



UNIVERSIDADE FEDERAL DO PARÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM
ECOLOGIA ÁQUÁTICA E PESCA



MÔNICA SILVA COELHO

AVALIAÇÃO DE MARCADORES MOLECULARES PARA IDENTIFICAÇÃO DE
PEIXES NEOTROPICAIS ATRAVÉS DA TÉCNICA DE METABARCODING

Belém – PA
2019

MÔNICA SILVA COELHO

**AVALIAÇÃO DE MARCADORES MOLECULARES PARA IDENTIFICAÇÃO DE
PEIXES NEOTROPICAIS ATRAVÉS DA TÉCNICA DE METABARCODING**

Tese apresentada ao Programa de Pós-Graduação em Ecologia Aquática e Pesca da Universidade Federal do Pará, como parte dos requisitos exigidos para obtenção do título de doutor em Ecologia Aquática e Pesca.

**Orientador: Dr. Leonardo dos Santos Sena
Co-orientador: Dr. Filipe José de O. Costa**

**Belém – PA
2019**

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD Sistema de Bibliotecas da Universidade Federal do Pará

Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

S586a Silva Coelho, Mônica
AVALIAÇÃO DE MARCADORES MOLECULARES PARA IDENTIFICAÇÃO DE PEIXES NEOTROPICAIS ATRAVÉS DA TÉCNICA DE METABARCODING / Mônica Silva Coelho. — 2019. 77 f.

Orientador(a): Prof. Dr. Leonardo dos Santos Sena
Coorientador(a): Prof. Dr. Filipe José de Oliveira Costa
Tese (Doutorado) - Programa de Pós-Graduação em Ecologia Aquática e Pesca, Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, 2019.

1. DNA ambiental. 2. Metabarcoding,. 3. eDNA. 4. Simulação de ambientes naturais. 5. Inventário da ictiofauna.
I. Título.

CDD 574.028

MÔNICA SILVA COELHO

AVALIAÇÃO DE MARCADORES MOLECULARES PARA IDENTIFICAÇÃO DE PEIXES NEOTROPICAIS ATRAVÉS DA TÉCNICA DE METABARCODING

BANCA EXMINADORA

DR. LEONARDO DOS SANTOS SENA
Universidade Federal do Pará – UFPA
Campus Belém – Belém – PA
Presidente

Co-orientador: DR. FILIPE JOSE DE OLIVEIRA COSTA
Universidade do Minho – UM
Campus Braga – Braga – Portugal

Banca: DR. MARCELO N. VALLINOTO DE SOUZA
Universidade Federal do Pará – UFPA
Campus Bragança – Belém – PA

Banca: DRA. RENATA COELHO. R. NORONHA
Universidade Federal do Pará – UFPA
Campus Belém – Belém – PA

Banca: DR. TIBERIO C. TORTOLA BURLAMAQUI
Instituto Evandro Chagas – CIT
Ananindeua – PA

Banca: DRA. JANAINA M. VASCONCELOS MASSAFRA
Universidade Federal do Pará – UFPA
Campus Belém – Belém – PA

Suplente: DR. MARCELO PETRACCO
Universidade Federal do Pará – UFPA
Campus Belém – Belém – PA

RESUMO

A Amazônia é considerada a última barreira de desenvolvimento socioeconômico no Brasil. Sua grande extensão e formações naturais dificultam a realização de inventários faunísticos devido ao alto custo envolvido em logística. Neste sentido, novas tecnologias não invasivas vêm sendo desenvolvidas para auxiliar no inventário das espécies, dentre elas, a técnica de obtenção de DNA a partir de amostras ambientais (eDNA) para identificação de vertebrados por metabarcoding, tanto em ambiente simulado (um sistema fechado de aquário) e ambiente natural. Apesar de ser um método promissor, a técnica de eDNA não foi tão testada em ecossistemas aquáticos complexos como a Amazônia. O objetivo deste estudo foi testar em comunidade simulada e natural a sensibilidade o método de eDNA para identificar espécies de peixes amazônicos utilizando uma abordagem de metabarcoding. Para isso, foram utilizados quatro *amplicons* modelados para os genes mitocondriais *cox1*, 16S e 12S para identificação do grupo Actinopterygii. Nossos resultados indicam que diferentes protocolos de sequenciamento podem influenciar nas taxas de recuperação de DNA das espécies alvos, reduzindo consideravelmente o número de sequências válidas. No entanto, o uso concomitante de quatro diferentes *amplicons* permitiu aumentar o número de espécies identificadas pela técnica de eDNA. O *primer* COA se mostrou a melhor opção dentre os demais utilizados, identificando 13 das 34 espécies do sistema de aquário, embora não tenha identificado com exatidão algumas espécies congêneres. O tamanho do *amplicon* influenciou diretamente na identificação das espécies, indicando que sequências com menos de 300 pb não são suficientes para amostrar a diversidade de peixes amazônicos. De qualquer forma, tanto o número de indivíduos como suas características fisiológicas e comportamentais parecem influenciar na quantidade de material genético presente em uma amostra de eDNA. Conclui-se que o protocolo utilizado pode recuperar a informação de algumas espécies de peixes em ecossistemas naturais, indicando que a técnica de eDNA pode ser utilizada se associada às demais técnicas tradicionais para melhorar os inventários e o monitoramento de espécies de peixes amazônicos.

Palavras-chave: DNA ambiental, eDNA, Metabarcoding, Inventário da ictiofauna, simulação de ambientes naturais, COX-1, 12S, 16S.

ABSTRACT

The Amazon region is considered the last frontier of socioeconomic development in Brazil. Its range and natural formation make it difficult to perform faunistic inventories because of the high cost involved. Thus, new non-invasive technologies have been developed to assist species inventory, including the technique of obtaining DNA from environmental samples (eDNA) in order to identify vertebrates through metabarcoding. Despite of being a promisor method, eDNA has not been tested to a great extent in complex aquatic environments, such as the Amazon. The aim of this study was to test the sensitivity of eDNA method in identifying species of Amazon fish using a metabarcoding approach both in a simulated environment (a closed system of aquariums) and in a natural system. We used four amplicons modelled to *cox1*, 16S and 12S mitochondrial genes to identify Actinopterygii group. Our results show that different sequencing protocols may influence DNA recovery rates of target species, considerably decreasing the number of valid sequences. However, the concurring use of four different amplicons allowed to increase the number of species identified with the eDNA method. COA primer was the best option among the other alternatives, identifying 13 of a total of 34 species in the aquarium system, although it has not identified with accuracy some congeners species. Amplicon size directly influenced species identification, indicating that sequences of less than 300 bp were not enough to sample the diversity of Amazon fish. Notwithstanding, both the number of individuals and their physiologic and behavioural characteristics seem to influence in the amount of genetic material found in an eDNA sample. We conclude that the protocol used may recover the information of some fish species in natural ecosystems, indicating that eDNA method may be used associated with other traditional methods to improve the inventory and monitoring of Amazon fish species.

Key-words: Environmental DNA, eDNA, ichthyofauna inventory, natural environment simulation, COA, 16S, 12S, Amazon fish, metabarcoding.

LISTA DE FIGURAS

Figura 1 – Fluxograma das etapas de trabalho desenvolvidas para a coleta e manipulação das amostras de eDNA.	17
Figura 2 – Sistema de aquários da empresa Arapaima Brazil Ltda utilizados neste experimento. Seta vermelha, linha de sistema de aquários selecionada; Seta amarela, cisterna de captação de água; Seta rosa, sistema de circulação e filtragem de água.	18
Figura 3 – Mapa indicativo da localização espacial do Rio Aporema, Amapá-Brasil. Círculos pretos representam os 15 pontos de coleta de amostras de eDNA.	20
Figura 4 – A esquerda, Sistema de filtragem do tipo Stericup acoplado a bomba a vácuo. A direita, cisterna de armazenamento de água utilizada para coleta de amostras de eDNA.	21
Figura 5 – Alinhamento dos quatro pares de primers indicando o local exato do fragmento de extensão para cada gene utilizando como modelo a espécie <i>Pygocentrus nattereri</i> . A = 12S, B = Citocromo Oxidase subunidade 1, C = 16S.	22
Figura 6 – Fluxograma das etapas de trabalho desenvolvidas para a elaboração da lista de espécies adquirida através de amostras de eDNA.	24
Figura 7 – Histograma de distribuição de sequências totais geradas pelas plataformas de sequenciamento para as amostras da comunidade simulada considerando cada um dos <i>primers</i> utilizados.	30
Figura 8 – Histograma de distribuição de sequências totais geradas pelas plataformas de sequenciamento para as amostras do rio Aporema considerando cada um dos <i>primers</i> utilizados.	30
Figura 9 – Classificação das sequências geradas após alinhamento. Em colunas, porcentagem de sequências não classificadas (NoHit) e classificadas (Hit, Filtro_100, Filtro_classe). Em círculos, porcentagem de sequências classificadas por grupos de organismos anterior (Hit) e, posterior ao filtro de remoção de 100bp (Remoção_100). Tabela ao centro, valores brutos das sequências analisadas.	31
Figura 10 – Distribuição dos valores de identidade das sequências removidas após a aplicação do filtro_100.	32
Figura 11 – Classificação das sequências geradas após alinhamento. Em colunas, porcentagem de sequências não classificadas (NoHit) e classificadas (Hit, Filtro_100, Filtro_classe). Em círculos, porcentagem de sequências classificadas por grupos de organismos anterior (Hit) e, posterior ao filtro de remoção de 100bp (Remoção_100). Tabela ao centro, valores brutos das sequências analisadas.	33
Figura 12 – Distribuição das sequências residuais após a aplicação do filtro_classe.	34
Figura 13 – Classificação das sequências geradas após alinhamento. Em colunas, porcentagem de sequências não classificadas (NoHit) e classificadas (Hit, Filtro_100, Filtro_classe). Em círculos, porcentagem de sequências classificadas por grupos de organismos anterior (Hit) e, posterior ao filtro de remoção de 100bp (Remoção_100). Tabela ao centro, valores brutos das sequências analisadas.	35
Figura 14 – Distribuição das sequências residuais após a aplicação do filtro_classe.	36

Figura 15 – Classificação das sequências geradas após alinhamento. Em colunas, porcentagem de sequências não classificadas (NoHit) e classificadas (Hit, Filtro_100, Filtro_classe). Em círculos, porcentagem de sequências classificadas por grupos de organismos anterior (Hit) e, posterior ao filtro de remoção de 100bp (Remoção_100). Tabela ao centro, valores brutos das sequências analisadas.	37
Figura 16 – Distribuição das sequências residuais após a aplicação do filtro_classe.	38
Figura 17 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.	39
Figura 18 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.	40
Figura 19 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.	41
Figura 20 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.	42
Figura 21 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.	43
Figura 22 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.	44
Figura 23 – Distribuição das sequências residuais após a aplicação do filtro_classe.	45
Figura 24 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.	46
Figura 25 – Distribuição das sequências residuais após a aplicação do filtro_classe.	46
Figura 26 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.	47
Figura 27 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.	48
Figura 28 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.	49
Figura 29 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.	49
Figura 30 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.	50
Figura 31 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.	50
Figura 32 – Distribuição das sequências residuais após a aplicação do filtro_classe.	51
Figura 33 – Distribuição das sequências residuais após a aplicação do filtro_classe.	52
Figura 34 – Cladogramas baseados nas sequências únicas das espécies encontradas nas amostras de eDNA.	52
Figura 35 – Contribuição de cada <i>primer</i> para o inventário das espécies da comunidade simulada.	54
Figura 36 – Classificação das sequências geradas após alinhamento. Em colunas, porcentagem de sequências não classificadas (NoHit) e classificadas (Hit, Filtro_100, Filtro_classe). Em círculos, porcentagem de sequências classificadas por grupos de organismos anterior (Hit) e, posterior ao filtro de remoção de 100bp (Remoção_100). Tabela ao centro, valores brutos das sequências analisadas.	55
Figura 37 – Distribuição das sequências atribuídas a espécie <i>Heros esfaciatus</i> em relação ao tamanho do fragmento, porcentagem de ID e número de espécies que alinharam as reads em NT.	63
Figura 38 – Alinhamento entre as sequências do gênero <i>Corydoras</i> amplificadas pelo <i>primer</i> 12S.	64

LISTA DE TABELAS

Tabela 1 – Lista das espécies e quantidade dos peixes alocados na comunidade simulada.....	18
Tabela 2 – Lista das espécies da comunidade simulada identificadas para cada <i>primer</i>	53
Tabela 3 – Lista das espécies descritas para o rio Aporema_Amapá/Brasil com os respectivos possíveis gêneros identificados pela técnica de eDNA.	56
Tabela 4 – Listas das possíveis ocorrências ainda não identificadas pelos pesquisadores do IEPA, porém, listada pela técnica de eDNA.	59
Tabela 5 – Classificação arbitrária das espécies da comunidade simulada quanto a; MOB=Mobilidade, TEMP= Temperamento solitário, TEMP_GRUPO=Temperamento em grupo, TECIDO=Tipo de tecido epidérmico, RESIST_TECIDO=Resistência do tecido. N=Número total de espécies; NSED=Número total de sequências consideradas para a espécie; * Espécie não identificada, porém, selecionada como candidata. + Espécie não identificada.	66

SUMÁRIO

1 INTRODUÇÃO.....	11
2 OBJETIVOS.....	16
2.1 OBJETIVO GERAL	16
2.2 OBJETIVOS ESPECÍFICOS	16
3 MATERIAL E MÉTODOS.....	17
3.1 COLETA E MANIPULAÇÃO DAS AMOSTRAS	17
3.1.1 Comunidade simulada.....	17
3.1.1.1 Seleção das espécies, coleta de tecido e aclimação	17
3.1.2 Ecossistema natural.....	19
3.1.1.2 Localização.....	19
3.1.2 Coleta das amostras	20
3.1.2.1 Comunidade simulada	20
3.1.2.2 Ecossistema natural	21
3.1.3 Extração do DNA.....	21
3.1.4 Amplificação do DNA	22
3.1.4.1 Seleção dos primers	22
3.1.5 Sequenciamento do DNA	23
3.2 MANIPULAÇÃO DE METADADOS	24
3.2.1 Remoção de sequências químéricas	24
3.2.2 Alinhamento Local - BLASTn	25
3.2.2.1 Elaboração do Banco Referencial.....	25
3.2.2.2 Alinhamento dos arquivos .fasta das amostras de eDNA.....	25
3.2.3 Seleção das sequências candidatas da comunidade simulada.....	26
3.2.5 Lista de espécies	27
3.2.6 Lista de espécie do rio Aporema.....	28
4 RESULTADOS	29
4.1 QUANTIDADE E QUALIDADE DAS SEQUÊNCIAS.....	29
4.2 IDENTIFICAÇÃO DAS ESPÉCIES DA COMUNIDADE SIMULADA.....	31
4.2.1 Distribuição das Sequências e Seleção das espécies candidatas.....	31
4.2.1.1 Primer 12S	31
4.2.1.2 Primer 16S	32
4.2.1.3 <i>Primer</i> COA	34
4.2.1.4 <i>Primer</i> COB.....	36
4.2.2. Validação das espécies candidatas.....	38

4.2.2.1 Primer 12S	38
4.2.2.2 <i>Primer</i> 16S.....	42
4.2.2.3 <i>Primer</i> COA	45
4.2.2.4 <i>Primer</i> COB.....	49
4.3 LISTA DE ESPÉCIES	52
4.4 APLICAÇÃO EM ECOSSISTEMA NATURAL.....	54
4.3.1 Distribuição das sequências	54
4.3.2 Validação das espécies.....	55
5 DISCUSSÃO	60
7 CONCLUSÕES	68
6 REFERÊNCIAS BIBLIOGRÁFICAS	69

1 INTRODUÇÃO

Os ecossistemas de água doce correspondem a apenas 3% do volume total de água da Terra e, são hoje considerados os mais ameaçados do planeta devido ao aumento constante da população humana, que leva ao aumento do desenvolvimento socioeconômico e, conseqüentemente, em diferentes usos dos recursos hídricos locais. Os fatores que levam a degradação destes ecossistemas estão relacionados principalmente ao uso indiscriminado dos corpos hídricos, podendo ser citados: a utilização dos mesmos como sumidouros de redes de esgoto, o que por sua vez, devido ao alto grau de conectividade da rede de tributários, facilita a dispersão rápida e generalizada de poluentes e sedimentos e, a captação da água para produção de energia hidrelétrica, irrigação, mineração e, sedentação animal¹

Mesmo sabendo que a enorme pressão imposta aos ecossistemas de água doce impacta diretamente a biodiversidade local, até o momento, não existem estudos que abranjam por completo toda a diversidade das espécies e o grau de ameaça a cada uma delas com inúmeras lagunas ainda existentes². A ictiofauna, por exercer diferentes serviços ecossistêmicos além de, ser importante fonte alimentar e de renda para a população humana, contribui como uma importante parcela da diversidade presente nestes ecossistemas³.

Existem hoje cerca de 16.000 espécies de peixes de água doce descritas, o que perfaz 48% das espécies do grupo e 25% das espécies de vertebrados. No entanto, por ocuparem apenas 1% da superfície terrestre, grande parte de seus habitats se encontram altamente impactados por processos antrópicos⁴.

A alta diversidade imposta ao grupo é determinada pela interação entre os processos hidrológicos, geomorfológicos, ecológicos e fisiológicos que atuam sobre cada organismo influenciando quais tipos (Famílias, Ordens, Gêneros, Espécies) podem sobreviver em determinado local¹.

A maior diversidade de peixes de água doce ocorre em ecossistemas neotropicais, com cerca de 5.400 táxons conhecidos⁵, porém, com uma taxa de descrição de novas ocorrências estimada para cada 3,5 dias, é esperado que o número real chegue em torno de 8.000 espécies⁶. Dentre as 51 famílias conhecidas, das quais 35 são endêmicas destas regiões, as mais representativas são Siluriformes, Characiformes, Cyprinodontiformes e Perciformes, sendo esta última a mais predominante⁵.

A bacia Amazônica detém quase metade da diversidade da ictiofauna dos ecossistemas neotropicais, o que representa cerca de 2.411 espécies sendo que, pelo menos 1.089 endêmicas⁷.

Dentre as principais ameaças a esta comunidade que incluem, as causas já citadas decorrentes do desenvolvimento humano, a implantação de megaprojetos de geração de energia com base em hidrelétricas seja, atualmente, a mais eminente. As hidrelétricas representam uma das maiores ameaças à biodiversidade aquática devido ao barramento do corpo hídrico o que acarreta em alterações hidrológicas generalizadas levando a, comprometimento das rotas migratórias, degradação de habitat, obstrução do fluxo gênico devido a redução da conectividade fluvial, favorecimento de espécies generalistas sobre espécies especialistas, aumento do risco de extinção de espécies endêmicas e, favorecimento da colonização por espécies exóticas⁸.

Atualmente, 416 barragens estão em operação ou em construção dentro da bacia e, mais 334 estão sendo propostas⁹, significando que apenas três afluentes da Amazônia estariam livres da influência de barragens nas próximas décadas¹⁰. Além da imensurável perda de biodiversidade, inúmeras espécies podem vir a ser extintas antes de serem conhecidas pela ciência¹¹⁻¹³. Agravando este processo, como principal medida regulatória para a implantação de empreendimentos potencialmente causadores de impactos ambientais, tais como os oriundos da construção de hidrelétricas, a legislação brasileira é condicionada pelas Normativas Conama nº1/86 e 237/87 e, Instrução Normativa Ibama nº146/07, que tratam do licenciamento ambiental de tais atividades pelo órgão público. Uma das obrigatoriedades é o diagnóstico da diversidade zoológica onde fica estabelecido que, para o início das operações os empreendimentos devem ser responsáveis por informar o inventário das espécies locais para que, medidas de mitigação sejam impostas. No entanto, o prazo definido pelo órgão público para a execução destes estudos é baseado na sazonalidade de cada região e não na capacidade de detecção da técnica utilizada. Esta lacuna da lei permite que estudos acerca da caracterização de comunidades biológicas inteiras, em sua grande maioria, sejam realizados em prazos inferiores a um ou dois anos o que, para muitos dos grupos taxonômicos, não seria adequado ao seu inventário para cálculos de impacto.

De fato, as grandes lacunas de conhecimento acerca da diversidade da ictiofauna na bacia Amazônica estão relacionadas aos altos custos para implantação de projetos de pesquisas básicas associadas ao baixo investimento governamental para o desenvolvimento das mesmas. Isto torna tanto inventários como monitoramentos morosos já que, devido à sua grande extensão, existe a presença de inúmeros microhabitats resultantes de diferenças topográficas e geológicas¹¹

Idealmente, a melhor técnica para se avaliar a estrutura de uma comunidade deve caracterizar 100% dos parâmetros ecológicos básicos que são expressos por composição das

espécies e abundâncias relativas¹⁴. No entanto, estes parâmetros são sensíveis ao esforço amostral e, determinar o número mínimo ou máximo de amostragens necessárias para identificar todos os integrantes de uma comunidade torna-se altamente dependente de fatores bióticos e abióticos¹⁵. Em se tratando da comunidade de peixes, os fatores abióticos são representados pelas inúmeras nuances que formam o habitat de cada espécie, ou seja, tamanho e profundidade do corpo hídrico associados aos tipos de formação vegetal, geológica e topográfica. Já, em relação aos fatores bióticos, existe a necessidade de saber o padrão de distribuição de cada espécie o que envolveria, as características de tamanho e fisiologia, capacidade de migração e comportamento¹⁵.

Dentre as técnicas mais utilizadas para o inventário da ictiofauna estão aquelas baseadas no sucesso de captura de indivíduos e podem ser citadas, o uso de redes de espera, peneiras, redes de arrasto, tarrafas, espinheis e, eletropesca¹⁶. No entanto, além de todas serem invasivas, considerando os padrões dos corpos hídricos amazônicos, para que se obtenha em curto tempo, um inventário completo de todas as espécies presentes se faz necessário a associação de todas elas. Este contraponto, acarreta em elevado custo-benefício já que, é necessário aquisição constante de material, número adequado de integrantes da equipe capacitados para o reconhecimento das espécies e disponibilidade de logística adequada ao número de vezes necessários para aplicação das técnicas.

Contudo, novas abordagem que permitam, de forma não invasiva, o fácil acesso à comunidade de peixes e que sejam desempenhadas com alta acurácia estão em constante aprimoramento. Atualmente, duas novas técnicas são apontadas como eficazes para complementar o inventário da ictiofauna: o uso de câmeras subaquáticas¹⁷ e a detecção de DNA das espécies em amostras ambientais (eDNA)¹⁸.

A técnica de eDNA, apesar de a mesma já vir sendo utilizada há mais de três décadas por microbiologistas¹⁹ somente após a publicação do manuscrito de Ficetola et al. (2008), iniciou-se o desenvolvimento de estudos com organismos vertebrados^{18,20-22}, o que levou a mesma ser apontada em 2013 pelo periódico *Trends in Ecology and Evolution*²³, como um dos 15 horizontes globais para a conservação das espécies.

Sem a necessidade de coleta do espécime, o material genético é obtido de uma amostra ambiental (água, solo, sedimentos) passível de conter restos celulares provindos da regeneração/decomposição de tecidos epidérmicos. Como a origem e tamanho da partícula contendo o material genético é incerta²⁴, este pode vir a sobreviver de horas a milhares de anos, dependendo da ação imposta pelo meio ambiente²⁵, se presente em uma amostra ambiental, pode ser detectado e a espécie identificada através de análises moleculares²⁶⁻²⁸. Desta maneira,

uma única amostragem de eDNA poderá conter representantes de diferentes organismos, tornando a técnica eficaz como apoio ao inventário e monitoramento dos mais diversos grupos taxonômicos^{9,10}.

Por se tratar de coleta de dados através de material genético exógeno, a possibilidade de reprodução de dados falso-positivos e/ou negativos tornam-se um dos pontos cruciais da técnica de eDNA e de difícil resolução, já que a forma de coleta e as ações do meio ambiente exercem influências diretas na qualidade da amostra²⁹⁻³¹.

Há diferentes técnicas de coleta e obtenção das amostras, como filtração e/ou centrifugação. Em casos de filtração, o tamanho do poro do filtro assim como manipulações laboratoriais quanto ao processo de extração e amplificação do DNA, podem incidir em resultados diferentes, gerando maior ou menor concentração de DNA^{32-35,35-41}.

Variáveis ambientais abióticas como, temperatura, pH, turbidez e, salinidade contribuem de formas diferentes nas taxas de degradação das moléculas de DNA para cada espécie, podendo estas permanecerem durante horas a dias no meio ambiente^{42,43}. Variáveis bióticas dentre elas a degradação do DNA pela atividade bacteriana e as ações de enzimas nucleases presentes nos corpos hídricos podem desnaturar o DNA exógeno em poucas horas^{24,44}, assim como, diferentes tipos de organismos podem ter taxas de liberação de DNA diferentes, devido a características intrínsecas de cada espécie^{26,45,46} além de que, podem ser mais ou menos abundantes^{47,48} e/ou migratórios^{22,41}.

O DNA pode também estar presente por ter sido carregado por agentes indiretos, tais como, parte da ingesta de outros animais, água de lastro de embarcações, resíduos da alimentação humana e contaminação de amostras durante os procedimentos laboratoriais^{37,49}.

Desta forma, o resultado de cada amostra deve analisar conjuntamente as informações básicas acerca do local de amostragem, suas características físico/químicas, possíveis níveis tróficos existentes, volume de água/sedimento coletado e, a amplitude da cobertura desejada quando ao nível taxonômico, sendo que este último irá depender exclusivamente do tipo de sequenciamento aplicado ao DNA, se *shotgun* (metagenômica – uso completo das sequências de DNA extraídas da amostra) ou *amplicon* (metabarcoding – uso de pequenas sequências de DNA espécie-específicas)³⁴.

Como já foi mencionado, em uma amostra de eDNA pode haver material genético de inúmeros grupos taxonômicos e, portanto, é usual em estudos focados em organismos mais complexos, tais como vertebrados, que o sequenciamento seja baseado em *amplicons*. Apesar de restringir a amplitude da técnica, como o DNA presente nas amostras referentes a estes organismos costuma estar altamente fragmentado, esta abordagem favorece o direcionamento

do estudo ao grupo/espécie alvo já que a quantidade de material genético em uma amostra ambiental é também reduzido³⁴.

O melhor modelo de *amplicom* ainda gera inúmeros questionamentos⁵⁰. Alguns autores alegam que o foco no DNA provindo do núcleo celular é mais eficaz, pois, devido ao menor número de cópias, tende a ser degradado mais rapidamente e seu encontro em uma amostra seria indicativo de material genético liberado recentemente no meio ambiente, reduzindo a possibilidade de cometimento de erros falso-positivos^{49,51}, no entanto, os *amplicons* modelados para os genes mitocondriais são os mais utilizados⁵²⁻⁵⁷. Dentre as regiões do DNA mitocondrial atualmente descritas podem ser citadas como as mais comuns o o gene Citocromo Oxidase subunidade 1 (*cox-1*)^{32,37,58-62}, o gene 12S^{46,49}, o gene 16S^{49-52,60,63}, o gene Citocromo *b*⁶⁴ e, a região controle D-loop³⁶.

Mesmo que a técnica de eDNA já esteja, aparentemente, preestabelecida em alguns países do Hemisfério Norte e sendo utilizada por diversos grupos de pesquisa para o monitoramento de populações naturais^{44,48,52,64,65}, estudos de base ainda são extremamente necessários para o estabelecimento da mesma em ecossistemas tão complexos quanto a bacia do rio Amazonas.

Até o momento, a grande maioria dos estudos realizados com eDNA voltados ao inventário e monitoramento da ictiofauna, ocorreram em regiões de clima temperado, com baixa diversidade de espécies e, considerando as regiões de clima subtropicais eles ainda são raros⁶⁶, com poucos se aventurando a inventariar áreas abertas com maior ocorrência de espécies e, portanto, maiores dificuldade quanto a identificação das espécies⁶⁷.

Desta forma, por se tratar de uma técnica recente, que ainda não tem estabelecidos protocolos básicos para que tanto os estudos já existentes como para que os futuros possam ser comparados entre si, é necessário elucidar as questões acima apontadas. Além disso, a aplicação da mesma ainda é enigmática em ecossistemas neotropicais, especialmente para os amazônicos, onde uma simples técnica já consagrada para outras regiões tende a não se adequar perfeitamente devido as fortes variações climáticas.

Portanto, estudos como este, que buscam entender tanto as possíveis variações bem como, as restrições da técnica de eDNA através de simulações das populações naturais, são de extrema importância e se caracterizam como uma iniciativa necessária para o desenvolvimento mesma na Amazônica.

2 OBJETIVOS

2.1 OBJETIVO GERAL

O objetivo geral deste estudo foi testar em comunidade simulada a sensibilidade da técnica de eDNA para identificar espécies de peixes amazônicos utilizando a abordagem de metabarcoding.

2.2 OBJETIVOS ESPECÍFICOS

Como objetivos específicos, nos propomos a:

- Avaliar em comunidade simulada a sensibilidade de quatro *amplicons* modelados para os genes *cox-1*, 16S e 12S quanto a identificação do grupo Actinopterygii;
- Avaliar em comunidade simulada a especificidade de quatro *amplicons* modelados para os genes *cox-1*, 16S e 12S quanto a classificação a nível de espécie;
- Avaliar em comunidade simulada a acurácia da técnica de eDNA em inventariar a ictiofauna;
- Testar os protocolos estabelecidos para a comunidade simulada em ecossistema natural.

