***. 45p. Dissertação (Mestrado em Aquicultura). Universidade Federal de Santa Catarina. Florianópolis, 2005

SILVA, A., *et al.* Pesca artesanal brasileira. Aspectos conceituais, históricos, institucionais e prospectivos.. 1. ed. Palmas: **Embrapa**, 2014. p. 32

SILVA, P.I., MARTINS, A.M., GOUVEA, E.G. *et al.* Development and validation of microsatellite markers for *Brachiaria ruziziensis* obtained by partial genome assembly of Illumina single-end reads. **BMC Genomics** **14**, 17 (2013). <https://doi.org/10.1186/1471-2164-14-17>

SUN C.H., ZHANG Y.N., ZENG X.S., *et al.* Mitogenome of *Knodus borki* (Cypriniformes: Characidae): genomic characterization and phylogenetic analysis. **Molecular Biology Reports**. v. 49, p.1741-1748, 2022.

TAVARES, R.A., PIEDRAS S.R.N., NUNES, M.D., ALMEIDA, D.B., MOREIRA, C.G.A., FERNANDES, J.M., FREITAS, S.F., MOREIRA, H.L.M., POUHEY, J.L.O.F., DIONELLO, N.J.L. Identification of microsatellite loci with amplification potential in “pejerrey” (*Odontesthes humensis*). **Arq. Bras. Med. Vet. Zootec**, v.66, n. 6, p.1941-1945, 2014

WEINGARTNER, M. **Aperfeiçoamento das técnicas de fertilização de ovócitos de dourado, *salminus brasiliensis* cuvier, 1816 (characiformes: characidae), utilizando sêmen fresco e congelado durante o processo de reprodução induzida**. 2010. Tese (Doutorado - Curso Aqüicultura do Centro de Ciências Agrárias), Universidade Federal de Santa Catarina, Florianópolis

VIVEIROS, A. T. M., OLIVEIRA, A. V., MARIA, A. N., ORFÃO, L. H., & SOUZA, J. C. (2009). Sensibilidade dos espermatozoides de dourado (*Salminus brasiliensis*) a diferentes soluções crioprotetoras. **Arquivo Brasileiro De Medicina Veterinária E Zootecnia**, 61(4), 883–889. <https://doi.org/10.1590/S0102-0935200900040001>

WENNE R. Microsatélites como Marcadores Moleculares com Aplicações na Exploração e Conservação de Populações de Animais Aquáticos. **Genes**. 2023; 14(4):808. <https://doi.org/10.3390/genes14040808>

YUAN, Z.; LIU, S.; ZHOU, T.; *et al.* Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. **BMC Genomics**. v.19, 141, 2018.