3 MATERIAL E MÉTODOS

3.1 COLETA E MANIPULAÇÃO DAS AMOSTRAS

Esta etapa consistiu na aquisição das amostras de eDNA e seguiu o fluxo de trabalho exibido na figura 1 sendo seus por menores detalhados a baixo.

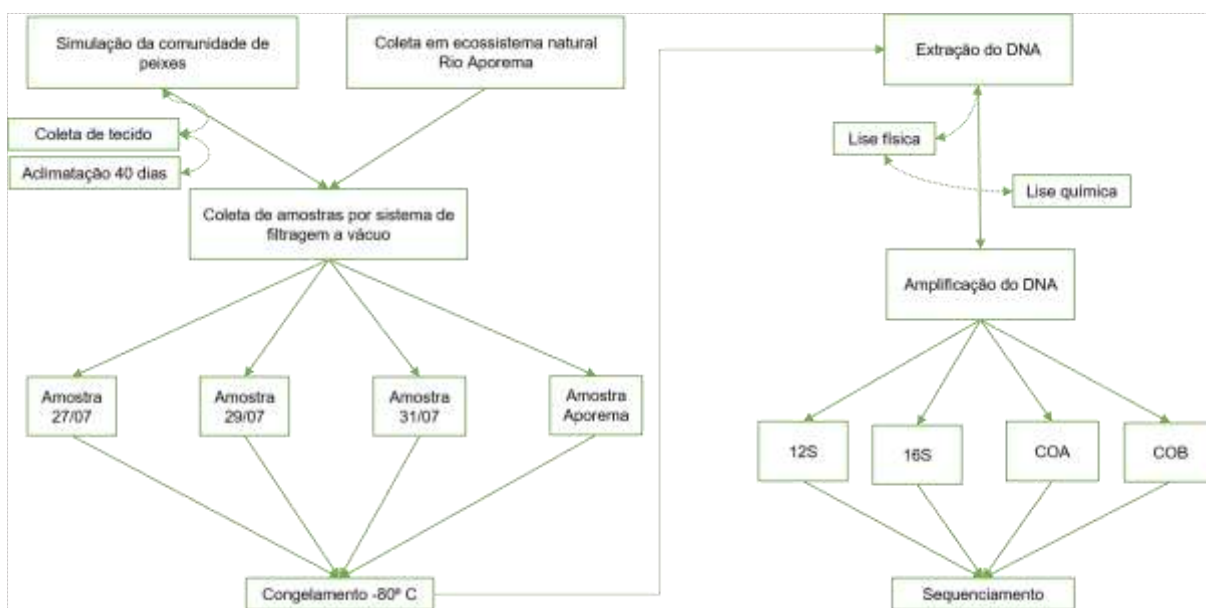


Figura 1 – Fluxograma das etapas de trabalho desenvolvidas para a coleta e manipulação das amostras de eDNA.

3.1.1 Comunidade simulada

3.1.1.1 Seleção das espécies, coleta de tecido e aclimação

Para elaborar o experimento, foram utilizadas as instalações da empresa Arapaima Brazil Ltda, as quais são estruturadas em sistemas integrados compostos por 56 aquários, cada um com capacidade de 120 litros de água cada um deles (Figura 2). A captação de água provém de cisternas com capacidade de 7.400 litros construídas exclusivamente para abastecer cada sistema. A circulação da água da cisterna para o sistema de aquários é feita através de dutos coletores e bombas de propulsão que promovem a distribuição e recaptação da água entre cada aquário, formando um sistema fechado de fornecimento de água. A água recaptada, ao retornar à cisterna, passa através de filtros biológicos anterior à sua recirculação para os aquários.



Figura 2 – Sistema de aquários da empresa Arapaima Brazil Ltda utilizados neste experimento. Seta vermelha, linha de sistema de aquários selecionada; Seta amarela, cisterna de captação de água; Seta rosa, sistema de circulação e filtragem de água.

A empresa tem seu principal foco no comércio de peixes ornamentais e chega a acondicionar cerca de 60 espécies de peixes exclusivamente amazônicos anualmente. Com a finalidade de reproduzir um pequeno curso d'água, 34 espécies de peixes, pertencentes a Classe Actinopterygii, foram selecionadas de acordo com a diversidade, ou seja, famílias, gêneros e espécies de maior ocorrência para a região amazônica⁷ (Tabela 1).

Tabela 1 – Lista das espécies e quantidade dos peixes alocados na comunidade simulada

DIVERSIDADE					N	
ORDEM	FAMILIA	SUBFAMILIA	GENERO	ESPECIE		
Characiformes	Acestrorhynchidae	Heterocharacinae	Heterocharax	<i>Heterocharax virgulatus</i>	451	
		Heterocharacinae	Gnathocharax	<i>Gnathocharax steindachneri</i>	502	
	Anostomidae	Anostominae	Pseudanos	<i>Pseudanos trimaculatus</i>	15	
	Characidae	Incertae sedis	Astyanax	<i>Astyanax fasciatus</i>	1	
			Hemigrammus	<i>Hemigrammus rhodostomus</i>	60	
			Hyphessobrycon	<i>Hyphessobrycon bentosi</i>	503	
				<i>Hyphessobrycon heterorhabdus</i>	51	
			Moenkhausia	<i>Moenkhausia oligolepis</i>	15	
			Cheirodontinae	Serrapinnus	<i>Serrapinnus kriegi</i>	4
			Chilodontidae	Incertae sedis	Chilodus	<i>Chilodus punctatus</i>
	Gasteropelecidae	Incertae sedis	Gasteropelecus	<i>Gasteropelecusylevis</i>	51	
	Serrasalmidea	Incertae sedis	Myloplus	<i>Myloplus rubripinnis</i>	10	
			Pygocentrus	<i>Pygocentrus nattereri</i>	1	
Serrasalmus			<i>Serrasalmus marginatus</i>	1		
Gymnotiformes	Apteronotidae	Incertae sedis	Apteronotus	<i>Apteronotus albifrons</i>	107	
Perciformes	Cichlidae	Cichlasomatinae	Acaronia	<i>Acaronia nassa</i>	2	
			Laetacara	<i>Laetacara dorsigera</i>	96	
			Mesonauta	<i>Mesonauta acora</i>	19	
		Cichlinae	Crenicichla	<i>Crenicichla johanna</i>	67	
		Satanoperca	<i>Satanoperca jurupari</i>	1		

DIVERSIDADE					N
ORDEM	FAMILIA	SUBFAMILIA	GENERO	ESPECIE	
Perciformes	Cichlidae	Geophaginae	Guianacara	<i>Guianacara dacrya</i>	10
			Apistogramma	<i>Apistogramma agassizii</i>	818
				<i>Apistogramma gephyra</i>	50
		Cichlasomatinae	<i>Apistogramma rupununi</i>	652	
			Heros	<i>Heros efasciatus</i>	4
			Pterophyllum	<i>Pterophyllum scalare</i>	25
	Polycentridae	Incertae sedis	Monocirrhus	<i>Monocirrhus polyacanthus</i>	9
Siluriformes	Callichthyidae	Corydoradinae	Corydoras	<i>Corydoras ephippifer</i>	30
				<i>Corydoras julii</i>	30
Siluriformes	Aspredinidae	Aspredininae	Platystacus	<i>Platystacus costatus</i>	6
Siluriformes	Loricariidae	Ancistrinae	Panaque	<i>Panaque armbrusteri</i>	2
				<i>Farlowella henriquei</i>	5
		Hypostominae	Hypostomus	<i>Hypostomus faveolus</i>	4
			Peckoltia	<i>Peckoltia vittata</i>	1

Após definição das espécies a serem utilizadas, cada uma foi alocada isoladamente em um dos aquários do sistema permanecendo lá sem quaisquer modificações por 40 dias. A aclimatação foi baseada no estudo de Ficetola et al. (2008) que indicou que o tempo suficiente para se equilibrar tanto as perdas por degradação de DNA quanto o ganho por sua produção, promovendo uma constante entre os dois valores, seria de ao menos sete dias. Um exemplar de cada uma das espécies foi coletado como modelo padrão para elaboração do banco de sequências referenciais a ser utilizado nas futuras análises. Desta forma, cada indivíduo foi eutanasiado com anestésico e subsequentemente acondicionado em sacos plásticos do tipo “ziplock”. As amostras foram então congeladas em freezer -20°C até o momento dos procedimentos laboratoriais.

3.1.2 Ecossistema natural

3.1.1.2 Localização

O Aporema é um tributário do rio Araguari localizado na região nordeste do estado do Amapá. Possui extensão de 75 quilômetros sendo que, grande parte de sua área está inserida na Reserva Ambiental Horto São Bento, mantida pela empresa Amapá Florestal e Celulose S/A. A empresa desenvolve inúmeros projetos ligados a conservação e manejo da biodiversidade local, sendo um deles o monitoramento de vertebrados aquáticos. Desta forma, como a ictiofauna local já vem sendo monitorada através das pesquisas realizadas pela parceria entre o Instituto de Pesquisas do Estado do Amapá (IEPA) e a empresa, quinze pontos amostrais anteriormente já inventariados foram selecionados para este projeto (Figura 3).

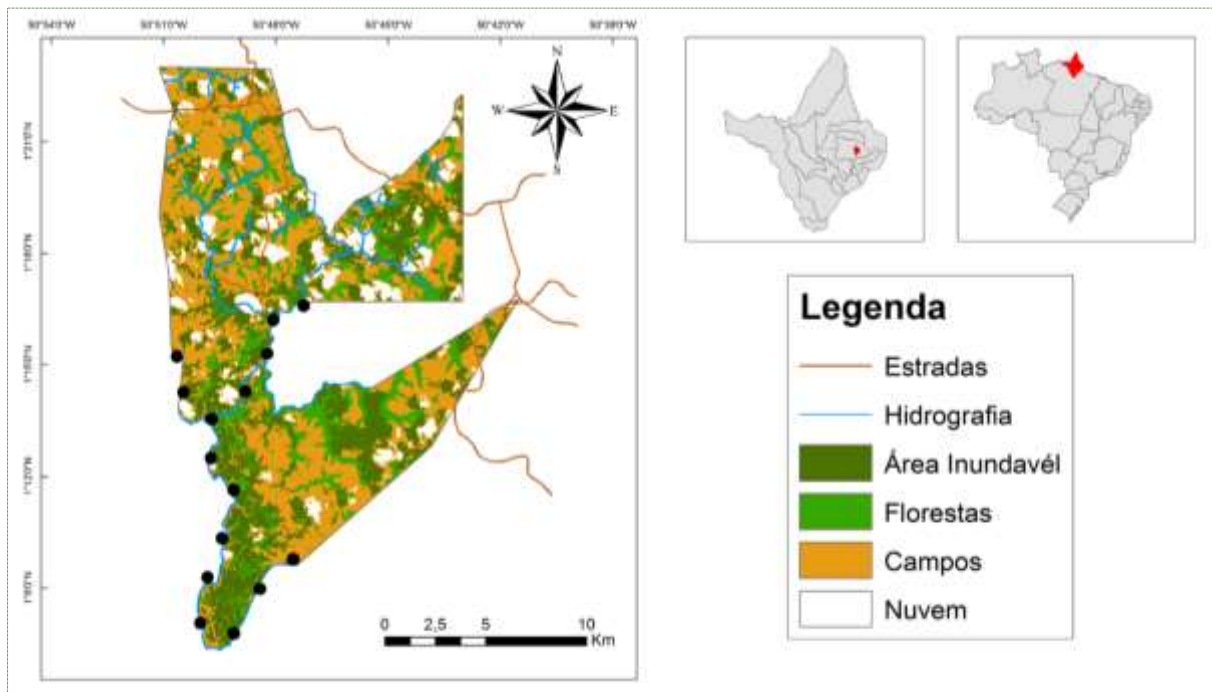


Figura 3 – Mapa indicativo da localização espacial do Rio Aporema, Amapá-Brasil. Círculos pretos representam os 15 pontos de coleta de amostras de eDNA.

3.1.2 Coleta das amostras

3.1.2.1 Comunidade simulada

Para evitar contaminação externa das amostras, nove garrafas com capacidade de 500ml foram lavadas e esterilizadas em autoclave anterior as coletas. A concentração de cada amostra, durou cerca de 30 minutos e foi realizada com o auxílio de sistema de filtragem do tipo Stericup 500 ml, membrana de $0,22\mu\text{m}$ – Millipore[®] e bomba a vácuo NOF-650 - Newpump[®], na velocidade de vazão de 55L/min e, vácuo de 650mmHg/85,5Kpa (Figura 4). A coleta de amostras ocorreu em três dias e três horários diferentes (27, 29 e 31 de julho de 2015; 8:00, 12:00 e 17:00 horas), totalizando novo amostras de cinco litros por período, removidos diretamente da cisterna de armazenamento.



Figura 4 – A esquerda, Sistema de filtração do tipo Stericup acoplado a bomba a vácuo. A direita, cisterna de armazenamento de água utilizada para coleta de amostras de eDNA.

Ao final de cada filtração, as membranas foram removidas com auxílio de lâminas de bisturi estéreis, acondicionadas em tubos plásticos de 50ml e imediatamente congeladas em freezer -20°C até o final de cada dia de amostragem. Para o transporte até o laboratório, as amostras foram acondicionadas em caixa de isopor envoltas por bolsas de gelo reutilizáveis e após, acondicionadas em freezer -80°C até o momento dos procedimentos laboratoriais.

3.1.2.2 Ecossistema natural

Para evitar contaminação externa das amostras, quinze galões com capacidade de 5 litros foram lavados e esterilizados em autoclave anterior a coleta. Tanto a concentração de cada amostra, que durou cerca de duas horas/amostra, quanto o transporte e conservação seguiram o mesmo protocolo já descrito para a comunidade simulada.

A coleta das amostras ocorreu nos dias 14 e 15 de fevereiro de 2017 onde foram coletados cinco litros de água removidos a cerca de 40cm de profundidade do canal central do rio totalizando 15 amostras.

3.1.3 Extração do DNA

O processo de extração do DNA das amostras de água se baseou em duas abordagens que foram: i) lise física, com a finalidade de fragmentação da membrana de polietileno; e ii) lise química, com a finalidade de romper a parede celular e liberar o DNA.

Para fragmentar as membranas de polietileno, foram inseridas 10 esferas de aço-inox de 3 mm de diâmetro aos tubos contendo as amostras. Os tubos foram então submersos em nitrogênio líquido por 10 segundos e agitados em vórtex por 30 segundos. Este procedimento foi repetido por três vezes para cada uma das amostras.

A lise química foi realizada seguindo o protocolo de purificação de amostras de DNA em grandes volumes fornecido pelo kit comercial E.Z.N.A. *Stool DNA Kit*, Omega Biotek[®],

sem qualquer alteração. Ao final, as amostras foram eluídas em 100µl de tampão fornecido pelo próprio kit e quantificadas com o fluorômetro Qubit Thermo Fischer Scientific®.

O processo de extração de DNA das amostras de tecido seguiu o mesmo protocolo de lise química já mencionada acima, porém, reproduzindo o protocolo de pequenas amostras fornecido pelo próprio kit sem qualquer alteração.

3.1.4 Amplificação do DNA

3.1.4.1 Seleção dos primers

A seleção dos *primers* para amplificação das amostras foi feita com base em estudos previamente já realizados com filogenia baseada em caracteres moleculares de peixes. Desta forma, foram selecionados quatro pares de *primers* específicos para a amplificação de genes mitocondriais sendo eles: i) FFd1/FRd2, adquirido de Ivanova et al. 2007⁶⁸ com capacidade de extensão de 650pb do gene Citocromo Oxidase subunidade 1, daqui em diante referido como COA; ii) L6730/H6855, adquirido de Miya e Nishida, 2000⁶⁹ com capacidade de extensão de 150pb do gene Citocromo Oxidase subunidade 1, daqui em diante referido como COB; iii) 12SV5F/12SV5R, adquirido de Riaz et al. 2011⁷⁰ com capacidade de extensão de 100pb do gene 12S, daqui em diante referido como 12S; iv) 16sarL/16sbrH, adquirido de Palumbi, 1991⁷¹ com capacidade de extensão de 650pb do gene 16S, daqui em diante referido como 16S (Figura 5).

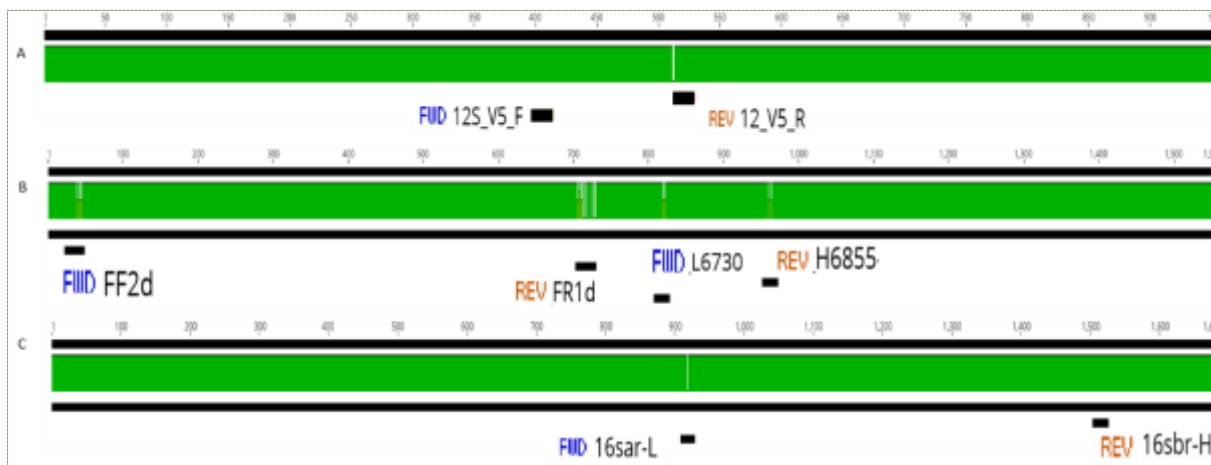


Figura 5 – Alinhamento dos quatro pares de primers indicando o local exato do fragmento de extensão para cada gene utilizando como modelo a espécie *Pygocentrus nattereri*. A = 12S, B = Citocromo Oxidase subunidade 1, C = 16S.

3.1.4.2 Procedimento de amplificação

As amostras foram amplificadas de forma padrão para a reação em cadeia da polimerase (PCR) utilizando o kit comercial FastStart™ High Fidelity PCR - Roche® para um volume final de 25µl onde, por reação foi adicionado 18,75µl de água pura, 2,5µl do Buffer 2, 0,5µl de DNTP (PCR Nucleotide mix - Roche®), 1µl de cada *primer*, 0,25µl do Buffer 1 (Polimerase) e, 1µl da amostra. As reações foram amplificadas no termociclador GeneAmp Thermal Cycler 9700 – Applied Biosystems®, conforme as temperaturas e tempos de desnaturação, anelamento e extensão determinados pela literatura consultada. Para avaliar a qualidade da amplificação, as amostras foram submetidas à eletroforese em gel de agarose 1% utilizando o E-Gel Electrophoresis System – ThermoFisher Scientific®.

3.1.5 Sequenciamento do DNA

Dois plataformas distintas foram utilizadas para o sequenciamento das amostras de água e tecidos sendo elas, Genome Sequencer GS FLX 454 – Roche® e, Ion Torrent PGM – Thermo Fisher Scientific®. A opção por distintas plataformas se baseou no tamanho de fragmento gerado, desta forma, os *primers* COA e 16S foram sequenciados na plataforma GS FLX 454 com capacidade de leitura de até 600pb e, os *primers* COB e 12S na plataforma Ion Torrent PGM com capacidade de leitura de até 400pb. Esta abordagem permitiu que as amostras fossem sequenciadas sem a necessidade de fragmentação evitando assim a formação de sequências quimeras devido ao processo de montagem do DNA posterior ao sequenciamento.

As duas plataformas têm protocolos similares e, portanto, para melhor caracterizar este tópico foi optado pela apresentação resumida dos principais procedimentos omitindo assim, o nome específico dos kits utilizados. Desta forma, a preparação para o sequenciamento das amostras procedeu da seguinte forma:

- Preparação da biblioteca: procedimento de reparado das pontas 3' e 5' das amostras para o recebimento de adaptadores e barcodes específicos, permitindo assim que amostra seja identificada posteriormente.
- Quantificação por qPCR para padronização equimolar: procedimento de quantificação das bibliotecas geradas por PCR em tempo real para definir a concentração final de DNA a ser utilizado na PCR emulsão.
- PCR de emulsão: procedimento de ligação de cada biblioteca a uma “bead” que é então, emulsionada para que fique aprisionada individualmente dentro de uma “bolha” de óleo onde irá ocorrer a reação de amplificação gerando assim, para cada biblioteca milhões de sequências idênticas.

▪ **Reação de Sequenciamento:** procedimento de quebra de emulsão e recuperação das “beads” que são então adicionadas a uma mistura de enzimas e nucleotídeos que, posteriormente são carregadas em uma placa PicoTiter (454) e/ou Chip Kit v2 BC (PGM Ion), onde ocorrerá a leitura das sequências pela plataforma de sequenciamento.

3.2 MANIPULAÇÃO DE METADADOS

Esta etapa consistiu no tratamento computacional dado as amostras de eDNA e seguiu o fluxo de trabalho exibido na figura 6 sendo seus por menores detalhados a baixo.

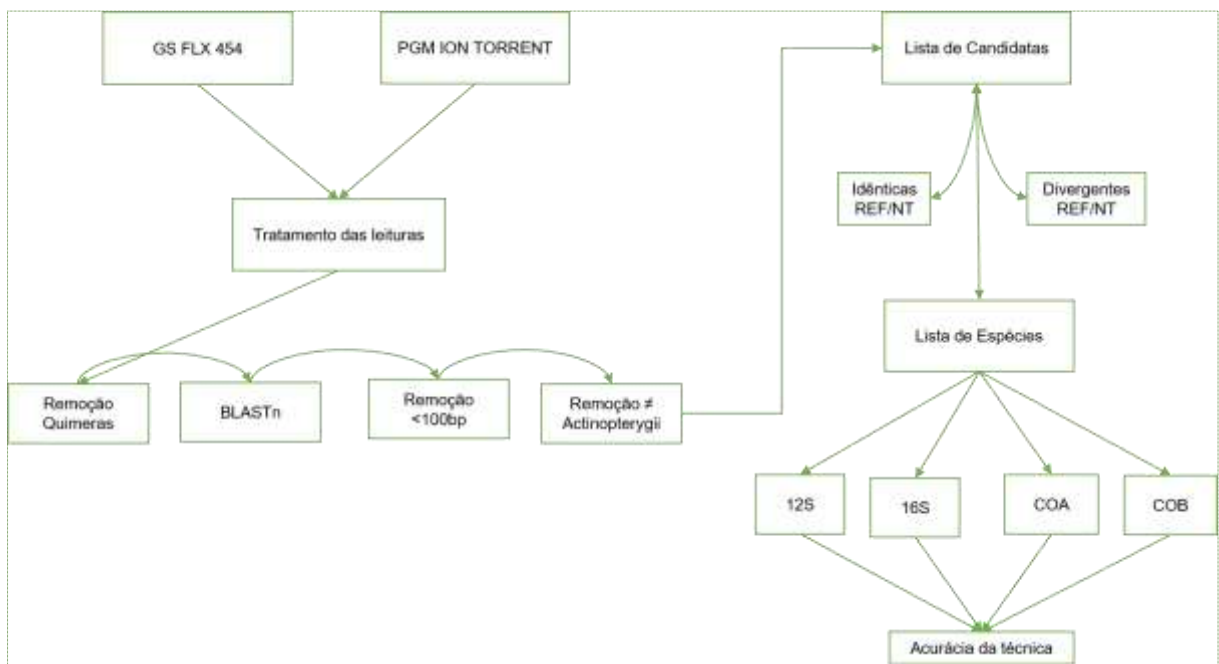


Figura 6 – Fluxograma das etapas de trabalho desenvolvidas para a elaboração da lista de espécies adquirida através de amostras de eDNA.

3.2.1 Remoção de sequências quiméricas

Como medida de padronização das leituras geradas pelas plataformas de sequenciamento, as amostras de eDNA tiveram os adaptadores e barcodes removidos (procedimento específico para a plataforma 454 utilizando a ferramenta “remoção de barcodes” contida no software Geneious®) e, seus arquivos do tipo .fastq transformados para .fasta pelo software Geneious®. Após este procedimento, foram removidas as potenciais sequências quiméricas através do uso da ferramenta “chirema” – Mothur®.

3.2.2 Alinhamento Local - BLASTn

3.2.2.1 Elaboração do Banco Referencial

Para compilação de um banco de sequências referentes a cada espécie presente na comunidade simulada, as amostras dos tecidos sequenciadas foram submetidas ao mesmo protocolo das amostras de eDNA, ou seja, remoção de adaptadores e barcodes, e transformação de arquivos .fastq para .fasta, como já descrito acima. Em seguida, foram removidas as sequências redundantes com a ferramenta “unique.seqs” – Mothur[®].

Para confirmação e validação das sequências únicas de cada espécie, foi realizado o referenciamento de cada uma delas utilizando a ferramenta “assembly – map to the reference” – Geneious[®]. Como referência, foi realizada busca na plataforma GenBank por sequências completas dos genes mitocondriais utilizados, ou seja, 12S, 16S e *cox-1* para cada uma das espécies da comunidade simulada. Na ausência de sequências para a espécie foi optado pelo nível taxonômico mais próximo, podendo ser este gênero ou subfamília. Após o referenciamento, a sequência consenso gerada foi extraída e alinhada com seus respectivos pares de *primers*, suas extremidades foram ajustadas ao tamanho de amplificação desejado e a sequência final foi recortada.

Para averiguar se a referência estava corretamente anotada, a sequência final foi alinhada na plataforma NCBI utilizando a ferramenta BLASTn com parâmetros selecionados para o “default” do algoritmo megablast (threshold 10, word_size 28, match/mismatch 1,-2, gap_costs linear). Foram avaliados neste procedimento os seguintes critérios de alinhamento da sequência referência: i) se houve alinhamento com a mesma espécie, ii) se houve alinhamento com o mesmo gênero, iii) se houve alinhamento com a mesma família e, iv) se houve alinhamento com a Classe Actinopterygii.

Nos casos onde não houve alinhamento com a mesma espécie, ou seja, identidade de 100%, foi utilizado como referência a porcentagem de alinhamento gerada, seguido do total de bases alinhadas e valor de cobertura (e-value). Este procedimento foi necessário devido a existência de algumas sequências geradas ainda não terem qualquer referência nucleotídica na base de dados do Genbank. Uma vez validadas as sequências referenciais para cada espécie, foi gerado o banco de dados “REF” utilizando a ferramenta “make.blastDB” – Mothur[®].

3.2.2.2 Alinhamento dos arquivos .fasta das amostras de eDNA

Anterior ao alinhamento das sequências das amostras de eDNA, todas as nove amostras da comunidade simulada, assim como as 15 amostras do rio Aporema, foram unificadas através

da ferramenta “merge.file” – Mothur[®], gerando um arquivo .fasta único para cada par de *primer* utilizado.

O alinhamento local do arquivo gerado para cada *primer* referente à comunidade simulada foi realizado através da ferramenta BLASTn com parâmetros selecionados para o “default” do algoritmo megablast. Como sequências referenciais foram utilizados o banco REF e o banco de sequências nucleotídicas do GenBank, daqui em diante referido como NT, enquanto que para o alinhamento dos arquivos gerados para os *primers* 12S e 16S referente ao rio Aporema foi utilizado apenas o banco NT.

Para obter a classificação taxonômica das sequências alinhadas por NT foi utilizada a ferramenta “Taxonkit” – NCBI. Desta forma, o resultado do banco NT continha, além das informações técnicas sobre a qualidade do alinhamento, as informações detalhadas quanto a classificação taxonômica de todos os organismos anotados.

3.2.3 Seleção das sequências candidatas da comunidade simulada

Para gerar a lista das sequências candidatas, ou seja, aquelas que potencialmente se referem as espécies da comunidade simulada, dois procedimentos foram realizados na tentativa de evitar tanto o cometido de classificação falso-positivas quanto falso-negativas, sendo estes: i) remoção das sequências de tamanho inferior a 100pb, daqui em diante referido como filtro_100, e ii) criação de dois grupos taxonômicos gerados a partir dos critérios de classificação do banco NT, sendo um formado por todas as sequências anotadas para Actinopterygii e outro formado pelos demais organismos celulares, daqui em diante referido como filtro_classe.

A remoção das sequências de tamanho menores do que 100pb foi arbitrária, por se considerar que fragmentos muito pequenos tendem a perder sinal filogenético e, possivelmente serem artefatos da técnica de sequenciamento.

A remoção ou não das sequências após aplicado o filtro_classe foi baseada em comparações entre os resultados dos alinhamentos dos dois bancos referenciais. Para realizar tais comparações, as seguintes informações foram utilizadas:

- read: código criado pela plataforma de sequenciamento para identificar cada uma das sequências lidas e, portanto, utilizada como conectora para unir os arquivos REF e NT;
- SP: táxon atribuído a cada read após o alinhamento do arquivo .fasta, nomeadas como SP_REF e SP_NT;
- ID: valor expresso em porcentagem que indica o quanto a read foi similar a sequência referência, nomeadas como ID_REF e ID_NT;

- ALG: número total de bases da read alinhadas com a sequência referência, nomeadas como ALG_REF e ALG_NT;

- SC: número de cobertura do alinhamento entre a read e a sequência referência levando em consideração o total de bases alinhadas e as penalidades sofridas entre gaps e ou deleção/substituição de bases, nomeadas como SC_REF e SC_NT.

3.2.4 Geração da lista de espécies da comunidade simulada

Para gerar a lista de espécies identificadas para cada *primer* os mesmos parâmetros da abordagem anterior foram empregados com a adição de remoção de todas as sequências com valores de e-value superior a 0,0001.

Desta forma, como algumas das espécies da comunidade simulada ainda não continham sequências mitocondriais publicadas para todos os *primers* na base de dados do Genbank e/ou, poderia existir a possibilidade de que a sequência REF divergisse devido a variações intraespecíficas de cada espécie. Para considerar um táxon validado ou não, foram utilizados os seguintes critérios:

- táxon validado por similaridade, ou seja, quando todas as reads foram coincidentes entre REF e NT;

- táxon validado com similaridade parcial, ou seja, quando mais de um táxon foi anotado por NT, porém, o táxon anotado para REF era um deles;

- táxon onde houve similaridade entre gêneros, ou seja, quando diferentes táxons foram anotados por NT, porém, o gênero anotado em REF era um deles e,

- táxons onde houve total dissimilaridade, as sequências foram realinhadas e a lista das 10 primeiras alinhadas, daqui em diante top_10, foi analisada em conjunto com os parâmetros ID/ALG/SC.

Nos casos onde não houve total similaridade entre os táxons, somente as reads coincidentes foram consideradas. Já para considerar um táxon não validado foram observados, além dos valores de ID/ALG/SC_REF em relação os de NT se a espécie estava abaixo da 5ª colocação da lista top_10.

3.2.5 Lista de espécies

Para avaliar se houve alta especificidade de cada um dos *primers* em diferenciar as espécies selecionadas, cada sequência única representativa foi alinhada com a ferramenta “MAFFT” e agrupada em árvore filogenética seguindo o melhor modelo proposto pelo método de Neighbor-Joining ajustado para 1000 réplicas.

A acurácia de cada *primer* foi avaliada através de tabela de contingência baseada em dados de presença/ausência onde, os resultados obtidos foram confrontados com a lista de espécies conhecida da comunidade simulada.

3.2.6 Lista de espécie do rio Aporema

Para gerar a lista de espécies do rio Aporema, todos os procedimentos já descritos foram empregados, ou seja, aplicação de filtro_100, filtro_classe e ID/ALG/SC/e-value. Os táxons candidatos foram contrastados com a lista de espécies de peixes local produzida pelo IEPA e, aquelas coincidentes a nível de gênero foram consideradas válidas.

4 RESULTADOS

4.1 QUANTIDADE E QUALIDADE DAS SEQUÊNCIAS

A quantidade de sequências geradas pelas plataformas de sequenciamento foi extremamente variável para todos os *primers* utilizados. Considerando que os parâmetros hipotéticos para avaliar o sucesso de uma reação de sequenciamento seriam, neste caso: i) que o padrão “ótimo” de uma reação bem sucedida para a plataforma GS FLX+ 454 produziria cerca de 2 milhões de sequências e, a PGM ION 5 milhões; ii) que as amostras foram equivalentemente distribuídas, e iii) que apenas duas amostras foram incluídas por reação de sequenciamento, as amostras representativas da comunidade simulada obtiveram um sucesso de, 0,14%, 2,72%, 52,75% e 64,37% do produto final esperado para 12S, 16S, COA e COB, respectivamente. Já, as amostras representativas do rio Aporema obtiveram um sucesso de 0,56% e 20,91% para 16S e 12S, respectivamente (Figuras 7 e 8).

A qualidade das sequências geradas pelas plataformas de sequenciamento foi considerada ruim para 16S, extremamente variável para COA e, uniforme para 12S e COB. Considerando que o parâmetro hipotético para avaliar a qualidade das sequências, neste caso, seria de que todas correspondessem ao mesmo tamanho de extensão do *primer* utilizado, para as amostras representativas da comunidade simulada, 67,27%, 0,15%, 18,96% e 57,53%, respectivamente, 12S, 16S, COA e COB, corresponderam ao padrão esperado. Já, as amostras representativas do rio Aporema tanto 12S como 16S obtiveram 0,01% de sequências dentro do padrão desejado (Figuras 7 e 8).

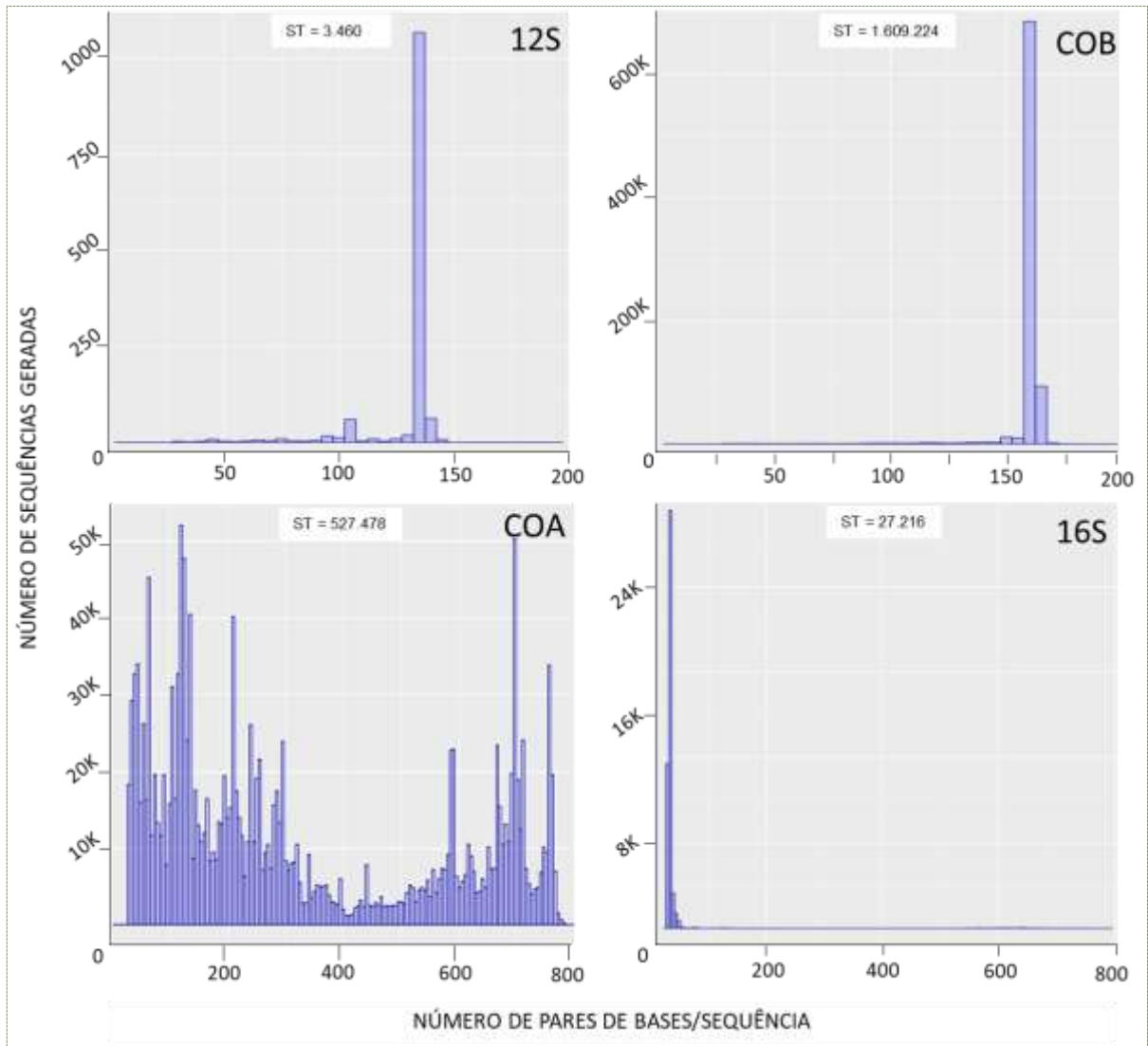


Figura 7 – Histograma de distribuição de sequências totais geradas pelas plataformas de sequenciamento para as amostras da comunidade simulada considerando cada um dos *primers* utilizados.

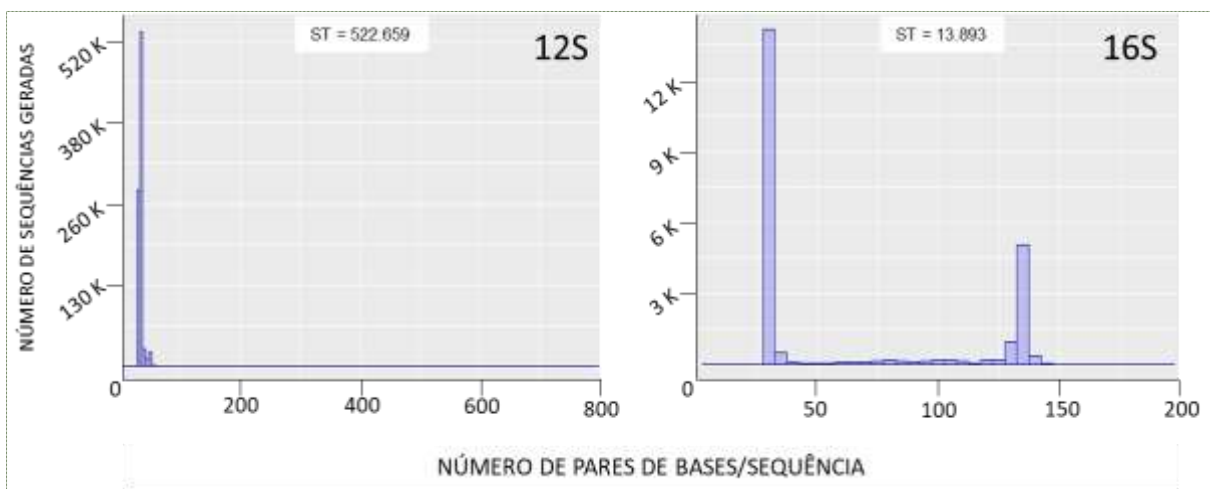


Figura 8 – Histograma de distribuição de sequências totais geradas pelas plataformas de sequenciamento para as amostras do rio Aporema considerando cada um dos *primers* utilizados.

4.2 IDENTIFICAÇÃO DAS ESPÉCIES DA COMUNIDADE SIMULADA

4.2.1 Distribuição das Sequências e Seleção das espécies candidatas

4.2.1.1 Primer 12S

Como resultado do alinhamento do banco NT, 84,3% das sequências obtidas foram similares a de algum tipo de organismo já descrito. Dentre os organismos classificados, 0,3% corresponderam a Bacteria e, 97,7% a Eukaryota sendo que dentre estes, 16% foram classificados como Actinopterygii. Quando aplicado o filtro_100, foram eliminadas 10% das sequências restando, 100% de Eukaryota sendo que, dentre estes 16% classificados como Actinopterygii. Quando aplicado o filtro_classe, foram eliminadas 83,5% das sequências e a porcentagem final considerada nas análises seguintes correspondeu a 14% do dado bruto (Figura 9). Em relação ao banco REF, houve alinhamento de 75,5% das sequências, após aplicado o filtro_100, foram eliminadas 5% das mesmas e a porcentagem foi reduzida para 71,7%. Quando aplicado o filtro_classe, foram eliminadas 84,5% das sequências e a porcentagem final considerada nas análises seguintes correspondeu a 15,7% do dado bruto (Figura 9).

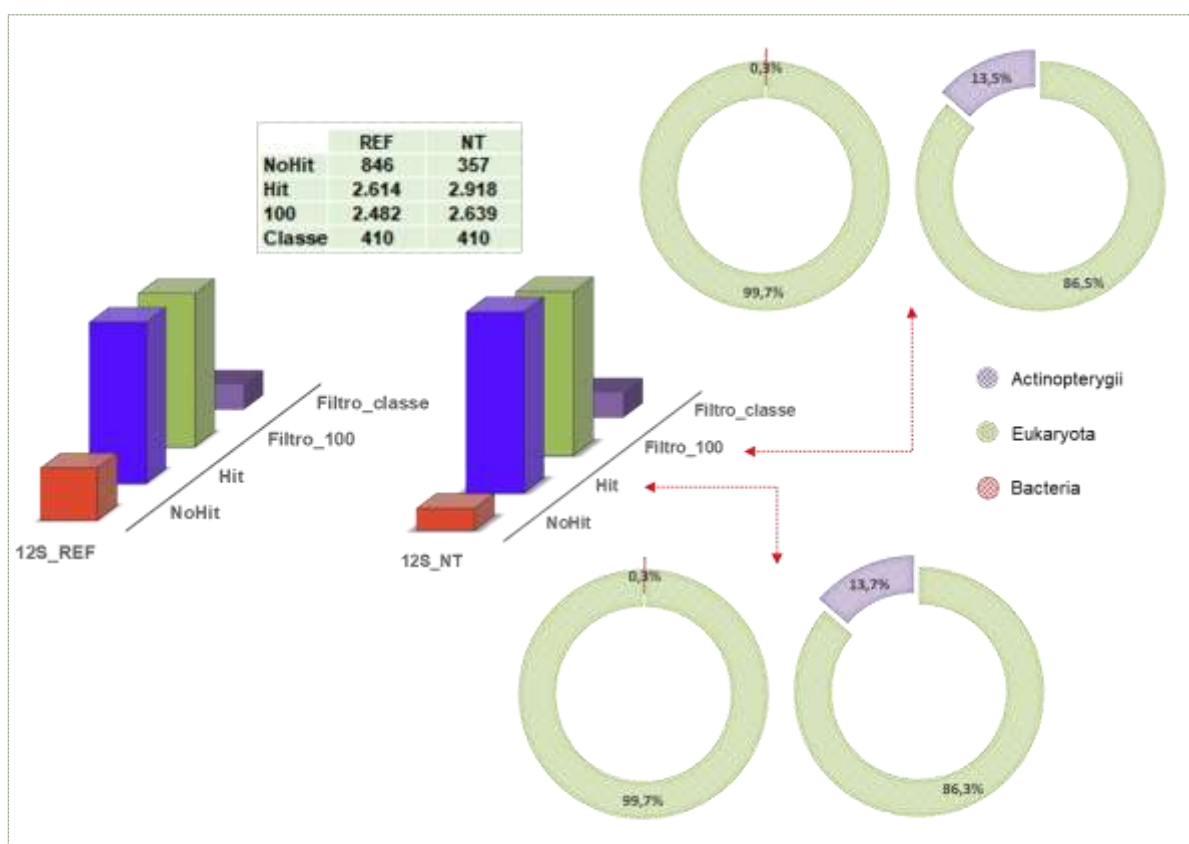


Figura 9 – Classificação das sequências geradas após alinhamento. Em colunas, porcentagem de sequências não classificadas (NoHit) e classificadas (Hit, Filtro_100, Filtro_classe). Em círculos, porcentagem de sequências classificadas por grupos de organismos anterior (Hit) e, posterior ao filtro_100). Tabela ao centro, valores brutos das sequências analisadas.

A verificação do resíduo gerado pelo filtro_classe indicou que as sequências eliminadas em ambos os bancos não seriam informativas para as análises seguintes, já que, além das espécies divergirem quanto a sua classificação taxonômica, os valores de ID_REF foram consideravelmente inferiores aos de ID_NT. Portanto, as espécies identificadas por REF como *Apterotonotus albifrons* (n=4; *Bos taurus*. $M_eID_NT=94,2\%$; $M_eSC_NT=201$), *Apistogramma agassizii* (n=2; *Tragelaphus eurycerus*. $M_eID_NT=93,4\%$; $M_eSC_NT=202$), *Astyanax bimaculatus* (n=1; *Bos taurus*. $ID_NT=97\%$; $SC_NT=214$), *Serrasalmus marginatus* (n=1; *Bos taurus* $ID_NT=93\%$; $M_eSC_NT=193$), *Corydoras ephippfer* (n=1982; *Bos taurus* $M_eID_NT=98\%$; $M_eSC_NT=230$) e, *Panaque armbrusteri* (n=1; *Cephalophus callipygus* $ID_NT=88,2\%$; $SC_NT=152$) foram desconsideradas (Figura 10).

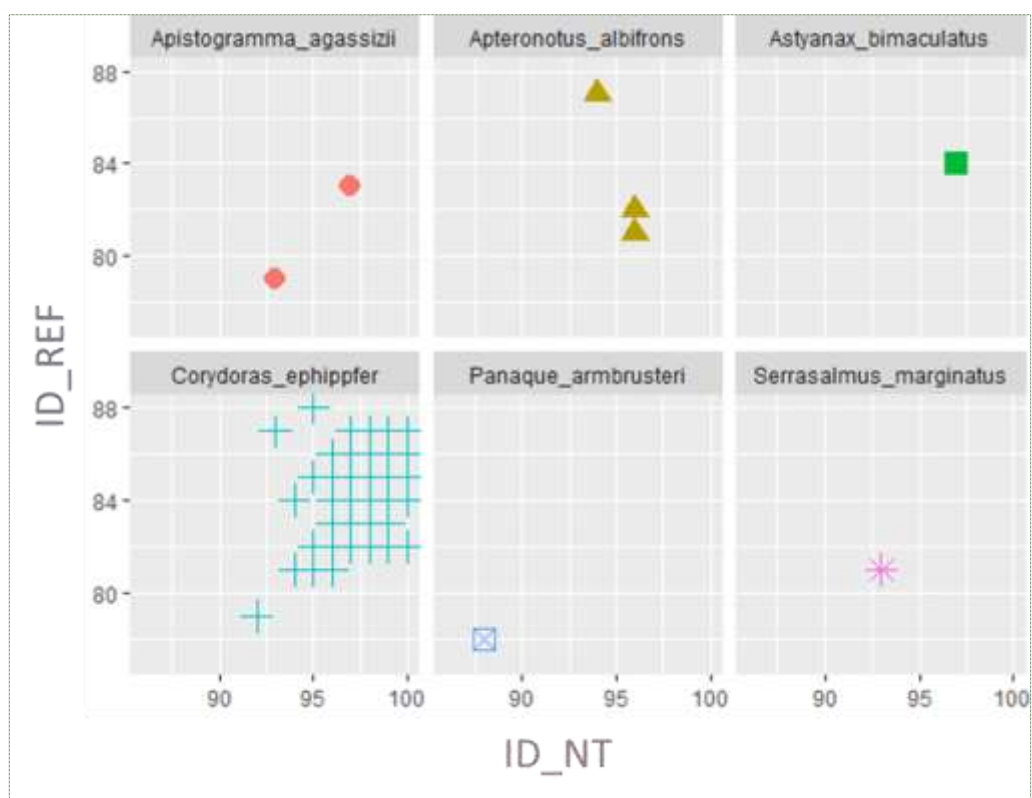


Figura 10 – Distribuição dos valores de identidade das sequências removidas após a aplicação do filtro_100.

4.2.1.2 Primer 16S

Como resultado do alinhamento do banco NT, 69,9% das sequências obtidas foram similares a de algum tipo de organismo já descrito. Dentre os organismos classificados, 7,2% corresponderam a Bacteria e, 92,8% a Eukaryota sendo que, dentre estes, 3,1% foram classificados como Actinopterygii. Quando aplicado o filtro_100, foram eliminadas 99% das sequências restando, 100% de Eukaryota sendo que dentre estes, 83% classificados como Actinopterygii. Quando aplicado o filtro_classe foram eliminadas 3,8% das sequências e a porcentagem final considerada nas análises seguintes corresponderam a 0,009% do dado bruto

(Figura 12). Em relação ao banco REF, 59,4% das sequências foram alinhadas, após aplicado o filtro_100 foram eliminadas 99% das mesmas e a porcentagem foi reduzida para 1%. Quando aplicado o filtro_classe foram eliminadas 16,7% das sequências e a porcentagem final considerada nas análises seguintes correspondeu a 0,008% do dado bruto (Figura 11).

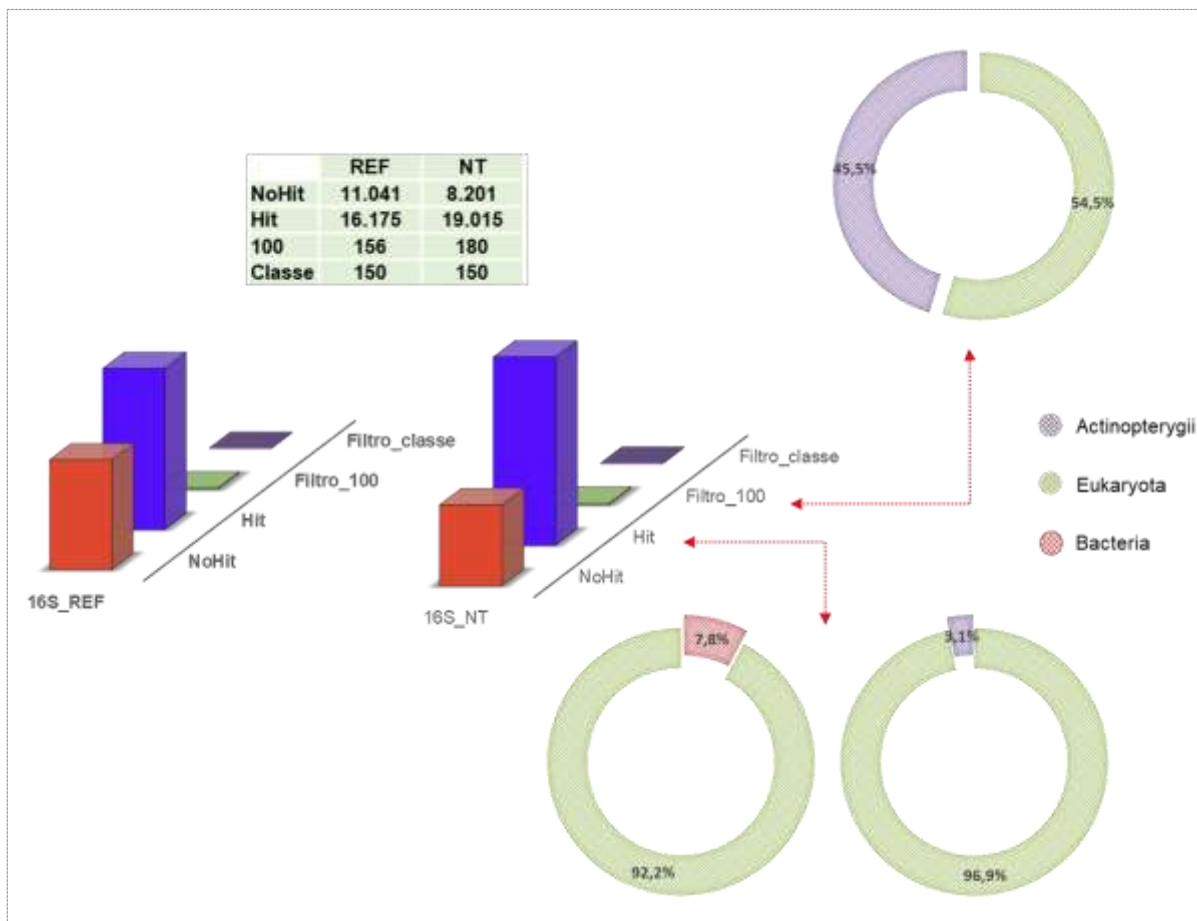


Figura 11 – Classificação das sequências geradas após alinhamento. Em colunas, porcentagem de sequências não classificadas (NoHit) e classificadas (Hit, Filtro_100, Filtro_classe). Em círculos, porcentagem de sequências classificadas por grupos de organismos anterior (Hit) e, posterior ao filtro_100. Tabela ao centro, valores brutos das sequências analisadas.

A verificação do resíduo gerado pelo filtro_classe indicou que as sequências eliminadas em ambos os bancos não seriam informativas para as análises seguintes e, portanto, *Hypostomus faveolus*, *Mesonauta acora* (n=1; *Bos taurus*; ID_NT=90,8%; SC_NT=734, ID_NT=99,7%; SC_NT=499, respectivamente) e, *Myloplus rubripinnis* (n=1; *Lissotriton vulgaris*; ID_NT=94,7%; SC_NT=213) mesmo apresentando ID_REF=100%, foram desconsideradas (Figura 12).

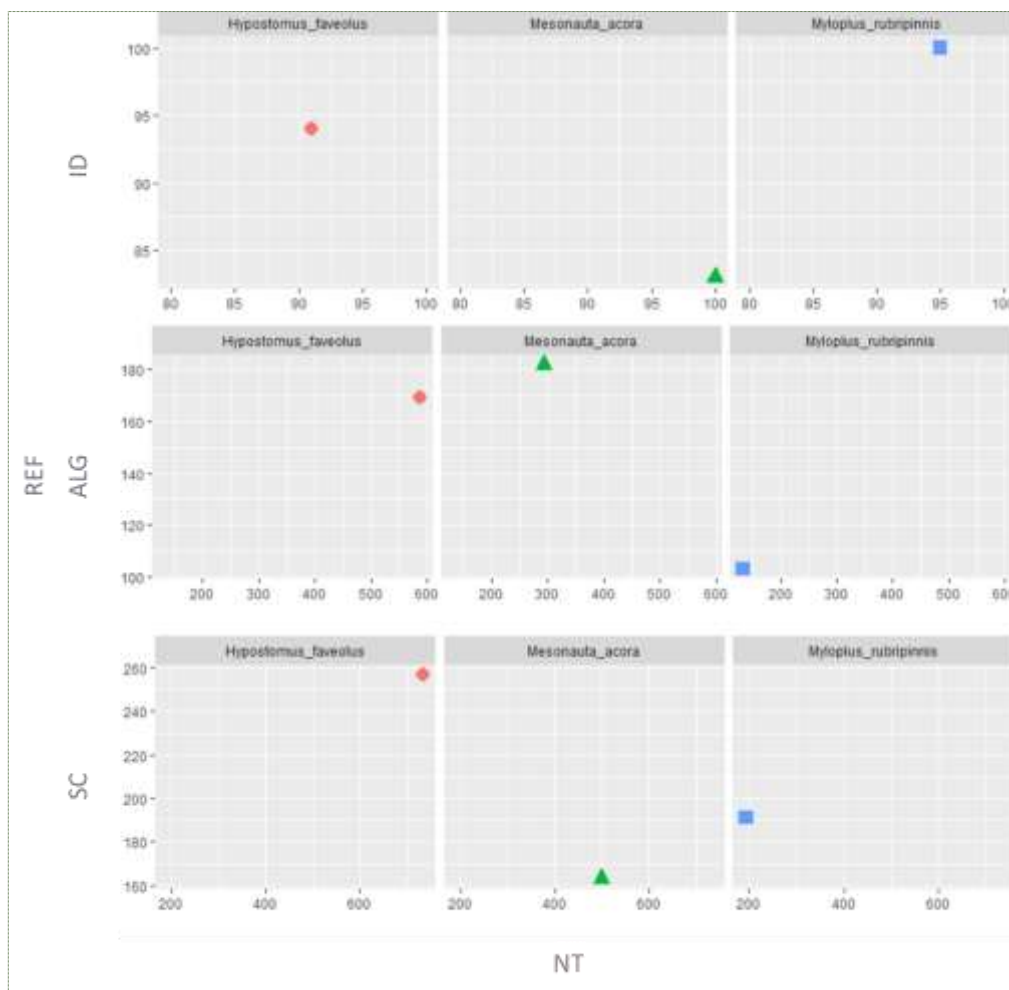


Figura 12 – Distribuição das seqüências residuais após a aplicação do filtro_classe.

4.2.1.3 Primer COA

Como resultado do alinhamento do banco NT, 81,9% das seqüências foram similares a de algum tipo de organismo já descrito. Dentre os organismos classificados, 80,9% corresponderam a Bacteria e 19,1% a Eukaryota, sendo que, dentre estes, 21,1% foram classificados como Actinopterygii. Quando aplicado o filtro_100, foram eliminadas 15% das seqüências, restando 88,1% de Bacteria e 12% de Eukaryota sendo que, dentre estes, 17% classificados como Actinopterygii. Quando aplicado o filtro_classe, foram eliminadas 98% das seqüências e a porcentagem final considerada nas análises seguintes correspondeu a 1,7% do dado bruto (Figura 14). Em relação ao banco REF, 16,4% das seqüências foram alinhadas e, após aplicado o filtro_100, foram eliminadas 72,2% das mesmas e a porcentagem foi reduzida para 4,9%. Quando aplicado o filtro_classe, foram eliminadas 72% das seqüências e a porcentagem final considerada nas análises seguintes correspondeu a 8,3% do dado bruto (Figura 13).

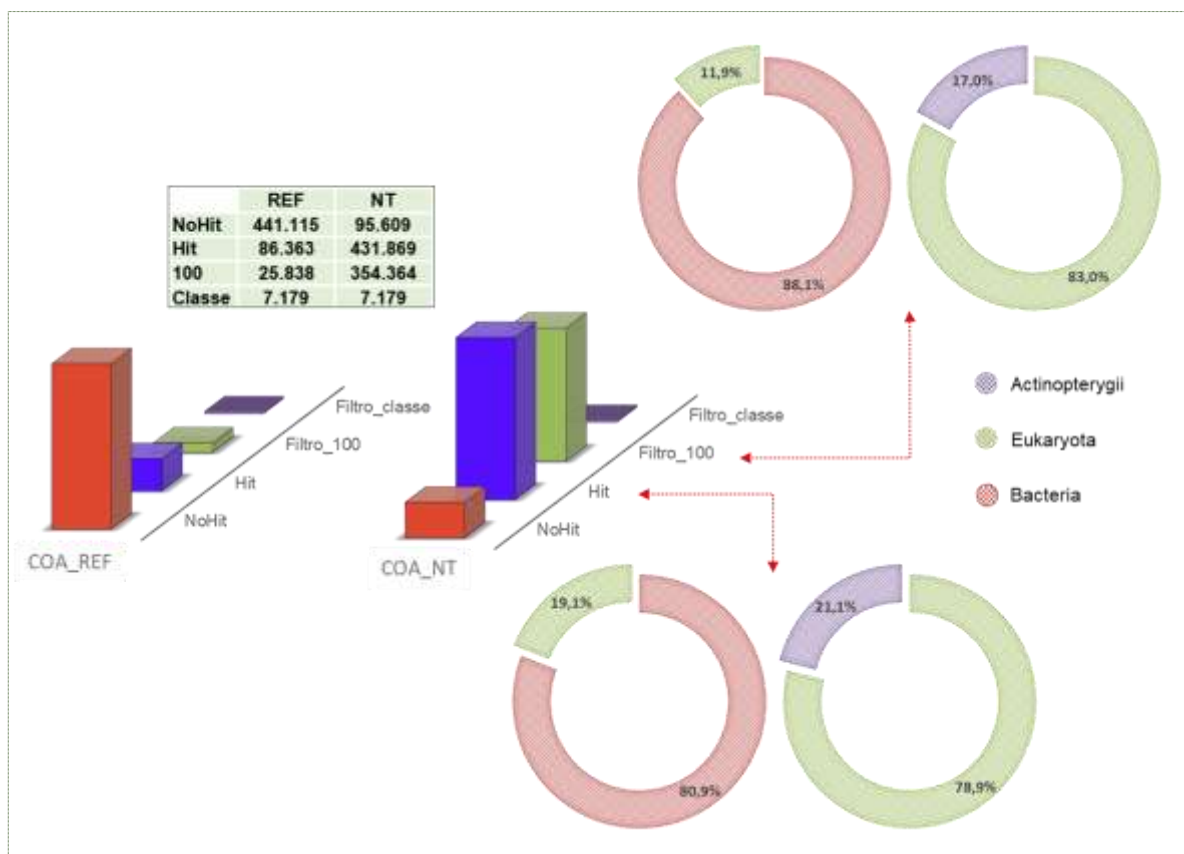


Figura 13 – Classificação das sequências geradas após alinhamento. Em colunas, porcentagem de sequências não classificadas (NoHit) e classificadas (Hit, Filtro_100, Filtro_classe). Em círculos, porcentagem de sequências classificadas por grupos de organismos anterior (Hit) e, posterior ao filtro_100. Tabela ao centro, valores brutos das sequências analisadas.

A análise do resíduo das sequências removidas pelo filtro_classe aplicado aos bancos indicou que 866 sequências foram eliminadas, sendo que destas, 750 não seriam informativas já que, além das espécies divergirem quanto a sua classificação taxonômica, a média de ID_REF foi de 78% enquanto de NT 89,1% e os respectivos valores SC foram de 233 para REF e 454 para NT.

Dentre as 116 sequências onde ID_REF foi superior a 90%, seis potenciais espécies candidatas foram avaliadas.

Monocirrhus polyacanthus (n=2; *Callipepla squamata*; $M_eID_NT=85$; $M_eSC_NT=194$; $M_eALG_NT=210$), *Apteronotus albifrons* (n=3; *Acidovorax* sp; $M_eID_NT=84,6\%$; $M_eSC_NT=485$), *Apistogramma gephyra* (7; *Novosphingobium* sp; $M_eID_NT=83\%$; $M_eSC_NT=400$) e, *Guianacara dacrya* (n=47; *Canis lupus familiaris*; $M_eID_NT=84,5\%$; $M_eSC_NT=134$) foram eliminadas por se tratar de diagnóstico inconclusivo (Figura 14).

As sequências referentes a *Apistogramma rupununi* (n=9; diversos táxons, $M_eID_NT=81,25\%$; $M_eSC_NT=138$) e, *Apistogramma agassizii* (n=27; diversos táxons,

$M_eID_NT=84,7\%$; $M_eSC_NT=134$) foram eliminadas, porém, consideradas como potenciais resultados falso-negativo (Figura 14).

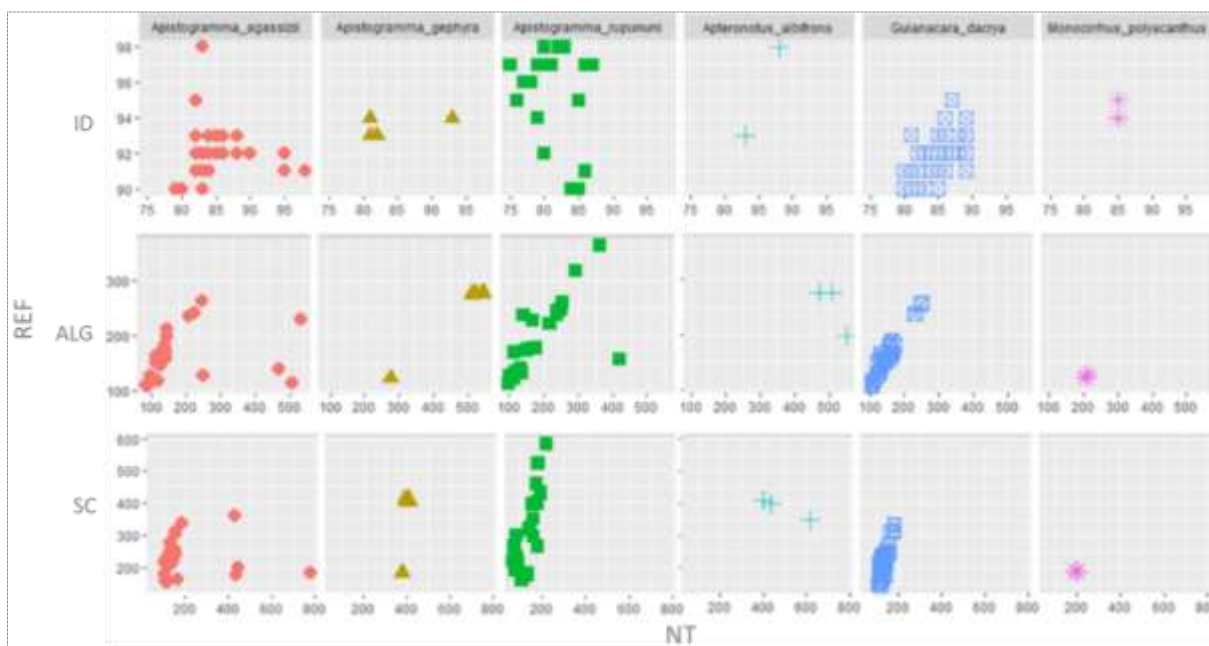


Figura 14 – Distribuição das sequências residuais após a aplicação do filtro_classe.

4.2.1.4 *Primer COB*

Como resultado do alinhamento do banco NT, 65,2% das sequências foram similares a de algum tipo de organismo já descrito. Dentre os organismos classificados, 100% corresponderam a Eukaryota, sendo que, dentre estes, 93,6% foram classificados como Actinopterygii. Quando aplicado o filtro_100, foram eliminadas 3,5% das sequências, restando 100% de Eukaryota, sendo que, dentre estes, 95,6% classificados como Actinopterygii. Quando aplicado o filtro_classe, foram eliminadas 4,4% das sequências e a porcentagem final considerada nas análises seguintes correspondeu a 60,1% do dado bruto (Figura 15). Em relação ao banco REF, 63,9% das sequências foram alinhadas, após aplicado o filtro_100, foram eliminadas 1% das mesmas e a porcentagem foi reduzida para 63,4%. Quando aplicado o filtro_classe, foram eliminadas 5,9% das sequências e a porcentagem final considerada nas análises seguintes correspondeu a 63,4% do dado bruto (Figura 15).

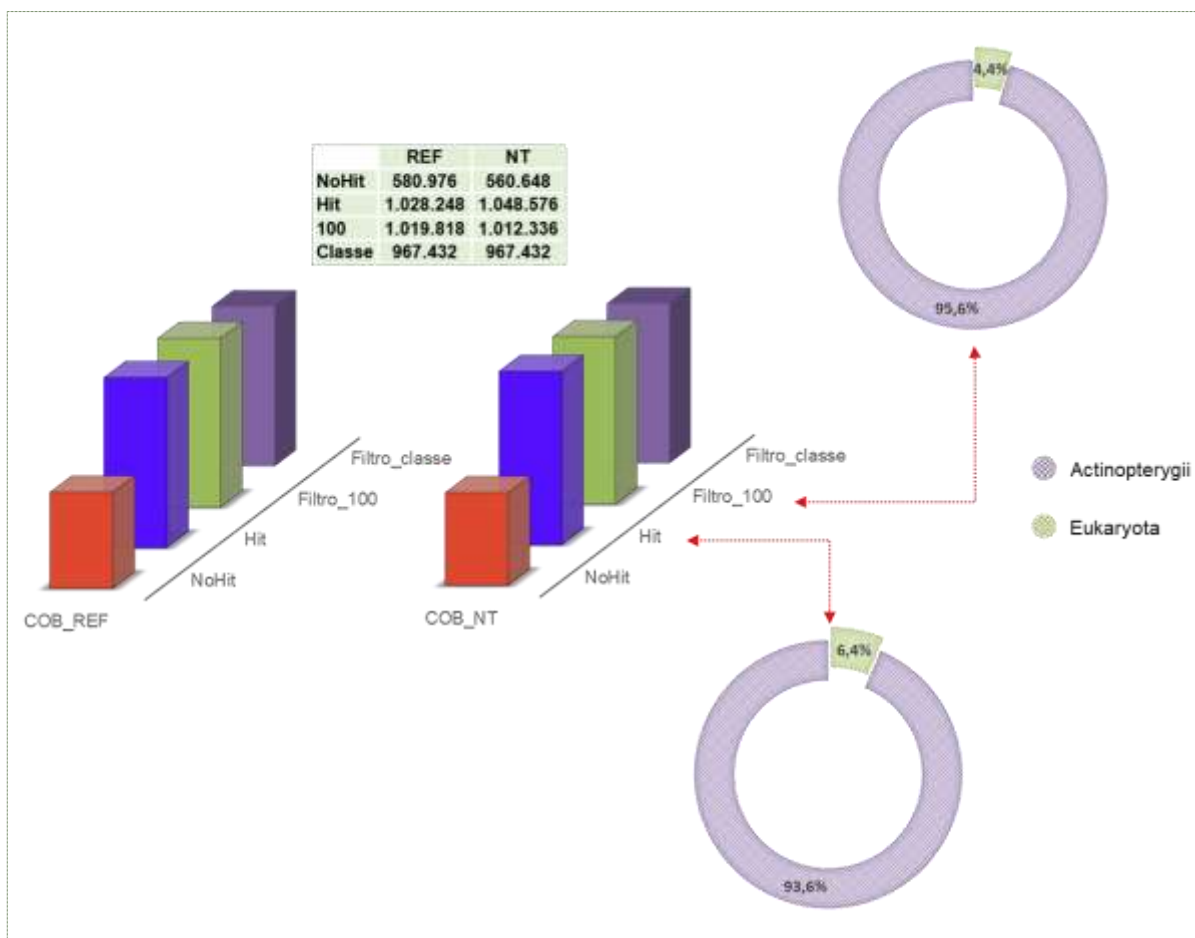


Figura 15 – Classificação das sequências geradas após alinhamento. Em colunas, porcentagem de sequências não classificadas (NoHit) e classificadas (Hit, Filtro_100, Filtro_classe). Em círculos, porcentagem de sequências classificadas por grupos de organismos anterior (Hit) e, posterior ao filtro de remoção de 100bp (Remoção_100). Tabela ao centro, valores brutos das sequências analisadas.

A análise do resíduo das sequências removidas pelo filtro_classe aplicado aos bancos indicou que 9.445 sequências foram eliminadas sendo que destas, 6.944 não seriam informativas já que, além de as espécies divergirem quanto a sua classificação taxonômica, a média dos valores de ID/SC_REF foram 87,2% e 184, enquanto os de NT 90,2% e 199, respectivamente, e foram eliminadas das análises.

Dentre as 2.503 sequências restantes onde ID_REF foram superiores a 90%, cinco potenciais espécies candidatas estavam incluídas. As sequências classificadas como *Apteronotus albifrons* e, *Peckoltia vittata* ($n=2$; *Urolophus halleri*; ID_NT=91,5%; SC_NT=210) foram eliminadas por se tratar de diagnóstico inconclusivo (Figura 16).

As sequências classificadas como *Farlowella henriquei* ($n=773$, *Pteromys volans*; M_e ID_NT=88,83%; SC_NT =188), *Panaque armbruster* ($n=1$, *Chelodina expansa*; ID_REF=92%; SC_NT=166) e, *Serrapinnus kriegi* ($n=1726$; diversos, ID_REF=86,2%, SC_REF = 162), foram consideradas como resultados falso-negativo (Figura 16).

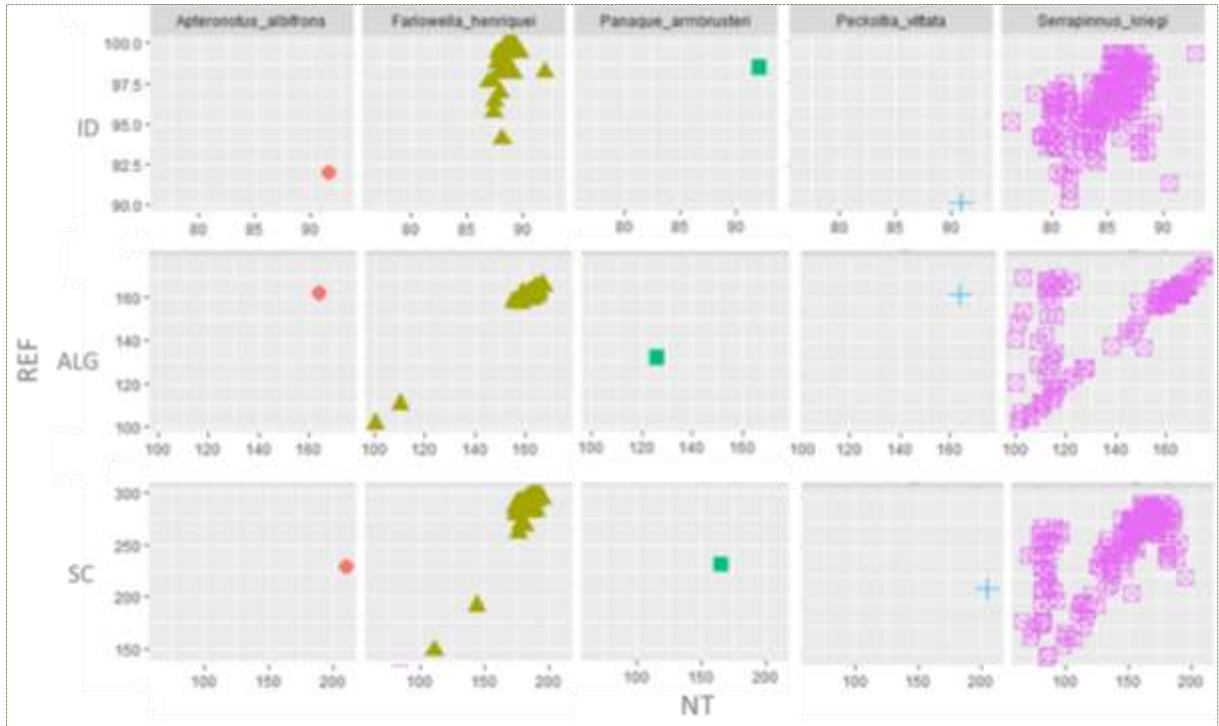


Figura 16 – Distribuição das seqüências residuais após a aplicação do filtro_classe.

4.2.2. Validação das espécies candidatas

4.2.2.1 Primer 12S

Como resultado do filtro_classe, 410 seqüências foram selecionadas como candidatas, sendo que 14 táxons foram definidos pelo banco REF.

Quatro táxons coincidiram todas as suas seqüências entre REF e NT, *Aptereronotus albifrons* (n=53(84); $M_eID_REF=99\%$; e-value=9.06e-68), *Corydoras ephippifer* (n=1(8); $M_eID_REF=97,7\%$; e-value=5.05e-65), *Corydoras julii* (n=1(2); ID_REF=100%; e-value=8.39e-68), *Gnathocharax steindachneri* (n=1(20); ID_REF=100%; e-value=3.26e-47), e foram considerados para a análise seguinte (Figura 17).

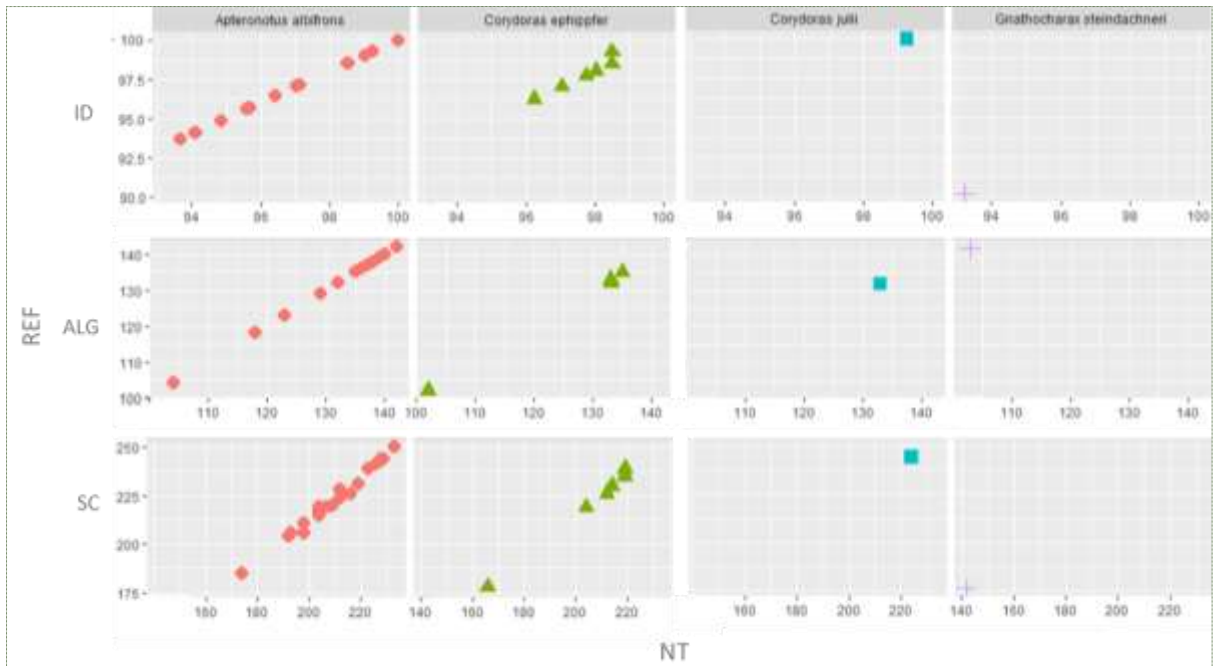


Figura 17 – Distribuição das seqüências selecionadas após a aplicação do filtro_classe.

Laetacara dorsigera (n=1(29); ID_REF=97%; e-value=1.64e-48) divergiu entre os bancos, porém, o táxon constava entre um dos anotados por NT, portanto, foi considerado nas análises seguintes (Figura 18).

Apistogramma rupununi(n=1(215); ID=100%; e-value=7.69e-68 x *Oreochromis sp*, ID_NT=94%), *Apistogramma agassizii* (n=9(62); ID_REF=97%; e-value=9.68e-57 x *Pterophyllum scalare*, ID_NT=96%), *Apistogramma gephyra* (n=1(4); ID_REF=98%; e-value= 8.43e-58 x *Etroplus maculatus*, ID_NT=94%), divergiram entre os bancos, mas, o gênero contava entre um deles e, ID_REF foi maior do que em NT, sendo assim considerados nas análises seguintes (Figura 18).

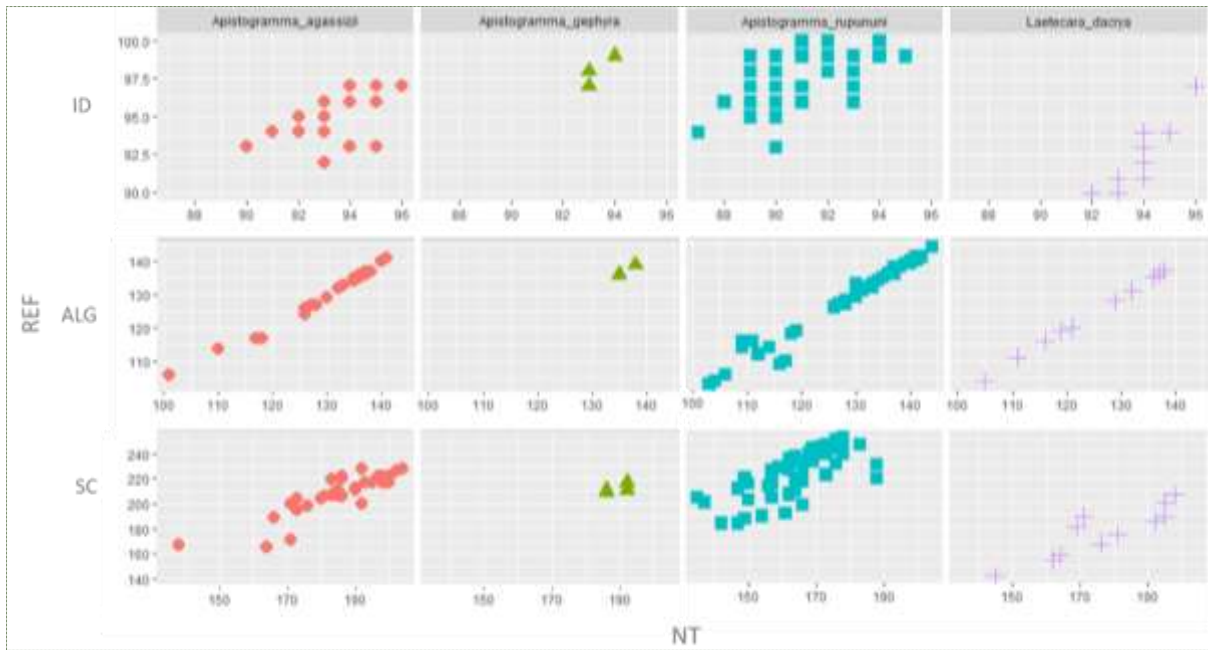


Figura 18 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.

Heros efasciatus (n=2; ID_REF=98,5% x *Cryptoheros cutteri*, ID_NT=96%), *Hemigrammus rhodostomus* (n=1; ID_REF=98% x *Paracheiroduon innesi*, ID_NT=97,7%) e, *Heterocharax virgulatus* (n=13; ID_REF=100%; x *Tytocharax tambopatensis*, ID_NT=94%) divergiram entre os bancos, no entanto, os táxons contrastantes são filogeneticamente próximos e, pelo que consta não existem sequências nucleotídicas para as espécies publicadas no Genbank como referência para o gene 12S, desta forma, foram considerados nas análises seguintes (Figura 19).

Myloplus rubripinnis (n=1; ID_REF=99%; e-value=5.18e-60 – SeqReq *Metynnis hypsauchen*; ID_NT=98%), foi considerado nas análises seguintes, pois, após verificação top_10, o táxon constava na lista (Figura 19).

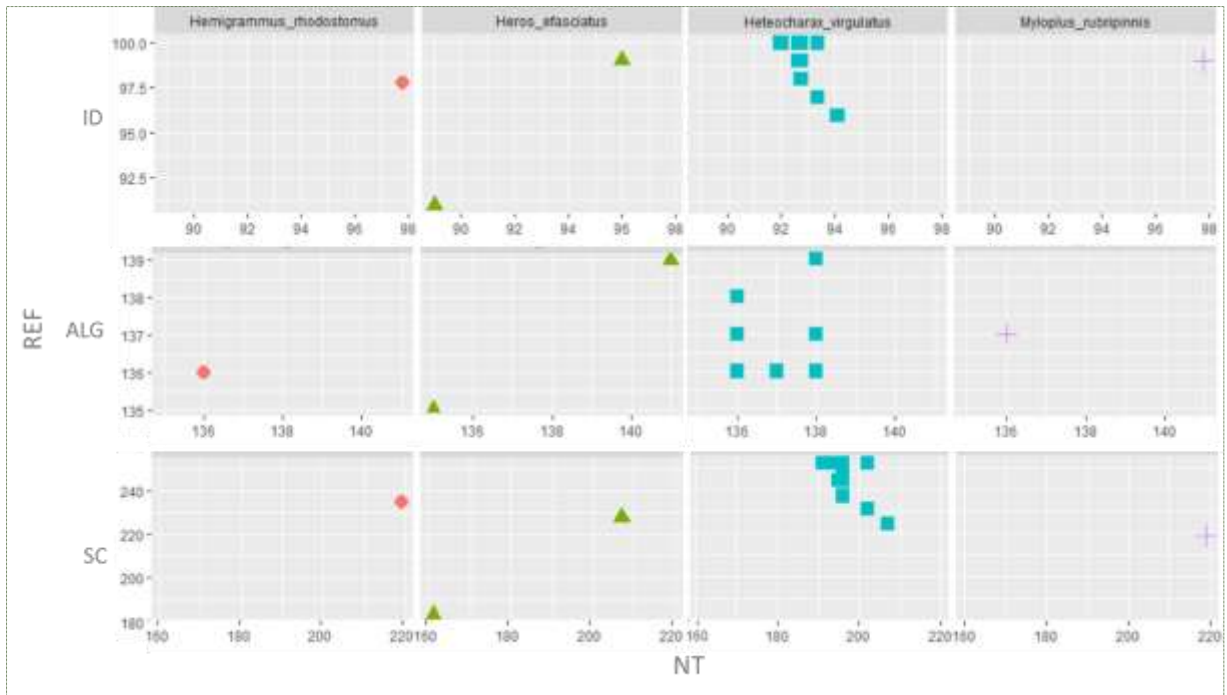


Figura 19 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.

Peckoltia vittata (n=1; ID_REF=93%; e-value=5.18e-60 x *Rineloricaria lanceolata*, ID_NT=100%) foi removida da lista de candidatas, pois seus valores de referência foram discrepantes em relação à contraparte e não constou entre as top_10 (Figura 20).

Pygocentrus natterei (n=3; ID_REF=100%; e-value=4.81e-70 x *Serrasalmus marginatus*, ID_NT=100%) foi removida da lista de candidatas, pois suas sequências foram similares em ID/ALG/SC às de *Serrasalmus marginatus*, espécie também presente na comunidade simulada indicando que possa existir problemas quando a resolução do *primer* em identificar a subfamília Serrasalminea (Figura 20).

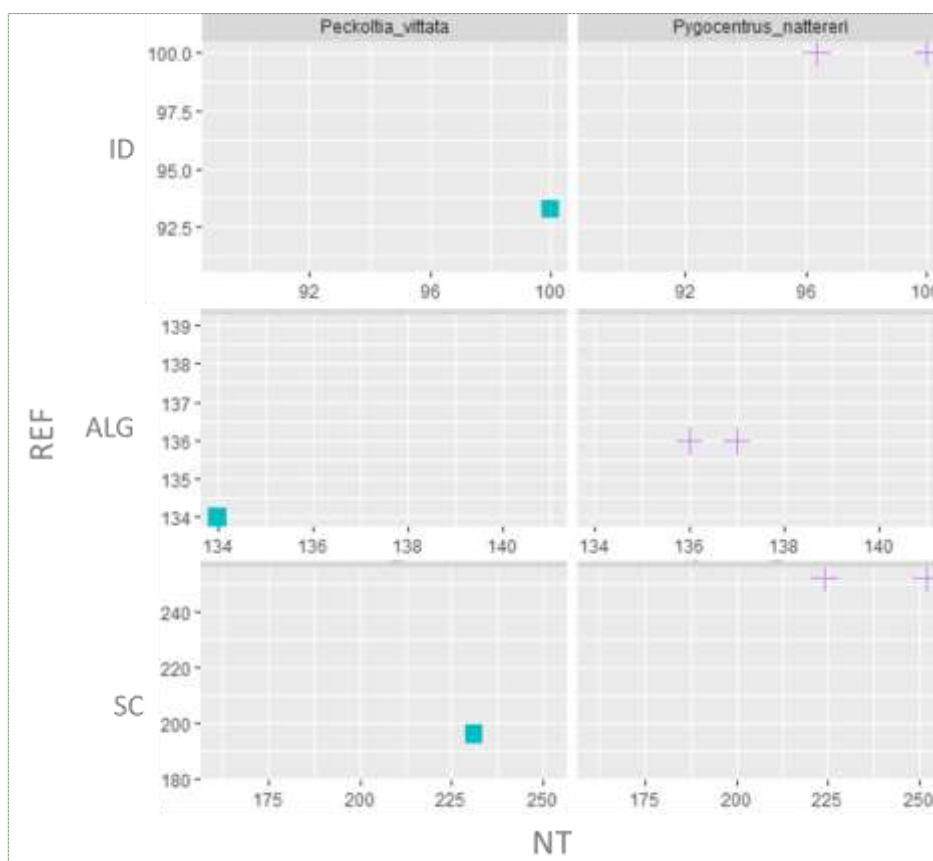


Figura 20 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.

4.2.2.2 Primer 16S

Como resultado do filtro_classe, 156 sequências foram selecionadas como candidatas, sendo que 13 táxons foram definidos pelo banco REF.

Cinco táxons coincidiram todas suas sequências entre REF e NT, *Apistogramma agassizii* (n=1(18); ID_REF=95%; e-value=9.23e-142), *Apteronotus albifrons* (n=2(9); ID_REF=95%; e-value=1.73e-11), *Guianacara dacrya* (n=2(9); ID_REF=94%; e-value=7.06e-178), *Peckoltia vittata* (n=1; ID_REF=95%; e-value=1.18e-61) e, *Panaque armbrusteri* (n=1(9); ID_REF=100%; e-value=7.33e-154) sendo considerados nas análises seguintes (Figura 21).

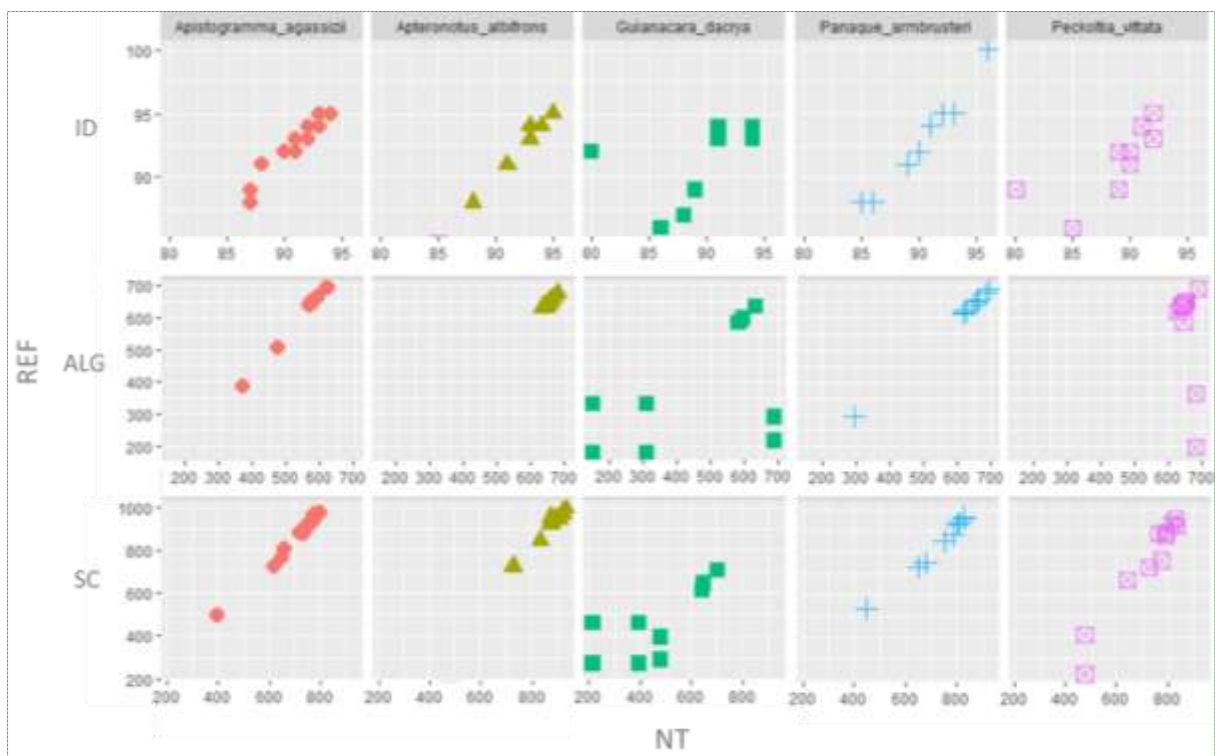


Figura 21 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.

Apistogramma gephyra (n=1(5); ID_REF=96%; e-value=3.59e-112), *Apistogramma rupununi* (n=2(17); ID_REF=94%; e-value=5.92e-164) e, *Heros efasciatus* (n=2(13); ID_REF= 98%; e-value=1.15e-170), divergiram entre os bancos, porém, os táxons constavam entre os anotados por NT e foram considerados nas análises seguintes (Figura 22).

Serrasalmus marginatus (n=1(4); ID_REF= 96%; e-value=1,01-e1; SeqReq – *Pygcentrus natereri*; ID_NT=95%) e, *Heterocharax virgulatus* (n=1(40); ID_REF=96%; e-value=6.19e-129; SeqREQ – *Hoplocharax goethei*; ID_NT=89%), foram considerados nas análises seguintes pois, após avaliação top_10 o táxon constava na lista (Figurar 22).

Satanoperca jurupari (n=2; ID_REF=88%; e-value=4.66e-155 – SeqReq *Apistogrammoides pucallpaensis*; ID_NT=89%), após verificação top_10, constava na lista, porém, devido aos baixos valores de ID/ALG, foi considerado indefinido (Figura 22).

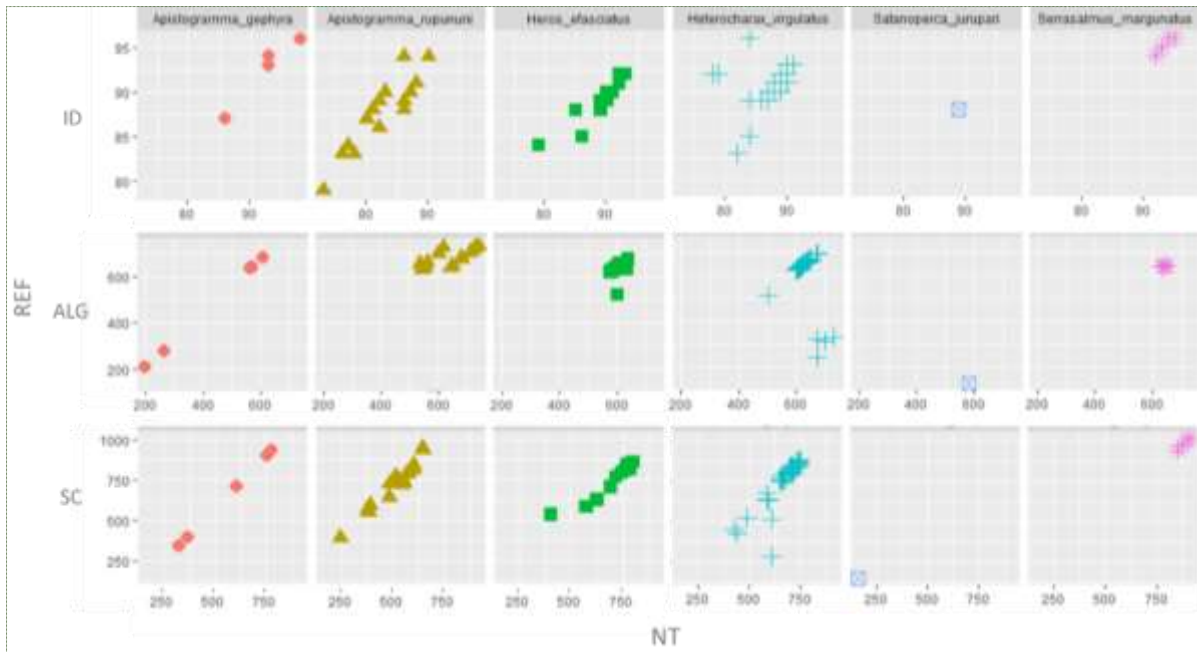


Figura 22 – Distribuição das seqüências selecionadas após a aplicação do filtro_classe.

Hypostomus faveolus (n=1(8); ID_REF=88%; e-value=4.75e-175) divergiu entre os bancos e, na avaliação top_10, suas seqüências não foram listadas e, portanto, foi desconsiderado da lista de espécies (Figura 23).

Pygocentrus nattereri (n=11; ID_REF=100%; e-value=6.06e-95 x *Hoplocharax goethei*; ID_NT=94%), apenas uma seqüência obteve ID desejável, portanto, com o tamanho fora do padrão e suas seqüências maiores identificadas como *Serralmus marginatus* foi desconsiderada (Figura 23).

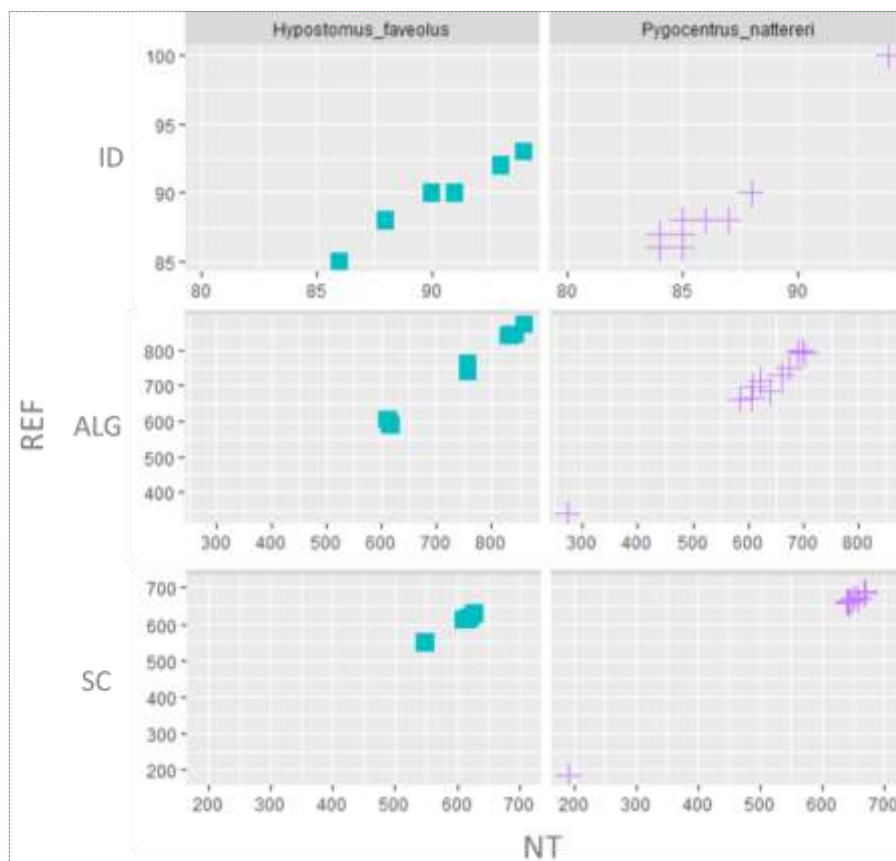


Figura 23 – Distribuição das seqüências residuais após a aplicação do filtro_classe.

4.2.2.3 Primer COA

Como resultado do filtro_classe, 7.179 seqüências foram selecionadas como candidatas, sendo que 25 táxons foram definidos pelo banco REF.

Cinco táxons coincidiram integralmente suas seqüências entre os bancos REF e NT, *Acaronia nassa* (n=2; ID_REF=99; e-value=1.73e-114), *Apteronotus albifrons* (n=11(3532); ID_REF=97,5; e-value=5.32e-61), *Farlowella henriquei* (n=1(9); ID_REF=97,5; e-value=8.54e-124), *Hyostomus faveolus* (n=1; ID_REF=96; e-value=6.33e-72), e, *Pseudanos trimaculatus* (n=1(4); ID_REF=100; e-value=7.39e-77) e foram considerados nas análises seguintes (Figura 24).

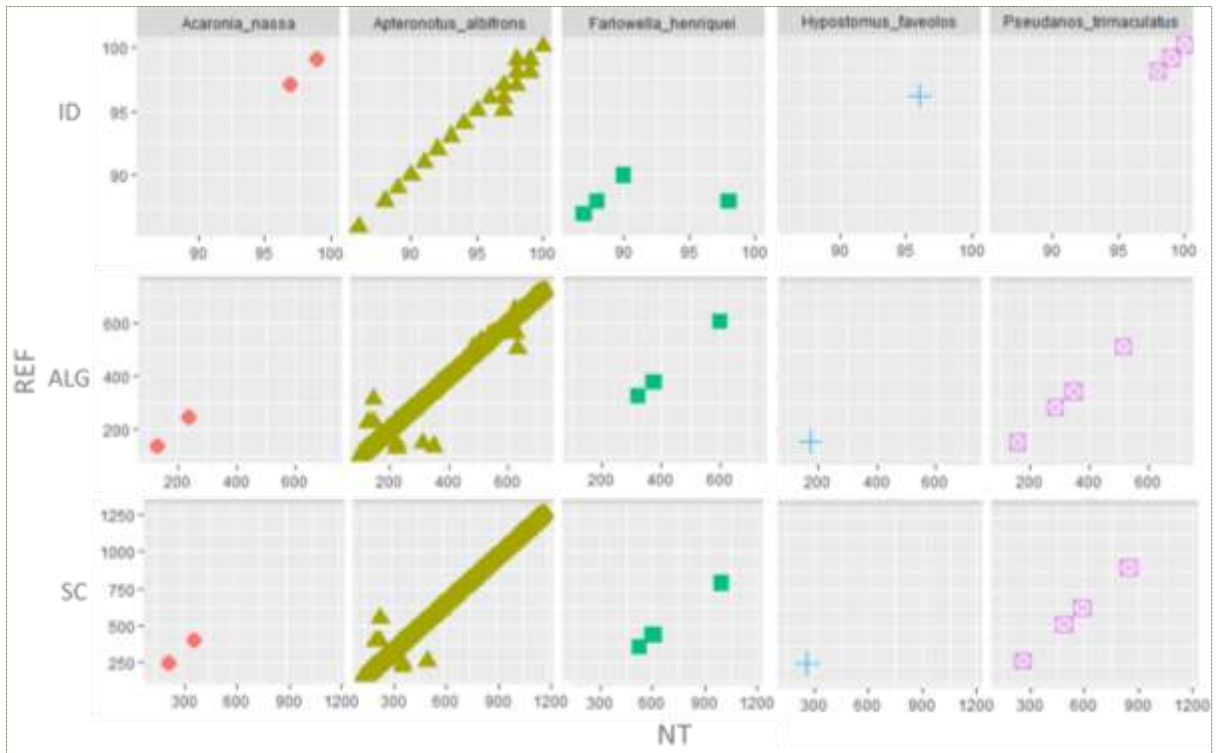


Figura 24 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.

Apistogramma agassizii (n=3(472); ID_REF=98%; e-value=6.18e-50, *Corydoras ephippifer* (n=13(35); ID_REF=98%; e-value=1.66e-144), *Pygocentrus nattereri* (n=3(209); ID_REF=100; e-value=2.4e-137), *Monocirrhus polyacanthus* (n=9(58); ID_REF=98; e-value=9.56e-76) divergiram entre os bancos, mas os táxons contavam entre os anotados por NT e foram considerados espécies válidas (Figura 25).

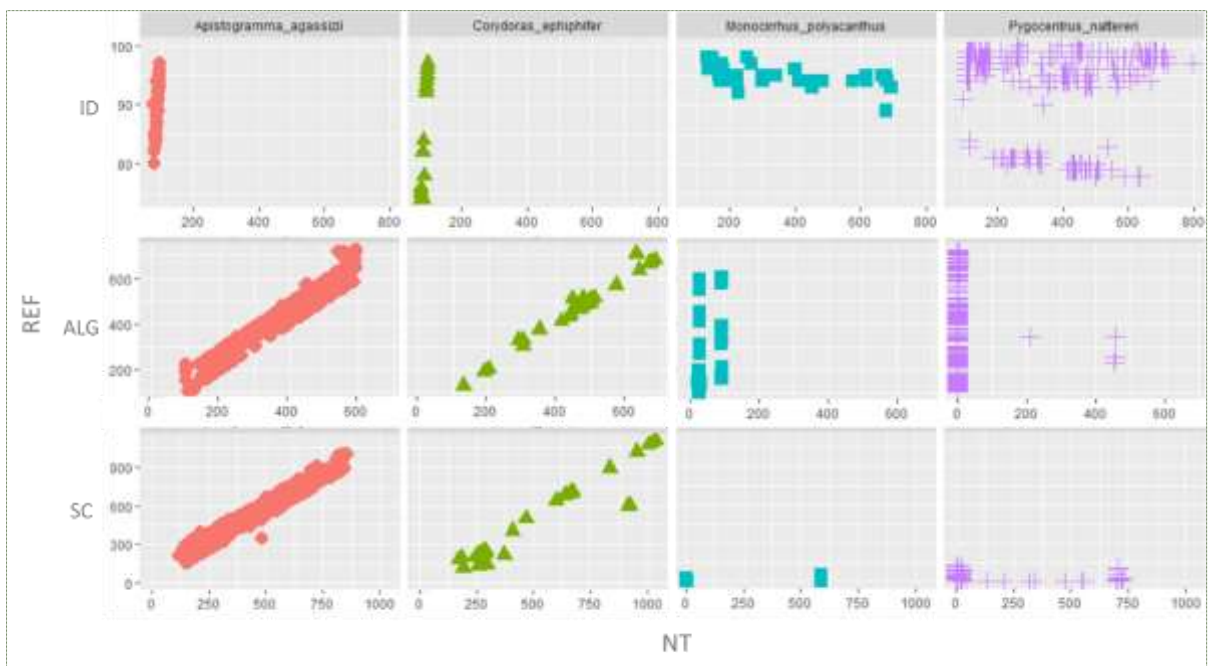


Figura 25 – Distribuição das sequências residuais após a aplicação do filtro_classe.

Heros efasciatus (n=1(549); ID_REF=99; evalue=1,02e-5 x *Heros severus*; ID=97,8), *Apistogramma rupununi* (n=1(236); ID=100%; e-value=2.05e-86 x *Glossogobius aureus*; ID_NT=87%), e, *Apistogramma gephyra* (n=1(388); ID_REF=97%; e-value=7.83e-81 x *Apistogramma agassizii*; ID_NT=97%), divergiram entre os bancos, porém, o gênero constava entre os anotados por NT. No entanto, os valores de ID_REF para as espécies *Apistogramma agassizii* e *Apistogramma gephyra* indicam que o possivelmente o *primer* COA tenha baixa resolução para este complexo. Desta forma, *Heros efasciatus* e o gênero *Apistogramma* foram considerados nas análises seguintes (Figuras 26).

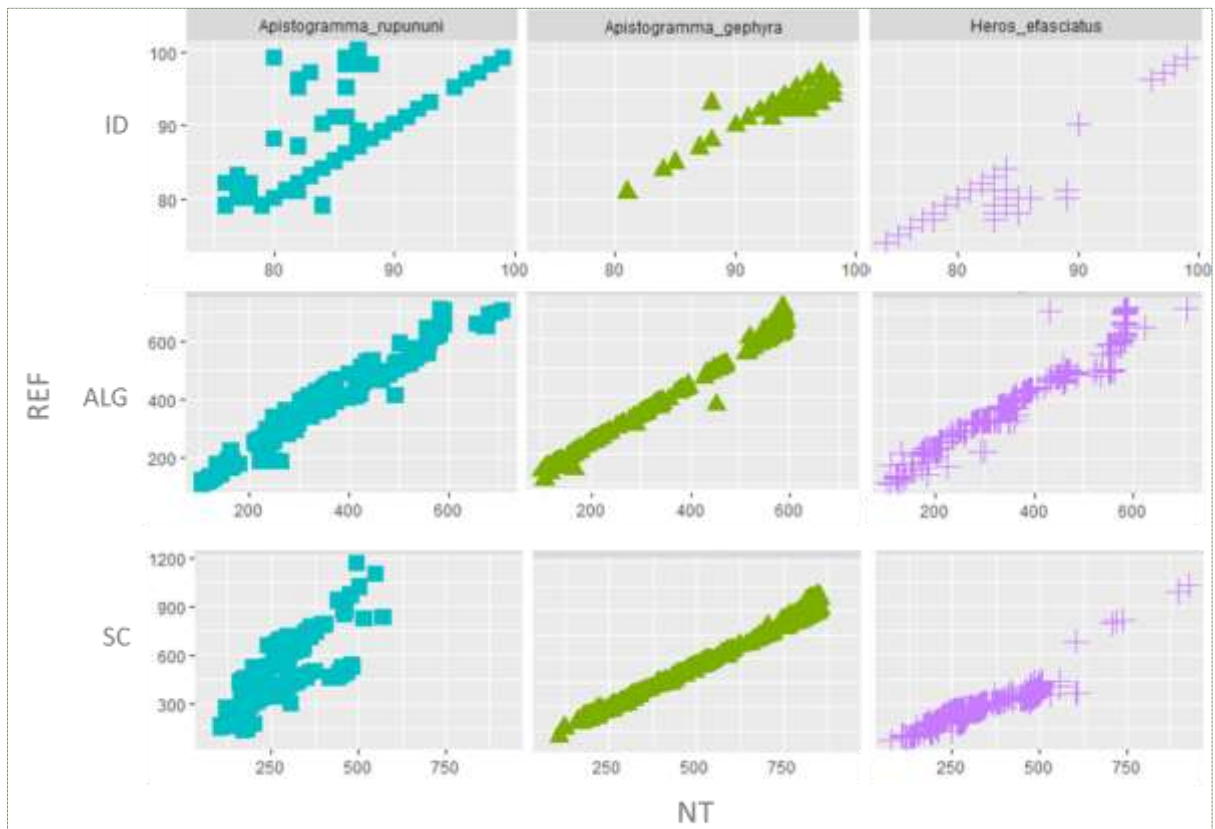


Figura 26 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.

Astyanax bimaculatus ($M_eID_REF=77,7\%$ x *Crenicichla britskii*, ID_NT=86,2%), *Gasteropelecus levis* ($M_eID_REF=83,3\%$ x *Crenicichla britskii*, ID_NT=96,2%), *Guianacara dacrya* ($M_eID_REF=85\%$ x *Crenicichla britskii*, ID_NT=86%), *Heterocharax virgulatus* ($M_eID_REF=100\%$ x *Pygocentrus nattereri*, ID_NT=100%), *Hemigrammus rhodostomus* ($M_eID_REF=76,1\%$ x *Crenicichla britskii*, ID_NT=85%), *Hyphessobrycon bentosi* ($M_eID_REF=82\%$ x *Pygocentrus nattereri*, ID_NT=98,7%), *Hyphessobrycon heterorhabdus* ($M_eID_REF=72,8\%$ x *Rachoviscus crassiceps*, ID_NT=83,3%), *Laetacara dorsigera* ($M_eID_REF=81,5\%$ x *Apistogramma agassizii*, ID_NT=86%), *Mesonauta acora* ($M_eID_REF=79,2\%$ x *Crenicichla britskii*, ID_NT=86,2%), *Myloplus rubripinnis*

($M_eID_REF=90\%$ x *Metynnis fasciatus*, $ID_NT=99\%$), *Peckoltia vittata* ($M_eID_REF=86\%$ x *Farlowella smithi*, $ID_NT=96\%$), *Pterophyllum scalare* ($M_eID_REF=91,1\%$ x *Atropus atropos*, $ID_NT=94,3\%$), *Chilodus punctatus* ($M_eID_REF=91,1\%$ x *Atropus atropos*, $ID_NT=94,3\%$) e *Serrapinnus kriegi* ($M_eID_REF=81\%$ x *Crenicichla alta*, $ID_NT=92$) foram removidas da lista de candidatas, pois seus valores de referência foram discrepantes em relação aos da contraparte, e suas sequências não retornaram entre as top_10. Outro fator que determinou a retirada destes táxons foi o tamanho das sequências em relação ao esperado, todas redor de 150 a 200pb (Figuras 27, 28 e 29).

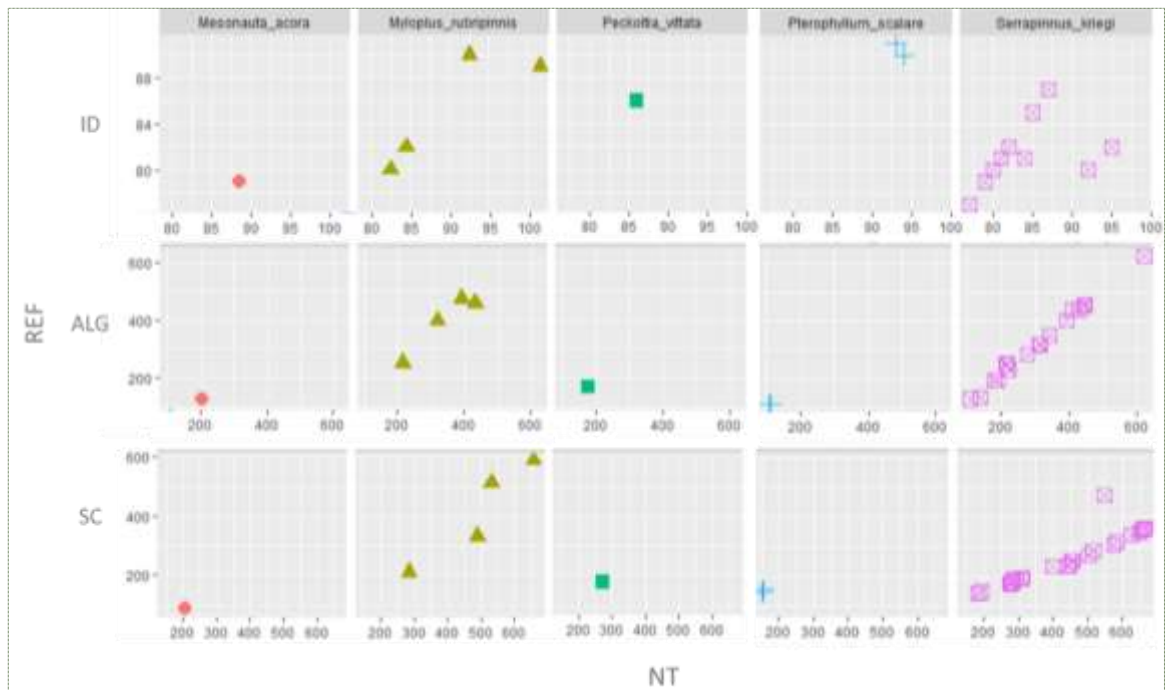


Figura 27 – Distribuição das sequências selecionadas após a aplicação do filtro_classe.

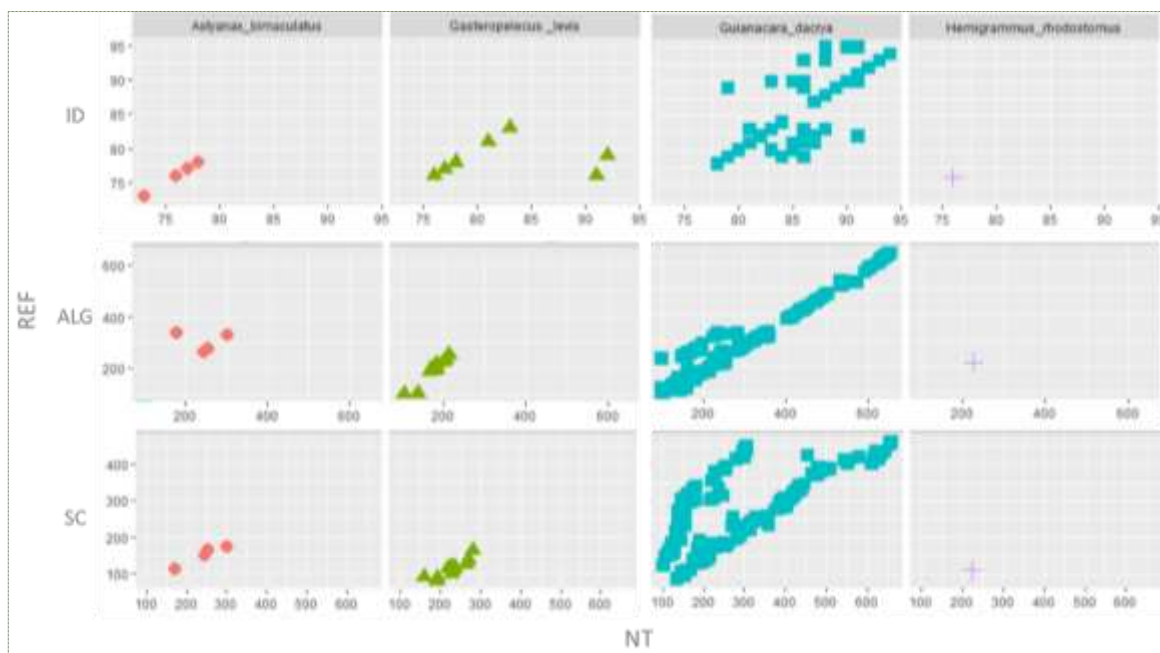


Figura 28 – Distribuição das seqüências selecionadas após a aplicação do filtro_classe.

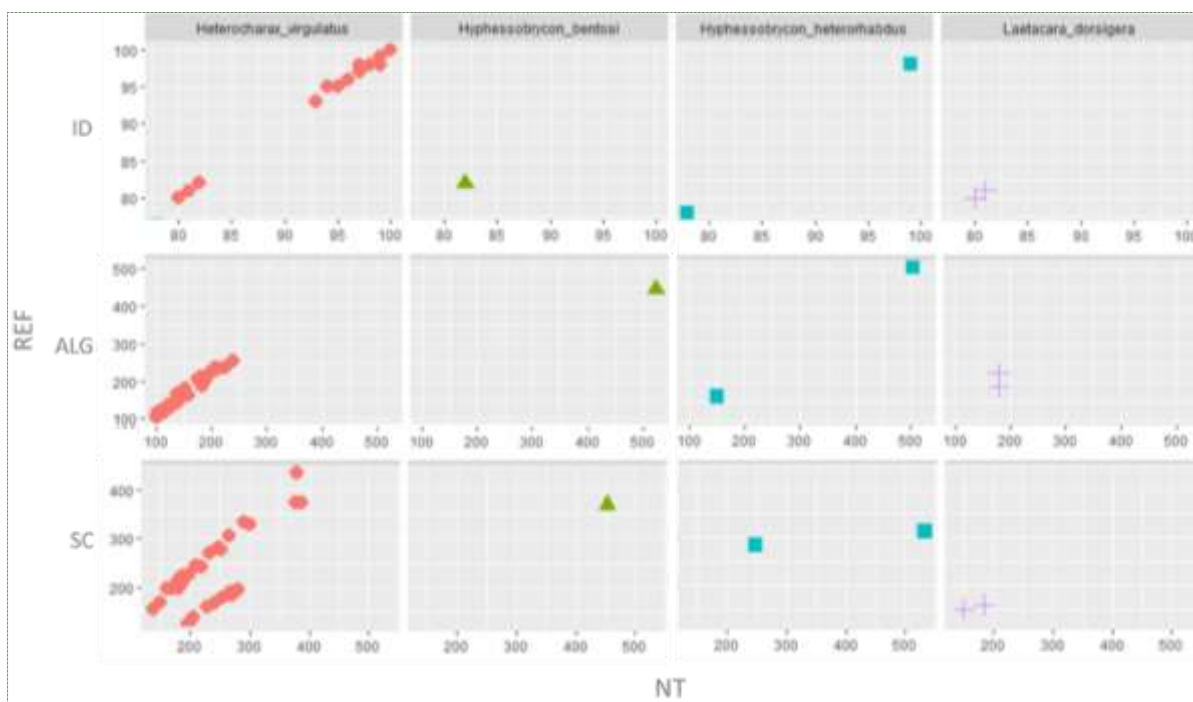


Figura 29 – Distribuição das seqüências selecionadas após a aplicação do filtro_classe.

4.2.2.4 Primer COB

Como resultado do filtro_classe, 967.432 seqüências foram selecionadas como candidatas, sendo que 15 táxons foram definidos pelo banco REF.

Um táxon coincidiu integralmente suas seqüências entre os bancos REF e NT, *Pterophyllum scalare* (n=1; ID_REF=96%; e-value= 1.2e-73), e, portanto, considerado nas análises seguintes (Figura 30).

Apterionotus albifrons (n=27(75055); ID_REF=100%; e-value=6.56e-76), *Gasteropelecus levis* (n=1(481); ID_REF=96%; 5.84e-47), *Chilodus punctatus* (n=3(168); ID_REF=99%; e-value=7.19e-76), *Corydoras ephippifer* (n=1(118); ID_REF=95%; e-value=5.6e-72), *Hemigrammus rhodostomus* (n=1(279); ID_REF=93%; e-value=4.48e-53), *Hypostomus faveolus*, (n=1 (120.746); ID_REF= 100%; evalue=8.99e-80) e, *Pygocentrus nattereri* (n=31(9851); ID_REF=100; e-value=7.19e-81) divergiram entre os bancos, porém, os táxons constavam entre os anotados por NT, e foram mantidos nas análises seguintes. (Figuras 31).

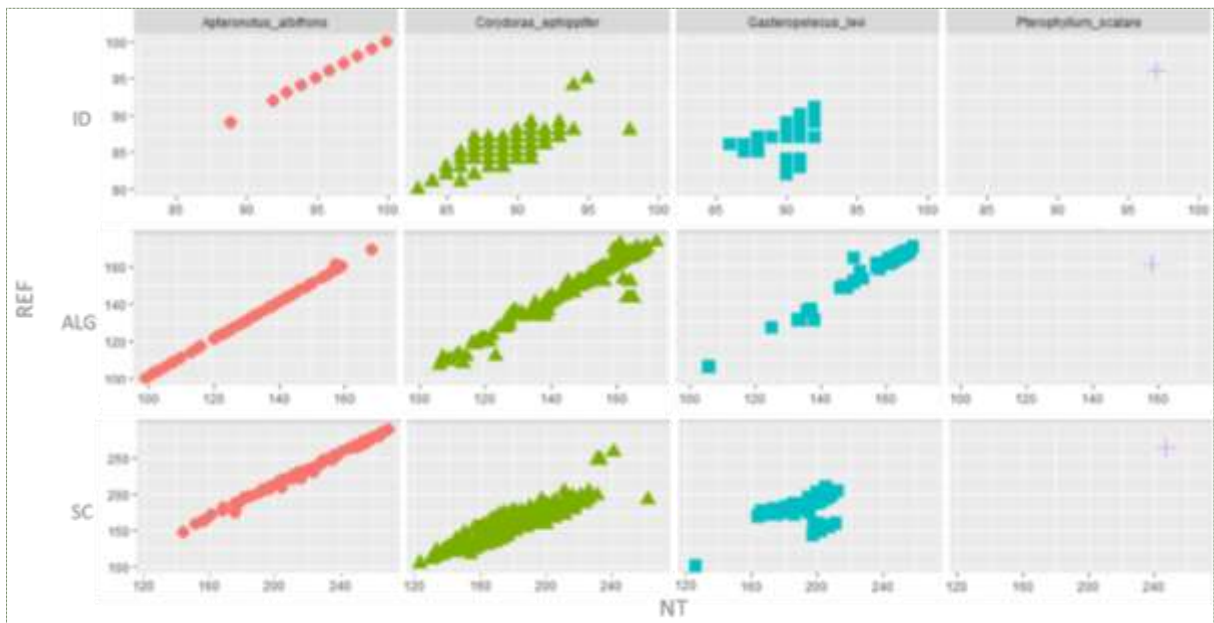


Figura 30 – Distribuição das seqüências selecionadas após a aplicação do filtro_classe.

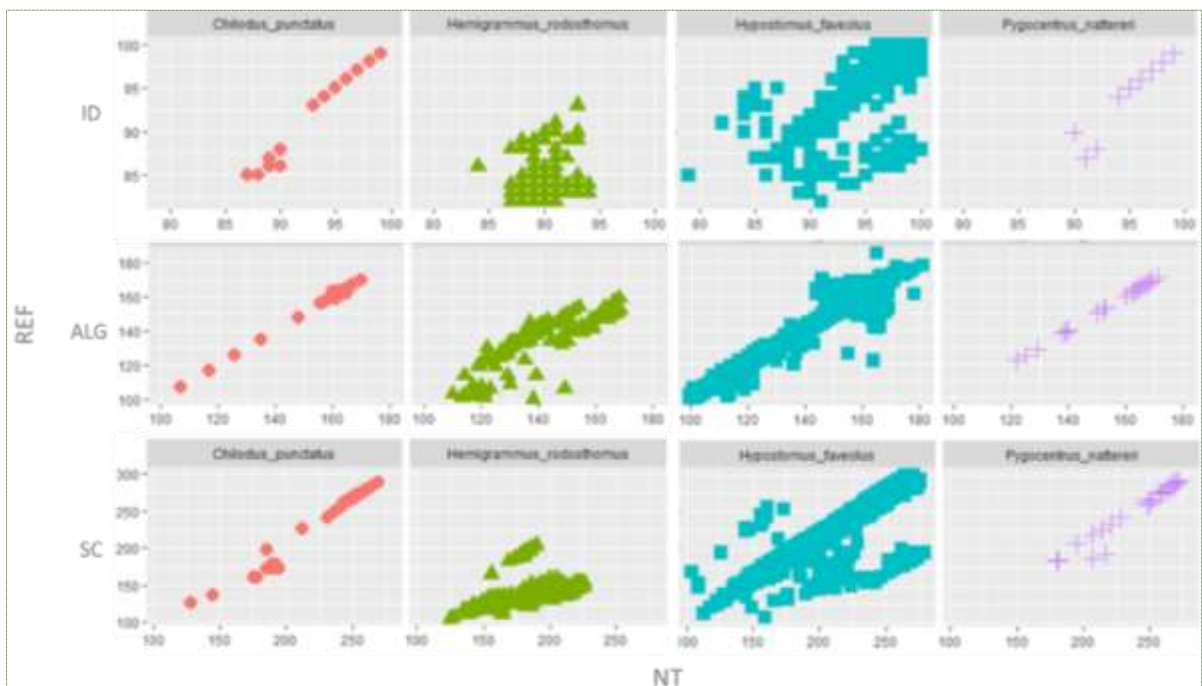


Figura 31 – Distribuição das seqüências selecionadas após a aplicação do filtro_classe.

Farlowella henriquei (n=13(680); ID_REF=100% x *Hemibagrus macropterus*, ID_NT=91,2%), *Panaque armbrusteri* (n=3(1506); ID_REF=100% x *Alburnus tarichi*, ID_NT=92%) divergiram entre os bancos, porém, pelo que se conhece, são as primeiras sequências anotadas para esta região do gene *cox-1*, portanto, com base em ID_REF, expressivamente maior do que ID_NT, foram considerados nas análises seguintes (Figura 32).

Peckoltia vittata (n=8(5855); e-value=5-e60; SEqREQ – *Ancistomus snethlageae*; ID_NT=98%) foi considerado nas análises seguintes, pois, após verificação entre as top_10, o táxon constava na lista (Figura 32).

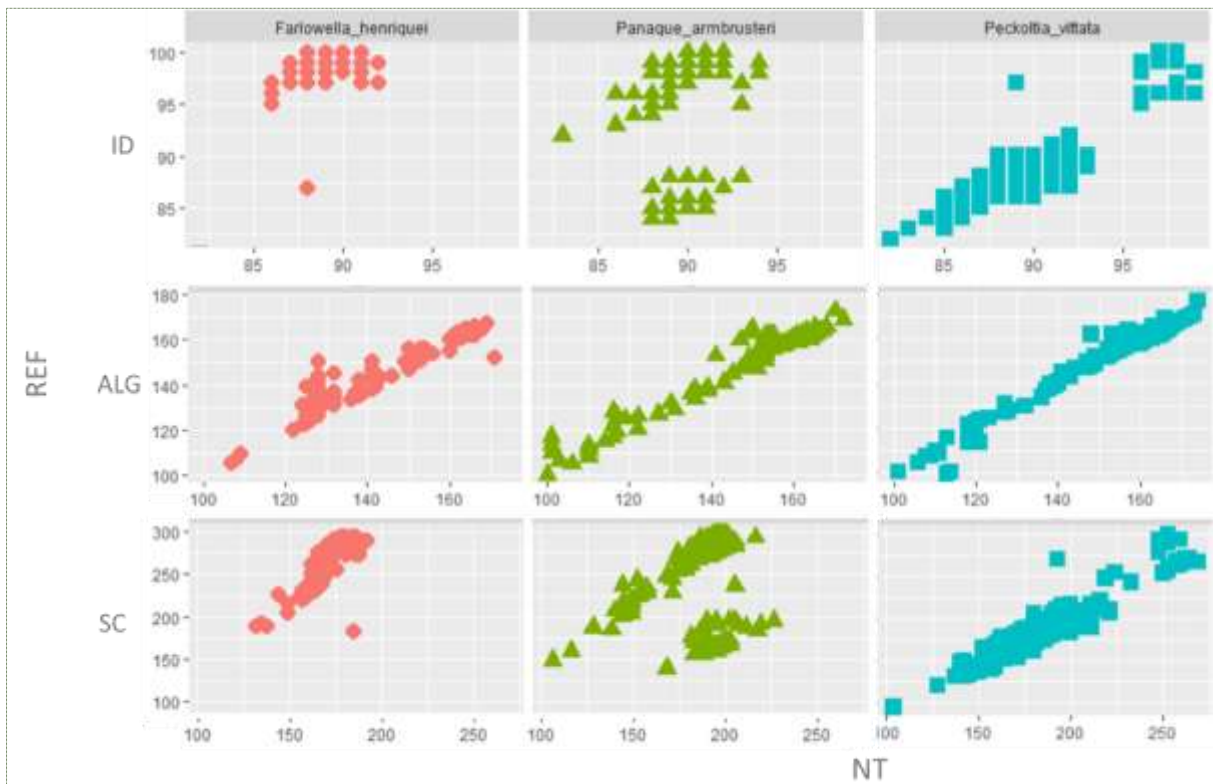


Figura 32 – Distribuição das sequências residuais após a aplicação do filtro_classe.

Quatro táxons foram desconsiderados, *Astyanax bimaculatus* (M_e ID_REF=85,3% x *Pellonula leonensis*, ID_NT=92%), *Apistogramma gephyra* (M_e ID_REF=84,3% x *Symphysodon aequifasciata*, ID_NT=89%), pois, ID_REF foram inferiores aos de ID_NT. *Serrapinus kriegi* (M_e ID_REF=100% x *Pygocentrus nattereri*, ID_NT=100%) não pode ser diferenciado de *Pygocentrus Nattereri*. *Apistogramma rupununi* (M_e ID_REF=99,2% x *Hypostomus sp*, ID_NT=100%), pois, suas sequências além de divergirem quanto ao táxon, se alinharam com membros de outra Ordem (Figura 33).

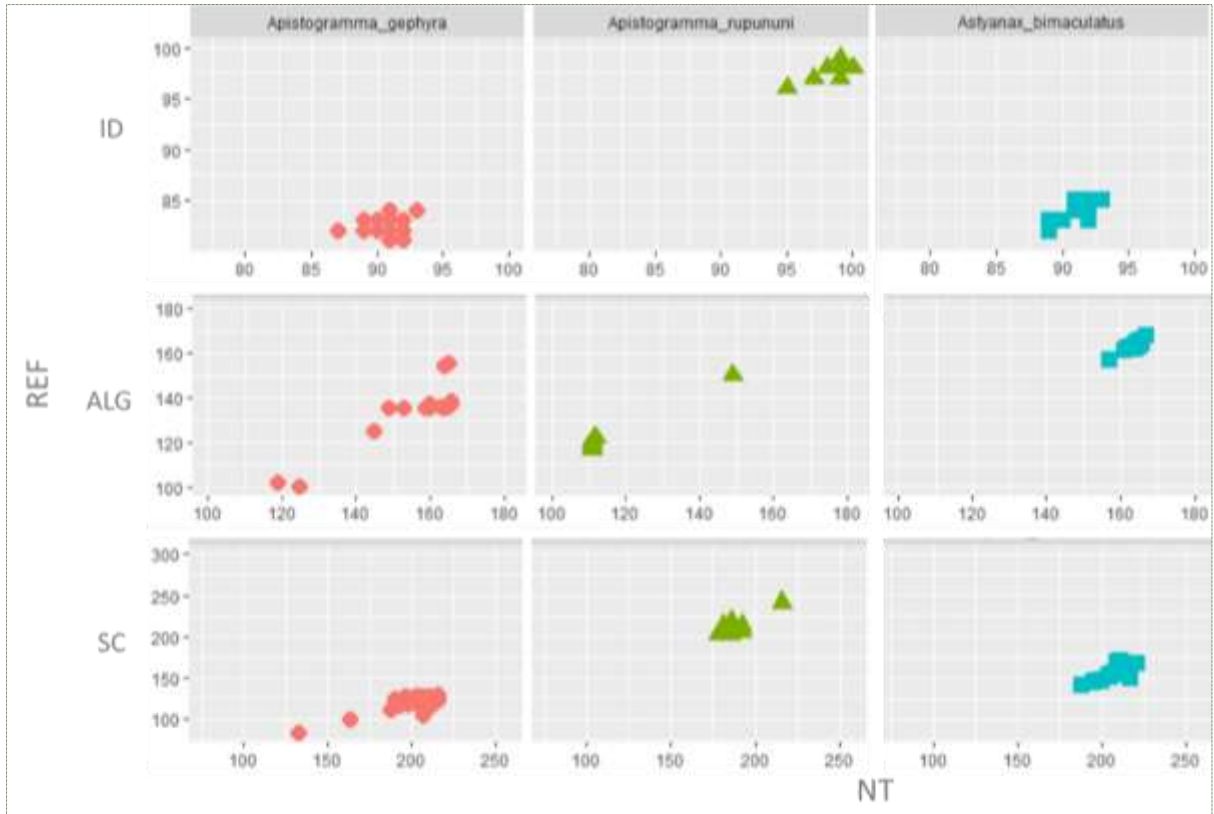


Figura 33 – Distribuição das sequências residuais após a aplicação do filtro_classe.

4.3 LISTA DE ESPÉCIES

Avaliando se houve problemas quanto a divergência das sequências únicas selecionadas por espécie e por *primer*, as árvores filogenéticas indicaram que, tanto COA como 16S não puderam classificar as espécies *Apistogramma rupununi*, *agassizii* e *gephyra*, portanto, para estes dois *primers* apenas o gênero foi considerado como identificado corretamente (Figura 34).

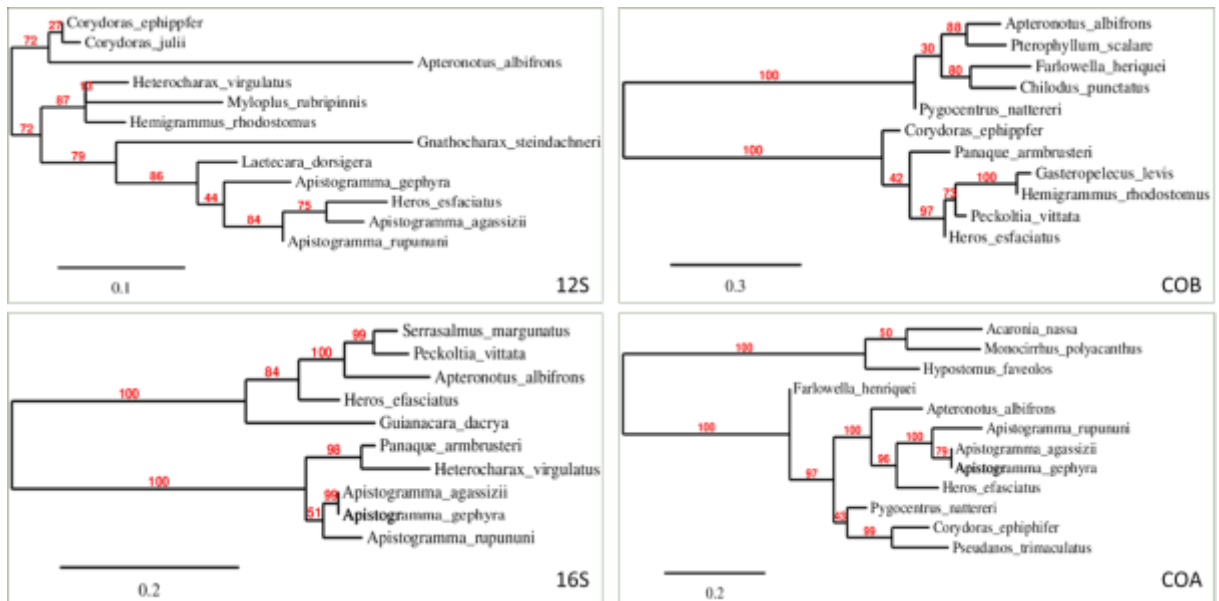


Figura 34 – Cladogramas baseados nas sequências únicas das espécies encontradas nas amostras de eDNA.

Como resultado final, foram identificadas 25 das 34 espécies da comunidade simulada sendo que *Apteronotus albifrons* foi a única espécie identificada por todos os *primers*. O *primer* 12S foi sensível ao sinal de 12 espécies, sendo 4 exclusivas resultando em uma acurácia de 0,6. O 16S identificou 10 espécies, sendo 2 exclusivas, resultando em uma acurácia de 0,5. O COA identificou 12 espécies, sendo 3 exclusivas e acurácia de 0,49, e o COB identificou 11 espécies, sendo 3 exclusivas aprestando acurácia de 0,51 (Tabela 2 – Figura 35). O teste Chi², que é definido ao se delimitar a matriz de contingência não demonstrou nenhuma diferença significativa entre os *primers*.

Tabela 2 – Lista das espécies da comunidade simulada identificadas para cada *primer*

ESPÉCIES	N	12S	ID	16S	ID	COA	ID	COB	ID
<i>Heterocharax virgulatus</i>	451	X	100	X	96	*	*	*	*
<i>Gnathocharax steindachneri</i>	502	X	100	*	*	*	*	*	*
<i>Astyanax bimaculatus</i>	1	*	*	*	*	*	*	*	*
<i>Hemigrammus rhodostomus</i>	60	X	98	*	*	*	*	X	95
<i>Hyphessobrycon bentosi</i>	503	*	*	*	*	*	*	*	*
<i>Hyphessobrycon heterorhabdus</i>	51	*	*	*	*	*	*	*	*
<i>Moenkhausia oligolepis</i>	15	*	*	*	*	*	*	*	*
<i>Pseudanos trimaculatus</i>	15	*	*	*	*	X	100	*	*
<i>Serrapinnus kriegi</i>	4	*	*	*	*	*	*	*	*
<i>Chilodus punctatus</i>	20	*	*	*	*	*	*	X	99
<i>Gasteropelecus levis</i>	51	*	*	*	*	*	*	X	96
<i>Myloplus rubripinnis</i>	10	X	99	*	*	*	*	*	*
<i>Pygocentrus nattereri</i>	1	*	*	*	*	X	100	X	100
<i>Serrasalmus marginatus</i>	1	*	*	X	96	*	*	*	*
<i>Apteronotus albifrons</i>	107	X	99	X	95	X	97,5	X	100
<i>Acaronia nassa</i>	2	*	*	*	*	X	99	*	*
<i>Apistogramma agassizii</i>	818	X	97	X	95	X	96	*	*
<i>Apistogramma gephyra</i>	50	X	98	X	96	X	97	*	*
<i>Apistogramma rupununi</i>	652	X	100	X	94	X	100	*	*
<i>Crenicichla johanna</i>	67	*	*	*	*	*	*	*	*
<i>Guianacara dacrya</i>	10	*	*	X	94	*	*	*	*
<i>Heros efasciatus</i>	4	X	98,5	X	92	X	99	*	*
<i>Laetacara dorsigera</i>	96	X	97	*	*	*	*	*	*
<i>Mesonauta acora</i>	19	*	*	*	*	*	*	*	*
<i>Pterophyllum scalare</i>	25	*	*	*	*	*	*	X	96
<i>Satanoperca jupari</i>	1	*	*	*	*	*	*	*	*

ESPÉCIES	N	12S	ID	16S	ID	COA	ID	COB	ID
<i>Monocirrhus polyacanthus</i>	9	*	*	*	*	X	98	*	*
<i>Platystacus costatus</i>	6	*	*	*	*	*	*	*	*
<i>Corydoras ephippifer</i>	30	X	98	*	*	X	95	X	95
<i>Corydoras julii</i>	30	X	100	*	*	*	*	*	*
<i>Farlowella henriquei</i>	5	*	*	*	*	X	96	X	100
<i>Hypostomus faveolus</i>	4	*	*	*	*	X	96	X	100
<i>Panaque armbrusteri</i>	2	*	*	X	95	*	*	X	100
<i>Peckoltia vittata</i>	1	*	*	X	95	*	*	X	100

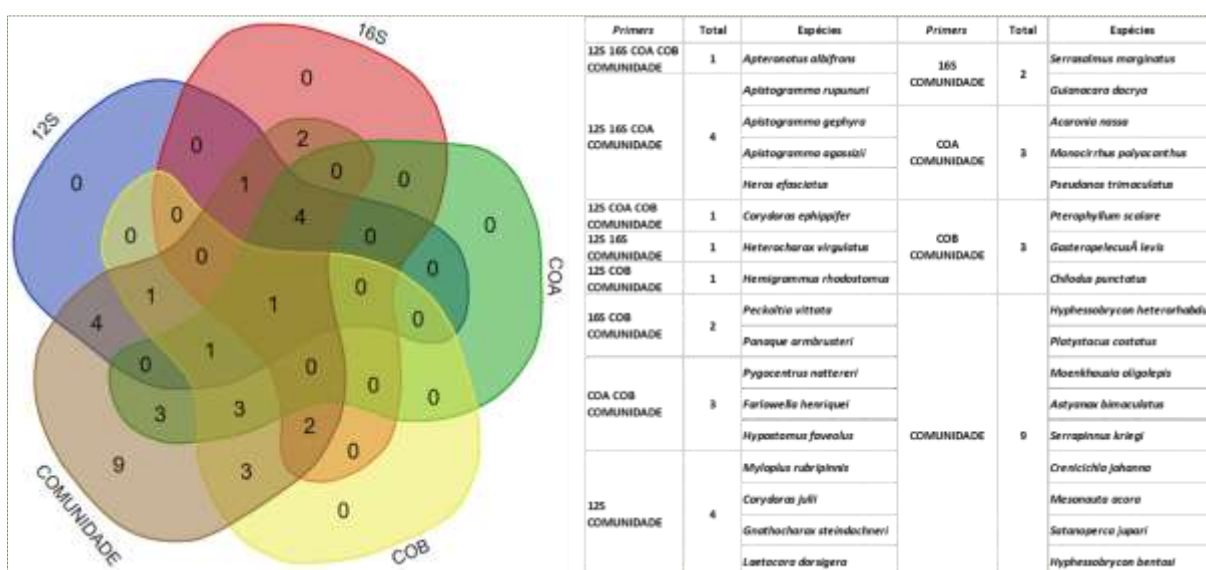


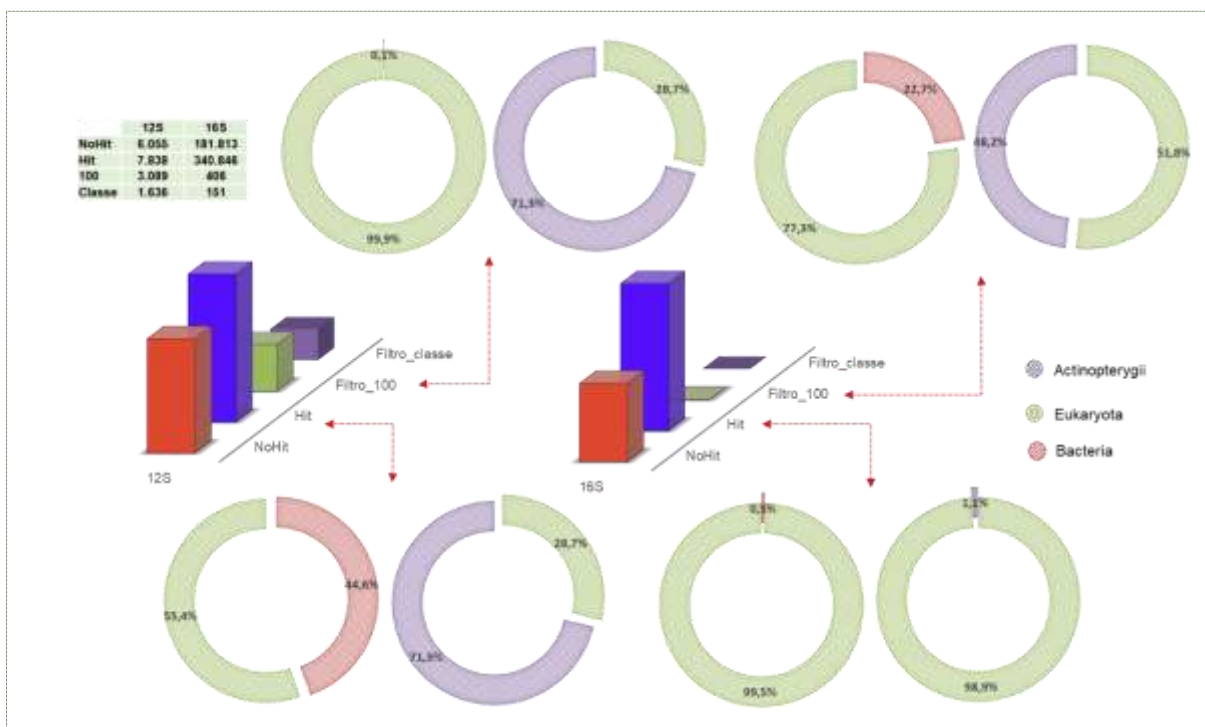
Figura 35 – Contribuição de cada *primer* para o inventário das espécies da comunidade simulada.

4.4 APLICAÇÃO EM ECOSISTEMA NATURAL

4.3.1 Distribuição das seqüências

Como resultado do alinhamento, neste caso apenas ao banco NT, para o *primer* 12S, 56,4% das seqüências foram similares a de algum tipo de organismo já descrito. Dentre os organismos classificados, 45% corresponderam a Bacteria e 55% corresponderam a Eukaryota sendo que dentre estes, 71,3% foram classificados como Actinopterygii. Quando aplicado o filtro_100 foram eliminadas 31% das seqüências, restando 100% de Eukaryota sendo que dentre estes, 66% classificados como Actinopterygii, não havendo nenhuma exclusão de seqüências classificadas para esta Classe. Quando aplicado o filtro_classe, foram eliminadas 63,3% das seqüências e a porcentagem final considerada nas análises seguintes correspondeu a 11,8% do dado bruto (Figura 36).

Em relação ao 16S, 65,2% das sequências foram similares a de algum tipo de organismo já descrito. Dentre os organismos classificados, 0,5% corresponderam a Bacteria e 99,5% corresponderam a Eukaryota, sendo que, dentre estes, 1% foram classificados como Actinopterygii. Quando aplicado o filtro_100 foram eliminadas 99% das sequências, restando 23% de Bacteria e 77% de Eukaryota, sendo que, dentre estes, 48,2% classificados como Actinopterygii, não havendo nenhuma exclusão de sequências classificadas para esta Classe. Quando aplicado o filtro_classe, foram eliminadas 37,2% das sequências e a porcentagem final considerada nas análises seguintes correspondeu a 0,03% do dado bruto (Figura 36).



4.3.2 Validação das espécies

A lista fornecida pelo IEPA dos peixes conhecidos para o rio Aporema compreende 116 espécies e 84 gêneros, sendo que, quando comparado os resultados das amostras de eDNA, 12S identificou 23 gêneros, enquanto 16S identificou 11 deles, representando uma resolução de 26,2% e 7,44%.

Quando considerado todos os dados em conjunto foi possível identificar em nível de espécie 8 das 121 e indicar que existem mais três possíveis ainda não capturadas, *Pimelodella cristata*, *Apteronotus albifrons* e *Geophagus brasiliensis* (Tabela 3). Outro dado secundário interessante que pode ser computado indica que existe a possibilidade de outros 18 gêneros, de

conhecida distribuição para o rio Araguari, o qual o Aporema é um dos tributários, e que podem ocorrer na região de estudo e ainda não foram observados pelos pesquisadores (Tabela 4).

Tabela 3 – Lista das espécies descritas para o rio Aporema_Amapá/Brasil com os respectivos possíveis gêneros identificados pela técnica de eDNA.

ESPÉCIES RIO APOREMA	12S	ID MAX	ID MIN	E-VALUE	SCORE	16S	ID MAX	ID MIN	E-VALUE	SCORE	N
<i>Acanthopoma sp.</i>											
<i>Acestrocephalus sp.</i>											
<i>Acestrorhynchus cf. microlepis</i>	<i>Acestrorhynchus sp</i>	96	90	3.16e-09	70	<i>Acestrorhynchus sp</i>	95	95	2.13e-73	286	3
<i>Acestrorhynchus falcatus</i>											
<i>Ageneiosus cf. inermis</i>	<i>Ageneiosus pardalis</i>	99	97	7.59e-22	113						2
*<i>Ageneiosus ucayalensis</i>	<i>Ageneiosus ucayalensis</i>	100	92	1.15e-40	176						26
<i>Agoniatés halecinus</i>											
<i>Amaralia hypsiura</i>											
<i>Anchovia aff. surinamensis</i>											
<i>Anchoviella cf. guianensis</i>											
<i>Ancistrus sp.</i>											
<i>Apistogramma cf. linkei</i>											
<i>Apteronotus sp.</i>	<i>Apteronotus albifrons*</i>	100	91	1.38e-57	233	<i>Apteronotus albifrons*</i>	100	94	2.25e-119	439	42
	<i>Apteronotus rostratus</i>	92	92	3.14e-25	125	<i>Apteronotus rostratus</i>	98	98	7.54e-74	288	2
<i>Astyanax aff. bimaculatus</i>	<i>Astyanax mexicanus</i>	98	96	2.09e-53	219						12
						<i>Astyanax maximus</i>	98	98	1.07e-96	363	1
<i>Auchenipterus aff. ambyiacus</i>											
<i>Awaous cf. flavus</i>											
<i>Bathygobius sp.</i>											
<i>Bivibranchia sp.</i>											
<i>Boulengerella cuvieri</i>											
<i>Brachyplatystoma filamentosum</i>											
<i>Bryconops aff. caudomaculatus</i>											
<i>Bryconops cf. alburnoides</i>	<i>Bryconops caudomaculatus</i>	96	91	9.99e-48	200						12
*<i>Bryconops cf. caudomaculatus</i>											
<i>Bryconops sp.</i>											
<i>Caquetaia cf. spectabilis</i>											
<i>Centromochlus cf. heckelii</i>											
<i>Cetopsis coecutiens</i>											
<i>Charax cf. gibbosus</i>	<i>Charax stenopterus</i>	94	91	3.55e-46	195						12
<i>Charax sp.</i>											
<i>Cichla ocellaris</i>	<i>Cichla ocellaris</i>	99	93	5.65e-56	228						2
<i>Cichla cf. pinima</i>											
<i>Creagrutus cf. cracentis</i>	<i>Creagrutus bolivari</i>	96	96	4.77e-31	144						1
	<i>Creagrutus melanzonus</i>	97	91	7.39e-52	214						21
<i>Crenicichla gr. johanna</i>											
<i>Crenicichla strigata</i>											
<i>Curimata inornata</i>	<i>Curimata mivartii</i>	98	93	2.09e-53	219						17
<i>Curimatella aff. immaculata</i>											
<i>Cyphocharax cf. goldingi</i>											
<i>Cyphocharax spiluropsis</i>											

ESPÉCIES RIO APOREMA	12S	ID MAX	ID MIN	E-VALUE	SCORE	16S	ID MAX	ID MIN	E-VALUE	SCORE	N
<i>Eigenmannia aff. trilineata</i>	<i>Eigenmannia limbata</i>	98	88	4.52e-09	69						69
	<i>Eigenmannia sp</i>	99	92	2.29e-51	212						47
						<i>Eigenmannia virescens</i>	98	98	1.89e-105	392	1
<i>Eleotris cf. pirois</i>											
<i>Geophagus cf. altifrons</i>	<i>Geophagus brasiliensis*</i>	100	98	1.41e-33	152	<i>Geophagus brasiliensis*</i>	92	92	2.89e-54	223	11
<i>Geophagus cf. proximus</i>	<i>Geophagus steindachneri</i>	96	96	3.23e-39	171						1
<i>Gladioglanis sp.</i>											
<i>Harttia cf. fowleri</i>											
<i>Hemigrammus cf. geisleri</i>											
<i>Hemiodus cf. quadrimaculatus</i>											
<i>Hemiodus unimaculatus</i>											
<i>*Hoplias malabaricus</i>	<i>Hoplias malabaricus</i>	100	93	1.59e-57	233						21
<i>Hoplias aimara</i>	<i>Hoplias intermedius</i>	100	93	2.5e-15	91						43
<i>Hoplias curupira</i>											
<i>*Hyphessobrycon cf. eques</i>	<i>Hyphessobrycon eques</i>	99	97	5.86e-55	224						5
<i>Hyphessobrycon sp.</i>	<i>Hyphessobrycon guarani</i>	90	88	2.22e-34	156						2
						<i>Hyphessobrycon herbertaxelrodi</i>	89	89	4.24e-42	181	1
<i>Hypoclinemus cf. mentalis</i>											
<i>Hypomasticus cf. julii</i>											
<i>Hypomasticus sp.</i>											
<i>Hypostomus sp</i>	<i>Hypostomus sp</i>	100	92	1.55e-56	229	<i>Hypostomus sp.</i>	97	97	1.01e-115	427	27
<i>Hypostomus cf. plecostomus</i>	<i>Hypostomus affinis</i>	96	88	9.58e-49	204						211
<i>Hypostomus gr. cochliodon</i>						<i>Hypostomus ancistroides</i>	97	97,5	1.95e-08	65.1	1
<i>Ilisha amazonica</i>											
<i>Jupiaba cf. acanthogaster</i>											
<i>Jupiaba polylepis</i>											
<i>*Krobia cf. guianensis</i>	<i>Krobia guianensis</i>	90	90	6.54e-39	171						1
<i>Laemolyta cf. taeniata</i>											
<i>Leporinus aff. fasciatus</i>	<i>Leporinus affinis</i>	100	94	4.64e-58	235						36
<i>Leporinus aff. maculatus</i>											
<i>Leporinus aff. parae</i>											
<i>Leporinus cf. melanostictus</i>											
<i>Limatulichthys cf. griseus</i>											
<i>Loricaria cataphracta</i>											
<i>Lycengraulis cf. batesii</i>											
<i>Megalodoras uranoscopus</i>											
<i>Mesonauta cf. guyanae</i>											
<i>Metynnis cf. maculatus</i>	<i>Metynnis hypsauchen</i>	95	93	3.26e-47	198	<i>Metynnis hypsauchen</i>	94	94	1.22e-98	370	4
<i>Microphilypnus ternetzi</i>											
<i>Moenkhausia cf. ceros</i>						<i>Moenkhausia comma</i>	96	96	1.05e-66	264	1
<i>Moenkhausia cf. gracilima</i>						<i>Moenkhausia oligolepis</i>	100	100	0.000226	56.5	1
<i>Moenkhausia copei</i>						<i>Moenkhausia xinguensis</i>	96	97	2.78e-90	343	1

ESPÉCIES RIO APOREMA	12S	ID MAX	ID MIN	E-VALUE	SCORE	16S	ID MAX	ID MIN	E-VALUE	SCORE	N
<i>Mylesinus paraschomburgkii</i>											
<i>Myleus aff. setiger</i>	<i>Myleus sp</i>	94	94	7.73e-27	130						1
	<i>Myleus schomburgkii</i>	93	93	4.48e-43	185						1
<i>Myloplus asterias</i>	<i>Myloplus rubripinnis</i>	100	92	3.02e-15	91						6
<i>Myloplus cf. lobatus</i>											
*<i>Myloplus cf. rubripinnis</i>											
<i>Myloplus rhomboidalis</i>											
<i>Mylossoma cf. duriventre</i>											
<i>Opsodoras aff. morei</i>											
<i>Osteoglossum bicirrhosum</i>											
<i>Pachypops fourcroyi</i>											
<i>Paratrygon cf. aireba</i>											
<i>Parauchenipterus porosus</i>											
*<i>Peckoltia cf. sabaji</i>						<i>Peckoltia sabaji</i>	98	98	4.59e-127	465	1
						<i>Peckoltia compta</i>	99,6	98	4.62e-126	461	5
						<i>Peckoltia vittata</i>	100	94	4.13e-145	525	5
						<i>Peckoltia oligospila</i>	99	95	6.02e-55	224	15
<i>Pellona castelnaena</i>											
<i>Petulanus cf. intermedius</i>											
<i>Pimelodella cf. howesi</i>	<i>Pimelodella cristata*</i>	98	95	7.23e-52	214						10
*<i>Pimelodus blochii</i>	<i>Pimelodus blochii</i>	100	98	1.62e-56	229						6
<i>Pimelodus ornatus</i>											
<i>Plagioscion montei</i>											
<i>Plagioscion squamosissimus</i>											
<i>Platynemichthys notatus</i>											
<i>Pristigaster sp.</i>											
<i>Propimelodus sp.</i>											
<i>Pseudacanthicus sp.</i>											
<i>Pseudopimelodus sp.</i>											
<i>Pyrrhulina cf. marilynae</i>											
<i>Retroculus cf. lapedifer</i>											
<i>Rhamphichthys aff. rostratus</i>											
<i>Rhamphichthys marmoratus</i>											
<i>Rivulus sp.</i>											
<i>Satanoperca cf. jurupari</i>											
<i>Schizodon cf. fasciatus</i>											
<i>Serrasalmus cf. maculatus</i>	<i>Serrasalmus eigenmanni</i>	92	92	3.01e-18	101						1
<i>Serrasalmus cf. rhombeus</i>						<i>Serrasalmus rhombeus</i>	100	96	1.85e-46	195	5
<i>Serrasalmus sp.</i>	<i>Serrasalmus sp</i>	100	94	3.4e-17	98	<i>Serrasalmus sp</i>	98	98	2.2e-138	502	55
<i>Steatogenys elegans</i>											
<i>Sternarchella sp.</i>											
<i>Tetragonopterus chalcus</i>											
<i>Triportheus auritus</i>											
RIO	Espécies	121		Gêneros	87						
12S				Gêneros	22				Resolução		26,19

Tabela 4 – Listas das possíveis ocorrências ainda não identificadas pelos pesquisadores do IEPA, porém, listada pela técnica de eDNA.

POSSIBILIDADES	ESPÉCIES	ID MAX	ID MIN	E-VALUE	SCORE	N
<i>Brachyhypopomus</i>	<i>Brachyhypopomus occidentalis</i>	98	83	6.49e-53	217	24
	<i>Brachyhypopomus sp</i>	96	96	3.06e-48	202	2
<i>Bryconamericus</i>	<i>Bryconamericus iheringii</i>	96	92	8,00E-50	207	12
	<i>Bryconamericus pachacuti</i>	95	95	7.49e-21	109	1
	<i>Bryconamericus sp</i>	96	96	2.79e-50	209	1
<i>Cheirodon</i>	<i>Cheirodon ibicuhensis</i>	96	95	2.47e-50	209	6
<i>Corydoras</i>	<i>Corydoras schwartzi</i>	99	99	5.22e-55	224	2
<i>Curimatopsis</i>	<i>Curimatopsis evelynae*</i>	99	92	1.97e-54	223	63
<i>Farlowella</i>	<i>Farlowella paraguayensis</i>	100	94	1.65e-56	229	51
	<i>Farlowella platorynchus</i>	93	93	1.18e-44	190	1
	<i>Farlowella reticulata</i>	97	97	2.43e-51	212	4
	<i>Farlowella taphorni</i>	96	96	5.95e-29	137	1
<i>Gymnorhamphichthys</i>	<i>Gymnorhamphichthys sp</i>	80	80	3.07e-15	93	2
<i>Hypobrycon</i>	<i>Hypobrycon leptorhynchus</i>	98	87	5.34e-31	144	196
<i>Knodus</i>	<i>Knodus heteresthes*</i>	96	92	8.87e-50	207	4
<i>Laetacara</i>	<i>Laetacara thayeri</i>	94	94	1.7e-34	156	2
<i>Nannacara</i>	<i>Nannacara anomala</i>	98	98	9.75e-61	243	1
<i>Polycentrus</i>	<i>Polycentrus schomburgkii</i>	100	98	1.24e-43	186	8
<i>Pterophyllum</i>	<i>Pterophyllum altum</i>	95	91	1.49e-40	176	30
<i>Pygocentrus</i>	<i>Pygocentrus nattereri</i>	100	97	1.76e-105	392	5
<i>Rineloricaria</i>	<i>Rineloricaria lanceolata</i>	100	94	4.46e-09	67	4
	<i>Rineloricaria fallax</i>	96	95	2.63e-49	205	2
	<i>Rineloricaria sp</i>	100	87	1.25e-35	159	10
<i>Sternopygus</i>	<i>Sternopygus dariensis</i>	93	91	4.52e-42	181	6
<i>Synbranchus</i>	<i>Synbranchus marmoratus</i>	98	98	2.21e-52	216	1
<i>Trachelyopterus</i>	<i>Trachelyopterus galeatus*</i>	100	96	1.61e-56	229	3
Gêneros						18

5 DISCUSSÃO

A falta de homogeneidade do produto gerado pela reação de sequenciamento de COA e 16S gerou um viés quanto a número de sequências finais para cada um deles. Uma das vantagens que se espera alcançar através do uso de tecnologias de sequenciamento de nova geração é a capacidade imensa de geração de cópias idênticas para cada sequência que conste em uma amostra, seja de DNA ou RNA. Esta robustez faz com que exista um número mínimo de réplicas capazes de confirmar que o produto gerado não é um artefato da técnica ou algo atribuído ao acaso⁷².

Ao que tudo indica, a etapa de seleção de tamanho de *amplicon* que consta no protocolo de sequenciamento da plataforma GLS FX+, tem baixa eficiência para remoção de todos os resquícios de amplificações inespecíficas que, neste caso, acabaram comprometendo a reação de sequenciamento pelo simples fato destes serem pequenos e, portanto, mais facilmente amplificados. Como a escolha por esta plataforma foi estabelecida exatamente para que *amplicons* com mais de 400 pares de bases fossem sequenciados sem a necessidade de fragmentação, COA e 16S tiveram baixo rendimento e qualidade em relação aos produtos produzidos para 12S e COB, que foram homogêneos quanto ao tamanho das sequências geradas.

O material genético de grandes organismos constantes em uma amostra de eDNA, tais como os peixes, são em sua maioria extremamente fragmentados e, neste caso, grandes *amplicons* passam a ser cada vez menos abundantes. Porém, tanto a amostra COA como a 16S apresentaram sequências informativas dentro do padrão de extensão esperado, portanto, a aplicação do mesmo protocolo utilizado para seleção de *amplicons* para 12S e COB poderia reproduzir melhores resultados. A diferença entre as técnicas é que a padronização dos *amplicons* a serem sequenciados na plataforma PGM ION se dá pela recuperação do fragmento desejado através de captura da banda amplificada observada em gel de agarose, e portanto, uma seleção muito mais curada.

Como a seleção dos *primers* foi baseada no modelo DNA barcode, onde, uma região específica, neste caso de um gene mitocondrial, é usada como parâmetro para se medir a distância genética entre diferentes organismos, a principal característica destes é a de que as variações intraespecíficas de dada espécie não sejam maiores do que sua variação interespecífica⁷³. Desta forma, uma região ideal seria aquela onde exista em suas extremidades 5' e 3' sequências de bases extremamente conservadas, como é o caso dos *primers* selecionados, o que garante o ancoramento do mesmo, e que, entre estas, exista uma sequência variável, a

qual permite avaliar as mutações nas bases nucleotídicas e, assim, definir os caracteres moleculares únicos para cada espécie existente⁷⁴. No entanto, fugindo a regra, devido ao grande número de classes, ordens, famílias, gêneros e espécies, uma região conservada pode não ser “universal” para a maioria dos organismos e, para se desenhar *primers* capazes de amplificar o DNA de todos eles se faz necessário que algumas de suas bases sejam degeneradas, facilitando assim o seu ancoramento nas extremidades 5’ e 3’⁷⁵.

Esta característica é encontrada em muitos dos *primers* desenhados para a região 5’ do gene *cox-1*, e explica a predisposição de COA em amplificar um maior número de bactérias em relação aos demais *primers* utilizados. Neste caso, a existência de homologies entre genes do complexo citocromo oxidase c (*cox-1*, *cox-2* e *cox-3*) presentes tanto em metazoários como em bactérias e arqueias faz com que esta região se torne extremamente “universal”⁷⁶, uma vez que a presença de homólogos do gene *cox-1*, diferentemente das subunidades *cox-2* e *cox-3* que podem ou não estar presentes em organismos unicelulares, encontra-se sempre presente em bactérias e arqueias^{77,78}.

Em COA, diferentemente de 12S e 16S, por se tratar de um *primer* específico para a comunidade de peixes, vieses como estes não são desejáveis, pois, uma amostra de eDNA pode conter milhões de partículas bacterianas e, portanto, reduzir a capacidade preditiva da técnica, já que proporcionalmente terão maior chance de replicação de seu DNA.

No entanto, este viés não foi observado. Mesmo com menor número de sequências proporcionais ao grupo Actinopterygii em relação a 12S e COB, COA pode definir 12 espécies e, chegou a elencar 26 entre as 34 presentes na comunidade simulada.

No geral, somente o COB teve alta afinidade para o grupo Actinopterygii com mais de 90% das sequências sendo atribuídas ao mesmo. O delineamento do *primer* COB, diferentemente do COA, foi feito sem degeneração de suas bases, o que o torna mais específico ao grupo em questão, além de que, sua localização de ancoramento está baseada na região 3’⁶⁹. Alguns estudos vêm demonstrando que esta região do gene *cox-1* ainda é negligenciada quando ao seu uso em estudos filogenéticos apresentando bons resultados quanto a discriminação de caracteres moleculares em grupos taxonômicos mal resolvidos pela da região 5’ do gene *cox-1*⁷⁹.

Houve clara diferença na capacidade de cada *primer* quanto a identificação das espécies. No entanto, o uso de dois bancos referenciais associado ao critério de filtros, auxiliaram para que resultados falso-positivo e/ou negativo não fossem computados nas análises, permitindo a identificação, com certo grau de acurácia, de 25 das 34 espécies da comunidade simulada. Esta abordagem, indiretamente, se mostrou uma medida eficaz para a tomada de decisão quanto a

validação ou não de cada espécie identificada por cada um dos *primers* utilizados, pois, em alguns casos, a ID_REF alcançada para algumas delas não chegou ao desejado de 99-100%.

Este artefato pode ser explicado por duas razões: ou i) houve falha na curadoria do banco referencial, onde possivelmente o método utilizado para alinhamento das sequências tenha gerado ruídos⁸⁰, embora este fato explicasse apenas as sequências referenciais de 16S, onde as variações de ID foram muito altas, porém, ao que tudo indica esta seja uma característica do gene 16S para algumas espécies⁸¹⁻⁸³ ou 2) devido ao fato de ter apenas utilizado um exemplar de cada espécie presente na comunidade simulada para modelar o banco referencial⁸⁴. Ao que tudo indica, a segunda hipótese seria a opção mais plausível, pois, para algumas espécies de peixes amazônicos, a variação intraespecífica de algumas delas pode ser ampla, ultrapassando neste caso os 2% preconizados para os padrões DNAbarcode⁸⁵.

Este viesse dificultou a identificação de espécies do gênero *Apistogramma* e da subfamília *Serrasalminea* tanto em COA como em 16S. Algumas espécies do gênero *Apistogramma* podem ter dentro de suas populações ampla variação intraespecífica, sendo que já foram reportadas distâncias genéticas entre as espécies desde grupo variando entre 0 a 8%⁸⁶, o mesmo ocorrendo entre as espécies da subfamília *Serrasalminea*⁸⁷. Neste estudo, estas espécies somente foram identificadas devido ao uso concomitante de outros dois *primers*, indicando que para uma maior robustez, assim como nos estudos de inferências filogenéticas, para a técnica de eDNA também se faz necessário o uso de diferentes regiões e, diferentes genes mitocondriais⁸⁸.

A Ordem Characiformes apresentou o menor número de espécies identificadas, quatro entre as sete. 12S e COB foram os que identificaram maior número de espécies dentre os quatro *primers* utilizados, no entanto, não distinguiram *Serrapinus kriegi* de *Pygocentrus nattereri*, assim como *Pygocentrus nattereri* de *Serrasalmus marginatus*. Ao analisar mais detalhadamente, este fato pode ser atribuído à recente radiação do grupo, quando para algumas regiões mais conservadas do DNA mitocondrial não mostram um grande número de mutações fixadas após a separação das linhagens, impedindo que as espécies fiquem bem diferenciadas entre si⁸⁹. No entanto, dentre as 13 espécies que COA não pode identificar corretamente, seis eram Characiformes, e todas as sequências representativas do grupo foram descartadas devido ao tamanho variar entre 150 a 250bp. Este achado, pode ser indicativo de que, ao menos para as regiões testadas, *amplicons* menores do que 300pb podem não ter adequada resolução para identificar todas espécies⁹⁰.

O *primer* 12S com apenas 100pb é um dos mais utilizados nas técnicas de recuperação de forma indireta de informação acerca das espécies, tais como as baseadas em

metabarcoding^{54,91,92}, no entanto, neste estudo, apenas pode elucidar 12 das 34 espécies presentes na comunidade simulada. COB com capacidade de extensão de 150pb, foi capaz de discriminar apenas 11 das 34 espécies.

Outro fator importante observado, foi que, ao avaliar as sequências de *Heros esfaciatus* em COA, estas foram atribuídas a diferentes espécies conforme a variação do tamanho da sequência, se estabilizando somente após 300bp (Figura 38). Obviamente os valores de ID estarão reduzidos, mas, como já mencionado acima, assim como os Characiformes, os ciclídeos amazônicos apresentam rápida radiação no continente sul americano, ocupando todas as regiões e sofrendo altas taxas de hibridização⁹³, sendo outro grupo de alta dificuldade para se estabelecer bons *amplicons* capazes de identificar todas as espécies, como já foi demonstrado acima para o gênero *Apistogramma* e observado que COB apenas pode identificar uma espécie da Ordem Perciformes.

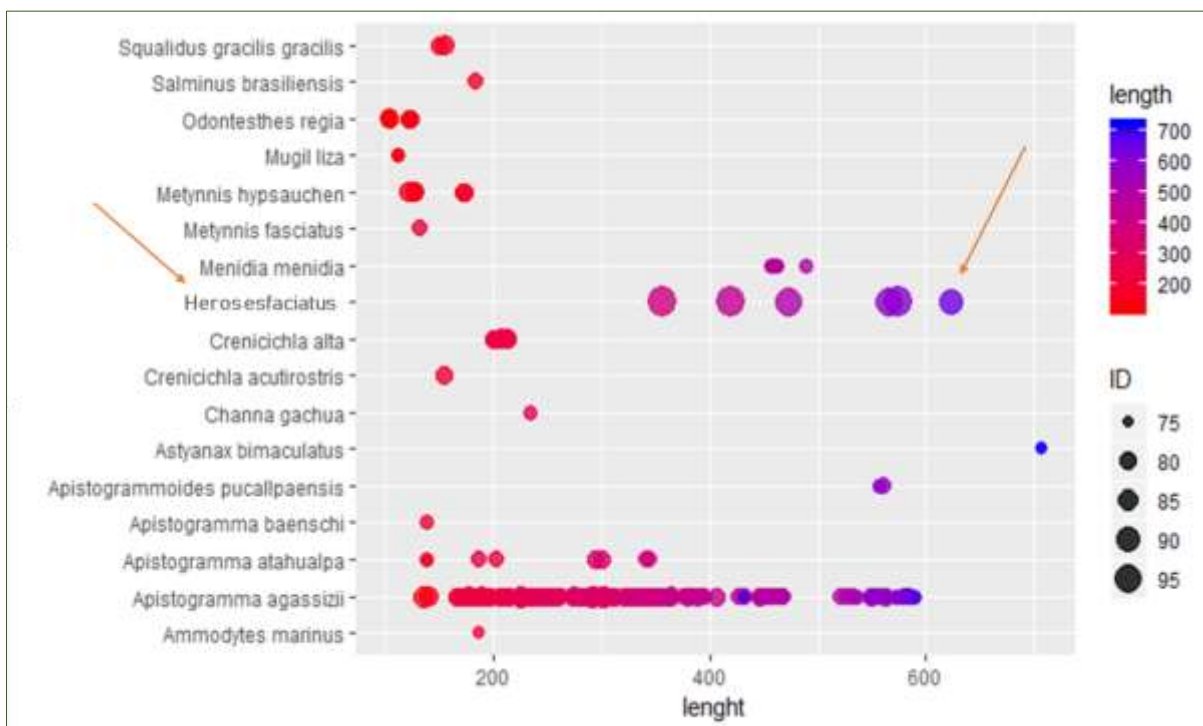


Figura 37 – Distribuição das sequências atribuídas a espécie *Heros esfaciatus* em relação ao tamanho do fragmento, porcentagem de ID e número de espécies que alinharam as reads em NT.

COA foi o único a identificar *Acaronia nasa*, mesmo com apenas duas sequências, estas passaram pelos filtros e seu tamanho foi de acordo com o esperado, não restando dúvidas de sua identificação. COA também pode identificar *Guianacara dacria* e, mesmo que tenham sido eliminadas por problemas quanto ao tamanho das sequências, foram listadas entre as espécies candidatas, *Mesonauta acora* e, *Crenicicla johanna*, as quais junto com *Satonoperca jupari*, eliminada pelo mesmo motivo, foram as únicas dentre os Perciformes a não serem identificadas.

Quando, analisadas as sequências removidas de COA, foi notado que muitas delas em NT indicavam ser representativas do gênero *Crenicicla* deixando uma dúvida quando a possibilidade desta espécie ter sido ou não identificada por este *primer*. Este fato reforça a ideia de que *amplicons* de maiores tamanhos aumentariam a amplitude da técnica de eDNA, já que foi demonstrado ser possível a coleta de grandes fragmentos de DNA.

A classe Siluriformes foi a melhor representada com apenas uma espécie não sendo identificada, *Platystacus costatus*. Ao que tudo indica, o gene *cox-1* tem boa resolução para grupo, com COB identificando o maior número de espécies.

O *primer* 12S identificou as duas espécies do gênero *Corydoras*. No entanto, *Corydoras julii* foi representada apenas por uma sequência e, avaliando detalhadamente ambas, a divergência se confere a apenas uma troca de base. Portanto, esta validação pode ser indicativa de falso-positivo, já que a subfamília *Callichthyidae* também faz parte das “espécies problemas” para a filogenia de peixes⁹⁴.



Figura 38 – Alinhamento entre as sequências do gênero *Corydoras* amplificadas pelo *primer* 12S.

Quando as 8 espécies não identificadas por nenhum dos *primers* foram avaliadas isoladamente, foi notado que correlações entre o número de indivíduos alocados (abundância) e o número de espécies identificadas podem não ser sempre equiparáveis como demonstrado por alguns estudos^{47,95,96-98}. Por exemplo, *Astyanax* (1) e *Serrapinus* (4), ambos menores em abundância na comunidade simulada, tiveram sequências identificadas, mesmo que sendo eliminadas a posteriori, enquanto *Crenicicla* (67) e *Moenkausia* (15) que estavam em maior número, se quer tiveram sequências registradas pelo banco REF, onde o tecido de cada uma delas estava presente, e/ou espécies congêneres quando consultado o banco NT. *Apteronotus*, foi identificado por todos os *primers*, no entanto, teve número de sequências proporcionalmente diferentes para cada um deles. Nas amostras COA, o número de sequências atribuídas a esta espécie representou 61% do total da amostra Actinopterygii, em COB, 7%, em 12S, 26% e, em 16S, 6,7%.

Como observação pessoal, acredita-se que a taxa de liberação de DNA possa estar correlacionada não somente com a abundância/biomassa, mas também, associada as características fisiológicas e comportamentais dos animais.

Anterior a coleta das amostras, as espécies foram aclimatadas e suas características comportamentais puderam ser observadas e, desta forma, arbitrariamente classificadas (Tabela 5). Obviamente estas suposições aplicam-se somente aos aspectos identificados dentro da esfera

da comunidade simulada. No entanto, os padrões de movimento, comportamento e, natureza do tecido podem explicar a baixa taxa de liberação de DNA para algumas das espécies não identificadas.

Pseudanos, *Pygocentryus*, *Serrasalmus*, *Apteronotus*, *Acaronia*, *Apistogramma* e, *Pterophylum* tiveram DNA detectado e são considerados ativos e agressivos entre si. Ao contrário, *Crenicicla* considerada também agressiva, quando em confinamento, assume caráter passivo diante dos outros, já que se torna extremamente territorialista tendendo a não se movimentar muito dentro dos aquários.

Dentre os animais considerados ativos, porém, de característica dócil, não entrando em disputas agressivas com os demais ao seu redor, apenas *Moenkausia* não foi identificada e, dentre os hipoativos que, compreende todos os Siluriformes, apenas *Platistacus* não foi identificado.

Avaliando estas duas observações, foi notado que dentre as nove espécies não identificadas, todas possuem escamas flexíveis e/ou ausência das mesmas (tecido do tipo “pele”). Talvez essa característica possa influenciar na taxa de degradação do DNA^{42,43}, já que, como já foi demonstrado, altas temperaturas, tais como as encontradas na região de Belém, onde o experimento foi realizado, pode aumentar a taxa de degradação do DNA por ação de bactérias e nucleases^{98,99}. *Apteronotus* possui também o mesmo tipo de tecido epidérmico classificado como ausência de escamas, mas sua natureza agressiva com os demais ao seu redor, em constante disputa por território faz com que aos animais sejam constantemente lesionados e, portanto, liberem maiores quantidades de tecido na água. Ainda que seja apenas uma hipótese, a abundância de eDNA depender de comportamentos agressivos entre indivíduos de uma mesma espécie em princípio poderia ser testada com um desenho experimental mais adequado no futuro.

Tabela 5 – Classificação arbitrária das espécies da comunidade simulada quanto a; MOB=Mobilidade, TEMP= Temperamento solitário, TEMP_GRUPO=Temperamento em grupo, TECIDO=Tipo de tecido epidérmico, RESIST_TECIDO=Resistência do tecido. N=Número total de espécies; NSED=Número total de sequências consideradas para a espécie; * Espécie não identificada, porém, selecionada como candidata. + Espécie não identificada.

ESPÉCIE	N	NSEQ	MOB	TEMP	TEMP GRUPO	TECIDO	RESIST TECIDO	LISTA
<i>Apistogramma agassizii</i>	818	13	Ativo	Agressivo	Recluso	Escama	Flexível	
<i>Apistogramma rupununi</i>	652	4	Ativo	Agressivo	Recluso	Escama	Flexível	
<i>Hyphessobrycon bentosi</i>	503	0	Ativo	Dócil	Recluso	Escama	Flexível	*
<i>Gnathocharax steindachneri</i>	502	2	Ativo	Dócil	Recluso	Escama	Flexível	
<i>Heterocharax virgulatus</i>	451	15	Ativo	Dócil	Recluso	Escama	Flexível	
<i>Apteronotus albifrons</i>	107	93	Hipoativo	Agressivo	Atacante	Pele	Flexível	
<i>Laetacara dorsigera</i>	96	2	Ativo	Dócil	Recluso	Escama	Dura	
<i>Apistogramma gephyra</i>	50	4	Ativo	Agressivo	Recluso	Escama	Flexível	
<i>Pterophyllum scalare</i>	25	2	Ativo	Agressivo	Atacante	Escama	Dura	
<i>Moenkhausia oligolepis</i>	15	0	Ativo	Dócil	Recluso	Escama	Flexível	+
<i>Guianacara dacrya</i>	10	2	Ativo	Dócil	Recluso	Escama	Dura	
<i>Myloplus rubripinnis</i>	10	2	Ativo	Dócil	Recluso	Escama	Dura	
<i>Crenicichla johanna</i>	67	0	Hipoativo	Agressivo	Recluso	Escama	Flexível	+
<i>Hemigrammus rhodostomus</i>	60	3	Hiperativo	Dócil	Recluso	Escama	Flexível	
<i>Gasteropelecus levis</i>	51	1	Ativo	Dócil	Recluso	Escama	Flexível	
<i>Hyphessobrycon heterorhabdus</i>	51	1	Hiperativo	Dócil	Recluso	Escama	Flexível	
<i>Corydoras ephippifer</i>	30	2	Hipoativo	Dócil	Recluso	Escama	Dura	
<i>Corydoras julii</i>	30	1	Hipoativo	Dócil	Recluso	Placa	Dura	
<i>Chilodus punctatus</i>	20	3	Hipoativo	Dócil	Recluso	Placa	Dura	
<i>Mesonauta acora</i>	19	0	Hipoativo	Dócil	Atacante	Escama	Flexível	*
<i>Pseudanos trimaculatus</i>	15	1	Ativo	Agressivo	Atacante	Escama	Dura	
<i>Heros efasciatus</i>	4	5	Ativo	Dócil	Atacante	Escama	Dura	
<i>Acaronia nassa</i>	2	2	Ativo	Agressivo	Atacante	Escama	Dura	
<i>Astyanax fasciatus</i>	1	0	Ativo	Dócil	Recluso	Escama	Flexível	*
<i>Satanoperca jurupari</i>	1	0	Ativo	Dócil	Recluso	Escama	Flexível	*
<i>Monocirrhus polyacanthus</i>	9	9	Hipoativo	Dócil	Atacante	Escama	Flexível	
<i>Platystacus costatus</i>	6	0	Hipoativo	Dócil	Recluso	Pele	Flexível	+
<i>Farlowella henriquei</i>	5	14	Hipoativo	Dócil	Recluso	Placa	Dura	
<i>Hypostomus faveolus</i>	4	3	Hipoativo	Dócil	Recluso	Placa	Dura	
<i>Serrapinnus kriegi</i>	4	0	Hiperativo	Dócil	Recluso	Escama	Flexível	*
<i>Panaque armbrusteri</i>	2	4	Hipoativo	Dócil	Recluso	Placa	Dura	
<i>Peckoltia vittata</i>	1	11	Hipoativo	Dócil	Recluso	Placa	Dura	
<i>Pygocentrus nattereri</i>	1	48	Hiperativo	Agressivo	Atacante	Escama	Dura	
<i>Serrasalmus marginatus</i>	1	1	Ativo	Agressivo	Atacante	Escama	Dura	

As análises comparativas entre a técnica de eDNA e a lista de espécie do rio Aporema mostraram que as abordagens realizadas para padronização das amostras de água foram adequadas e com sucesso replicadas em ambiente natural.

Houve recuperação de 22 dos 84 gêneros descritos para a região. Mesmo que *a priori*, o resultado não se mostre tão satisfatório, com apenas 4,7% das espécies sendo identificadas, vale ressaltar que o “checkpoint” final ocorreu às cegas, sem que amostras de tecido tivessem sido coletadas e um banco referencial elaborado.

O estado do Amapá é um dos que compõem a Amazônia legal com menor número de publicações científicas nesta área, com a maioria de suas espécies ainda sem correta identificação morfológica e pouquíssima informação molecular em bases de dados como o Genbank. Desta forma, acredita-se que os resultados possam ser promissores, além de que, apenas dois *primers* foram testados.

Avaliando um estudo realizado na Guiana Francesa (região próxima ao estado do Amapá)⁶⁷, também utilizando o gene 12S com um tamanho similar de fragmento, observou-se que, assim como demonstrado na comunidade simulada, também houve redução no número de espécies identificadas das ordens Characiformes e Siluriformes. Em nosso estudo, foi possível identificar 28% das espécies de Characiformes na comunidade simulada, enquanto no estudo de Cilleros et al (2018), o sucesso amostral foi de 49% para este grupo. Dentre os Siluriformes, para a comunidade simulada em nosso estudo, 29% das espécies puderam ser identificadas, enquanto no referido estudo foram identificadas 38% delas. Considerando os Perciformes, o *primer* 12S obteve melhor desempenho em ambos os estudos, sendo que na comunidade simulada foram identificadas 41% das espécies deste grupo e na Guiana Francesa houve recuperação de 81%. Outro ponto importante é que é esperado que ocorram os gêneros *Serrasalmus* em ambos os locais, *Pygocentrus* no Amapá e *Pygopristi* na Guiana Francesa, mas, corroborando os resultados em nossa comunidade simulada, o *primer* 12S não obteve resolução suficiente para distinção entre as duas espécies, e apenas com o uso do gene 16S foi possível a identificação desses gêneros tanto em aquários como em vida livre.

No entanto vale destacar que, ao consultar os pesquisadores do IEPA, 18 novos gêneros foram considerados como de possível ocorrência e, ainda não capturados por técnicas tradicionais. Desta forma, o uso de eDNA auxiliou de forma complementar, sendo um indicativo de que a biodiversidade da área ainda está longe de ser totalmente conhecida.

7 CONCLUSÕES

Por se tratar de um estudo experimental, com a finalidade de avaliar os prós e contras da técnica de eDNA voltada a identificação de peixes neotropicais, pode-se concluir que:

- Diferentes protocolos de sequenciamento podem influenciar nas taxas de recuperação de DNA das espécies alvos reduzindo consideravelmente o número de sequências válidas.
- O uso concomitante de quatro diferentes *amplicons* permitiu aumentar o número de espécies identificadas pela técnica de eDNA.
- A abordagem de comparação entre dois bancos de dados possibilitou a redução de inclusão de resultados falso-positivos e/ou negativos.
- O *primer* COA se mostrou a melhor opção dentre os demais utilizados, identificando 13 das 34 espécies do sistema de aquário, ainda que não tenha diferenciado com exatidão algumas espécies congêneres.
- O tamanho do *amplicom* influenciou diretamente na identificação das espécies, indicando que sequências com menos de 300pb não são suficientes para amostrar a diversidade de peixes amazônicos.
- Tanto o número de indivíduos como suas características fisiológicas e comportamentais podem influenciar na quantidade de material genético presente em uma amostra de eDNA.
- O protocolo utilizado de forma experimental pode recuperar a informação de algumas espécies de peixes em ecossistemas naturais, indicando que a técnica de eDNA pode ser utilizada associada as demais técnicas tradicionais de inventário e monitoramento das espécies de peixes amazônicos.

6 REFERÊNCIAS BIBLIOGRÁFICAS

1. Darwall, W. *et al.* The *Alliance for Freshwater Life* : A global call to unite efforts for freshwater biodiversity science and conservation: The *Alliance for Freshwater Life* : A global call to unite efforts for freshwater biodiversity science and conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems* **28**, 1015–1022 (2018).
2. Dudgeon, D. *et al.* Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews* **81**, 163 (2006).
3. Holmlund, C. M. & Hammer, M. Ecosystem services generated by fish populations. *Ecological Economics* **29**, 253–268 (1999).
4. Van Der Laan, R., Eschmeyer, W. N. & Fricke, R. Family-group names of Recent fishes. *Zootaxa* **3882**, 1 (2014).
5. Reis, R. E. Conserving the freshwater fishes of South America: Conserving South American Freshwater Fishes. *International Zoo Yearbook* **47**, 65–70 (2013).
6. Ferraris, C. J. Phylogeny and Classification of Neotropical Fishes. *Copeia* **2000**, 314–315 (2000).
7. Reis, R. E. *et al.* Fish biodiversity and conservation in South America. *Journal of Fish Biology* **89**, 12–47 (2016).
8. Fitzgerald, D. B. *et al.* Diversity and community structure of rapids-dwelling fishes of the Xingu River: Implications for conservation amid large-scale hydroelectric development. *Biological Conservation* **222**, 104–112 (2018).
9. Winemiller, K. O. *et al.* Balancing hydropower and biodiversity in the Amazon, Congo, and Mekong. *Science* **351**, 128–129 (2016).
10. Fearnside, P. M. Tropical dams: To build or not to build? *Science* **351**, 456–457 (2016).
11. Pelayo-Villamil, P. *et al.* Completeness of national freshwater fish species inventories around the world. *Biodiversity and Conservation* **27**, 3807–3817 (2018).
12. Lees, A. C., Peres, C. A., Fearnside, P. M., Schneider, M. & Zuanon, J. A. S. Hydropower and the future of Amazonian biodiversity. *Biodiversity and Conservation* **25**, 451–466 (2016).
13. Castello, L. & Macedo, M. N. Large-scale degradation of Amazonian freshwater ecosystems. *Global Change Biology* **22**, 990–1007 (2016).
14. Cao, Y., Larsen, D. P. & Hughes, R. M. Evaluating sampling sufficiency in fish assemblage surveys: a similarity-based approach. *Canadian Journal of Fisheries and Aquatic Sciences* **58**, 1782–1793 (2001).

15. Angermeier, P. L. & Smogor, R. A. Estimating number of species and relative abundances in stream-fish communities: effects of sampling effort and discontinuous spatial distributions. *Canadian Journal of Fisheries and Aquatic Sciences* **52**, 936–949 (1995).
16. BC Environment & Resources Inventory Committee (Canada). *Fish collection methods and standards*. (Ministry of Resources Inventory Committee, 1997).
17. Stat, M. *et al.* Combined use of eDNA metabarcoding and video surveillance for the assessment of fish biodiversity: Fish Surveying. *Conservation Biology* (2018). doi:10.1111/cobi.13183
18. Ficetola, G. F., Miaud, C., Pompanon, F. & Taberlet, P. Species detection using environmental DNA from water samples. *Biology Letters* **4**, 423–425 (2008).
19. Ogram, A., Sayler, G. S. & Barkay, T. The extraction and purification of microbial DNA from sediments. *Journal of Microbiological Methods* **7**, 57–66 (1987).
20. Jerde, C. L., Mahon, A. R., Chadderton, W. L. & Lodge, D. M. “Sight-unseen” detection of rare aquatic species using environmental DNA. *Conservation Letters* **4**, 150–157 (2011).
21. Goldberg, C. S., Pilliod, D. S., Arkle, R. S. & Waits, L. P. Molecular Detection of Vertebrates in Stream Water: A Demonstration Using Rocky Mountain Tailed Frogs and Idaho Giant Salamanders. *PLoS ONE* **6**, e22746 (2011).
22. Dejean, T. *et al.* Persistence of Environmental DNA in Freshwater Ecosystems. *PLoS ONE* **6**, e23398 (2011).
23. Reid, G. M., Contreras MacBeath, T. & Csatádi, K. Global challenges in freshwater-fish conservation related to public aquariums and the aquarium industry: Freshwater Fishes and Their Conservation: Global Challenges. *International Zoo Yearbook* **47**, 6–45 (2013).
24. Turner, C. R. *et al.* Particle size distribution and optimal capture of aqueous microbial eDNA. *Methods Ecol Evol* **5**, 676–684 (2014).
25. Willerslev, E. Diverse Plant and Animal Genetic Records from Holocene and Pleistocene Sediments. *Science* **300**, 791–795 (2003).
26. Thomsen, P. F. *et al.* Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples. *PLoS ONE* **7**, e41732 (2012).
27. Bohmann, K. *et al.* Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution* **29**, 358–367 (2014).
28. Biggs, J. *et al.* Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (*Triturus cristatus*). *Biological Conservation* **183**, 19–28 (2015).

29. Moyer, G. R., Díaz-Ferguson, E., Hill, J. E. & Shea, C. Assessing Environmental DNA Detection in Controlled Lentic Systems. *PLoS ONE* **9**, e103767 (2014).
30. Ficetola, G. F. *et al.* Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources* **15**, 543–556 (2015).
31. Ficetola, G. F., Taberlet, P. & Coissac, E. How to limit false positives in environmental DNA and metabarcoding? *Molecular Ecology Resources* **16**, 604–607 (2016).
32. Banks, J. C., Demetras, N. J., Hogg, I. D. & West, D. W. Monitoring brown trout (*Salmo trutta*) eradication in a wildlife sanctuary using environmental DNA. *13* (2016).
33. Takahara, T., Minamoto, T. & Doi, H. Effects of sample processing on the detection rate of environmental DNA from the Common Carp (*Cyprinus carpio*). *Biological Conservation* **183**, 64–69 (2015).
34. Turner, C. R., Uy, K. L. & Everhart, R. C. Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation* **183**, 93–102 (2015).
35. McKee, A. M., Spear, S. F. & Pierson, T. W. The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biological Conservation* **183**, 70–76 (2015).
36. Foote, A. D. *et al.* Investigating the Potential Use of Environmental DNA (eDNA) for Genetic Monitoring of Marine Mammals. *PLoS ONE* **7**, e41781 (2012).
37. Collins, R. A., Armstrong, K. F., Holyoake, A. J. & Keeling, S. Something in the water: biosecurity monitoring of ornamental fish imports using environmental DNA. *Biological Invasions* **15**, 1209–1215 (2013).
38. Davy, C. M., Kidd, A. G. & Wilson, C. C. Development and validation of environmental DNA (eDNA) markers for detection of freshwater turtles. *PloS one* **10**, e0130965 (2015).
39. Eichmiller, J. J., Miller, L. M. & Sorensen, P. W. Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. *Molecular Ecology Resources* **16**, 56–68 (2016).
40. Hunter, M. E. *et al.* Environmental DNA (eDNA) Sampling Improves Occurrence and Detection Estimates of Invasive Burmese Pythons. *PLOS ONE* **10**, e0121655 (2015).
41. Jane, S. F. *et al.* Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Mol Ecol Resour* **15**, 216–227 (2015).
42. Strickler, K. M., Fremier, A. K. & Goldberg, C. S. Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biological Conservation* **183**, 85–92 (2015).

43. Eichmiller, J. J., Best, S. E. & Sorensen, P. W. Effects of Temperature and Trophic State on Degradation of Environmental DNA in Lake Water. *Environmental Science & Technology* **50**, 1859–1867 (2016).
44. Evans, N. T. *et al.* Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Mol Ecol Resour* **16**, 29–41 (2016).
45. Pilliod, D. S., Goldberg, C. S., Arkle, R. S. & Waits, L. P. Factors influencing detection of eDNA from a stream-dwelling amphibian. *Mol Ecol Resour* **14**, 109–116 (2014).
46. Forsström, T. & Vasemägi, A. Can environmental DNA (eDNA) be used for detection and monitoring of introduced crab species in the Baltic Sea? *Marine Pollution Bulletin* **109**, 350–355 (2016).
47. Takahara, T., Minamoto, T., Yamanaka, H., Doi, H. & Kawabata, Z. Estimation of Fish Biomass Using Environmental DNA. *PLoS ONE* **7**, e35868 (2012).
48. Takahara, T., Minamoto, T. & Doi, H. Using Environmental DNA to Estimate the Distribution of an Invasive Fish Species in Ponds. *PLoS ONE* **8**, e56584 (2013).
49. Goldberg, C. S. *et al.* Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution* **7**, 1299–1307 (2016).
50. Freeland, J. R. The importance of molecular markers and primer design when characterizing biodiversity from environmental DNA. *Genome* **60**, 358–374 (2017).
51. Minamoto, T. *et al.* Nuclear internal transcribed spacer-1 as a sensitive genetic marker for environmental DNA studies in common carp *Cyprinus carpio*. *Mol Ecol Resour* **17**, 324–333 (2017).
52. Bylemans, J. Monitoring freshwater fish communities with environmental DNA (eDNA) metabarcoding. (2018). doi:10.13140/rg.2.2.32976.15364
53. Cristescu, M. E. From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution* **29**, 566–571 (2014).
54. Marquina, D., Andersson, A. F. & Ronquist, F. New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. *Molecular Ecology Resources* (2018). doi:10.1111/1755-0998.12942
55. Li, F. *et al.* Application of Environmental DNA Metabarcoding for Predicting Anthropogenic Pollution in Rivers. *Environmental Science & Technology* (2018). doi:10.1021/acs.est.8b03869

56. Ardura, A. *et al.* eDNA and specific primers for early detection of invasive species – A case study on the bivalve *Rangia cuneata*, currently spreading in Europe. *Marine Environmental Research* **112**, 48–55 (2015).
57. Cannon, M. V. *et al.* In silico assessment of primers for eDNA studies using PrimerTree and application to characterize the biodiversity surrounding the Cuyahoga River. *Scientific Reports* **6**, (2016).
58. Bronnenhuber, J. E. & Wilson, C. C. Combining species-specific COI primers with environmental DNA analysis for targeted detection of rare freshwater species. *Conservation Genetics Resources* **5**, 971–975 (2013).
59. Goldberg, C. S., Sepulveda, A., Ray, A., Baumgardt, J. & Waits, L. P. Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). *Freshwater Science* **32**, 792–800 (2013).
60. Piaggio, A. J. *et al.* Detecting an elusive invasive species: a diagnostic PCR to detect Burmese python in Florida waters and an assessment of persistence of environmental DNA. *Mol Ecol Resour* **14**, 374–380 (2014).
61. Egan, S. P. *et al.* Rapid Molecular Detection of Invasive Species in Ballast and Harbor Water by Integrating Environmental DNA and Light Transmission Spectroscopy. *Environmental Science & Technology* **49**, 4113–4121 (2015).
62. Ikeda, K., Doi, H., Tanaka, K., Kawai, T. & Negishi, J. N. Using environmental DNA to detect an endangered crayfish *Cambaroides japonicus* in streams. *Conservation Genetics Resources* **8**, 231–234 (2016).
63. Thomsen, P. F. *et al.* Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology* **21**, 2565–2573 (2012).
64. Janosik, A. M. & Johnston, C. E. Environmental DNA as an effective tool for detection of imperiled fishes. *Environmental Biology of Fishes* **98**, 1889–1893 (2015).
65. Dhar, B. & Ghosh, S. K. Mini-DNA barcode in identification of the ornamental fish: A case study from Northeast India. *Gene* **627**, 248–254 (2017).
66. Valdez-Moreno, M. *et al.* Using eDNA to biomonitor the fish community in a tropical oligotrophic lake. (2018). doi:10.1101/375089
67. Cilleros, K. *et al.* Unlocking biodiversity and conservation studies in high-diversity environments using environmental DNA (eDNA): A test with Guianese freshwater fishes. *Molecular Ecology Resources* (2018). doi:10.1111/1755-0998.12900
68. Ivanova, N. V., Zemlak, T. S., Hanner, R. H. & Hebert, P. D. N. Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes* **7**, 544–548 (2007).

69. Miya, M. & Nishida, M. Use of Mitogenomic Information in Teleostean Molecular Phylogenetics: A Tree-Based Exploration under the Maximum-Parsimony Optimality Criterion. *Molecular Phylogenetics and Evolution* **17**, 437–455 (2000).
70. Riaz, T. *et al.* ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research* **39**, e145–e145 (2011).
71. Romano, S. THE SIMPLE FOOL’S GUIDE TO PCR.
72. Rozowsky, J. *et al.* PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nature Biotechnology* **27**, 66–75 (2009).
73. Kress, W. J., García-Robledo, C., Uriarte, M. & Erickson, D. L. DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology & Evolution* **30**, 25–35 (2015).
74. Kocher, T. D. *et al.* Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences* **86**, 6196–6200 (1989).
75. Linhart, C. & Shamir, R. The Degenerate Primer Design Problem: Theory and Applications. *Journal of Computational Biology* **12**, 431–456 (2005).
76. Mioduchowska, M., Czyż, M. J., Gołdyn, B., Kur, J. & Sell, J. Instances of erroneous DNA barcoding of metazoan invertebrates: Are universal *cox1* gene primers too “universal”? *PLOS ONE* **13**, e0199609 (2018).
77. Castresanal, J., Lubben, M., Saraste, M. & Higgins, D. G. Evolution of cytochrome oxidase, an enzyme older than atmospheric oxygen. 10
78. Roe, A. D. & Sperling, F. A. H. Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution* **44**, 325–345 (2007).
79. Rach, J. *et al.* The marker choice: Unexpected resolving power of an unexplored CO1 region for layered DNA barcoding approaches. *PLOS ONE* **12**, e0174842 (2017).
80. Prada, C. F. & Boore, J. L. Gene annotation errors are common in the mammalian mitochondrial genomes database. *BMC Genomics* **20**, (2019).
81. Lv, J. *et al.* Assessment of four DNA fragments (COI, 16S rDNA, ITS2, 12S rDNA) for species identification of the Ixodida (Acari: Ixodida). *Parasites & Vectors* **7**, 93 (2014).
82. Vences, M., Thomas, M., Van der Meijden, A., Chiari, Y. & Vieites, D. R. Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. *Frontiers in zoology* **2**, 5 (2005).

83. Xia, Y. *et al.* COI is better than 16S rRNA for DNA barcoding Asiatic salamanders (Amphibia: Caudata: Hynobiidae): DNA BARCODING OF ASIATIC SALAMANDERS. *Molecular Ecology Resources* **12**, 48–56 (2012).
84. Meyer, C. P. & Paulay, G. DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLoS Biology* **3**, e422 (2005).
85. Rach, J., DeSalle, R., Sarkar, I. ., Schierwater, B. & Hadrys, H. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proceedings of the Royal Society B: Biological Sciences* **275**, 237–247 (2008).
86. Britzke, R. Relações filogenéticas do gênero *Apistogramma* (Teleostei, Cichlidae) e filogeografia da espécie *Apistogramma agassizii*. 195
87. Machado, V. N. *et al.* One thousand DNA barcodes of piranhas and pacus reveal geographic structure and unrecognised diversity in the Amazon. *Scientific Reports* **8**, (2018).
88. Menning, D., Simmons, T. & Talbot, S. Using redundant primer sets to detect multiple native Alaskan fish species from environmental DNA. *Conservation Genetics Resources* (2018). doi:10.1007/s12686-018-1071-7
89. Oliveira, C. *et al.* Phylogenetic relationships within the speciose family Characidae (Teleostei: Ostariophysi: Characiformes) based on multilocus analysis and extensive ingroup sampling. *BMC Evolutionary Biology* **11**, (2011).
90. Kocher, A. *et al.* Evaluation of short mitochondrial metabarcodes for the identification of Amazonian mammals. *Methods in Ecology and Evolution* **8**, 1276–1283 (2017).
91. Shehzad, W. *et al.* Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *undefined* (2012). Available at: /paper/Carnivore-diet-analysis-based-on-next-generation-to-Shehzad-Riaz/227a2015eabbbde1dda37f51ac61aec70d5311c7. (Accessed: 9th February 2019)
92. Li, J. *et al.* Development of an environmental DNA method for monitoring fish communities: ground truthing in diverse lakes with characterised fish faunas. *bioRxiv* (2018). doi:10.1101/394718
93. al, V. E., et. Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/12649486?dopt=Abstract>. (Accessed: 9th February 2019)
94. Mariguela, T. C., Alexandrou, M. A., Foresti, F. & Oliveira, C. Historical biogeography and cryptic diversity in the Callichthyinae (Siluriformes, Callichthyidae). *Journal of Zoological Systematics and Evolutionary Research* n/a-n/a (2013). doi:10.1111/jzs.12029

95. Doi, H. *et al.* Use of Droplet Digital PCR for Estimation of Fish Abundance and Biomass in Environmental DNA Surveys. *PLOS ONE* **10**, e0122763 (2015).
96. Pilliod, D. S., Goldberg, C. S., Arkle, R. S. & Waits, L. P. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences* **70**, 1123–1130 (2013).
97. Stoeckle, M. Y., Soboleva, L. & Charlop-Powers, Z. Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary. *PloS one* **12**, e0175186 (2017).
98. Pilliod, D. S., Goldberg, C. S., Arkle, R. S. & Waits, L. P. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences* **70**, 1123–1130 (2013).
99. Renshaw, M. A., Olds, B. P., Jerde, C. L., McVeigh, M. M. & Lodge, D. M. The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol–chloroform–isoamyl alcohol DNA extraction. *Mol Ecol Resour* **15**, 168–176 (2015).
100. Yamanaka, H. *et al.* A simple method for preserving environmental DNA in water samples at ambient temperature by addition of cationic surfactant. *Limnology* **18**, 233–241 (2017).