RELATÓRIO FINAL

Levantamento da Biodiversidade em Sistemas Agroflorestais

Cliente:

BELTERRA

Contato Bio Bureau:
Juliana Americo
juliana.americo@biobureau.com.br

Versão 1.0

Data:
13/03/2026

SUMÁRIO

SUMÁRIO **2**
APRESENTAÇÃO **3**
MÉTODOS **4**
Coleta de amostras ambientais 4
Extração de DNA ambiental 7
PCR, construção de bibliotecas e sequenciamento 7
Análise Bioinformática: controle de qualidade e classificação taxonômica das sequências 8
RESULTADOS 9
Extração de DNA ambiental 9
PCR, construção de bibliotecas e sequenciamento 9
PCR e Construção de bibliotecas 9
Sequenciamento e bioinformática 10
Composição taxonômica 11
Esforço amostral e profundidade de sequenciamento 14
ESTATÍSTICA ECOLÓGICA 15
Diversidade alfa 15
Composição de comunidades e separação entre usos do solo 16
GLOSSÁRIO 19
MATERIAL SUPLEMENTAR 23
Curvas de rarefação por amostra e marcador genético 23
Análises multivariadas da composição de comunidades em amostras de solo e água 24
Análises multivariadas aplicadas às amostras de solo com SAFs separados em maduro e recente 25
Análises multivariadas aplicadas às amostras de solo excluindo áreas preservadas 26

APRESENTAÇÃO

O objetivo do contrato foi o levantamento da biodiversidade de metazoários e plantas em amostras ambientais (água e solo) de sistemas agroflorestais (SAFs) em diferentes estágios de desenvolvimento em comparação com áreas de reserva legal e áreas degradadas.

Este relatório apresenta as atividades realizadas nos seguintes aspectos:

- **Métodos aplicados:** descrição geral do processo, com ênfase no desenho amostral.

- **Resultados e conclusão**: apresentação dos resultados obtidos com as análises realizadas e conclusões.

Observação:
Os termos técnicos destacados em **negrito e verde** ao longo deste relatório estão definidos no Glossário, disponível ao final do documento.

MÉTODOS

Coleta de amostras ambientais

A coleta das amostras foi realizada pela equipe da Belterra Agroflorestas entre os dias 02 e 03 de setembro de 2024, seguindo um protocolo fornecido pela Bio Bureau para preservar a integridade do **DNA ambiental (eDNA)**.

As coletas ocorreram em reservas legais, sistemas agroflorestais (SAFs) com 2, 3 e 5 anos de implantação e áreas degradadas nas Fazendas Cultrosa, Girassol, Sussuarana e Eva, localizadas no município de Camamu, no estado da Bahia.

A área total amostrada compreende cerca de 200 hectares, distribuídos conforme o tamanho das áreas estudadas:

- Fazenda Cultrosa (22,3 km²): 8 amostras,
- Fazenda Sussuarana (0,562 km²): 5 amostras,
- Fazenda Girassol (0,505 km²): 3 amostras,
- Cachoeira Acaraí (área externa): 2 amostras.

Ao todo, foram coletadas 20 amostras, sendo 5 de água e 15 de solo. A Tabela 1 apresenta o detalhamento completo das amostras, incluindo local de coleta, descrição da área, data e coordenadas geográficas. A classificação das amostras em reserva, degradada, SAF maduro e SAF recente foi realizada com base nas informações fornecidas pela contratante. As amostras de água coletadas dentro das fazendas foram classificadas como “rural” e a amostra de água coletada na cidade de Camamu foi classificada como “urbana”. A Figura 1 ilustra a localização georreferenciada dos pontos amostrados, diferenciando substratos aquosos (círculos azuis) e de solo (círculos amarelos).

Tabela 1. Descrição e informações das amostras de água e solo coletadas para posterior processamento por *metabarcoding*. Para cada amostra, são apresentados o nome das amostras, a fazenda onde estão localizadas, a descrição da área de coleta, a data de amostragem, latitude e longitude. Amostras de água são indicadas pelo sufixo 'AQ' em seu código, amostras de solo, pelo sufixo 'SL'.

Amostra	Fazenda	Classificação da área	Descrição	Data da coleta	Latitude (S)	Longitude (O)
BD040AQ	Cultrosa	Rural	Área de reserva, igarapé sem aparente influência de corrente	02/09/2024	13°56'32,259''	39°11'5,029''
BD041AQ	Cultrosa	Rural	SAF maduro	02/09/2024	13°57'16,305''	39°10'19,143''
BD042AQ	Girassol	Rural	Cachoeira em área de reserva com forte influência de corrente	02/09/2024	13°56'7,015''	39°8'38,197''
BD043AQ	Sussuarana	Rural	Riacho em área de reserva	03/09/2024	14°0'20,858''	39°9'39,141''
BD044AQ	Cachoeira Acaraí	Urbano	Fora da fazenda, lago próximo à cachoeira sem aparente influência da pressão das quedas d'água	03/09/2024	13°58'18,86''	39°8'13,003''
BD045SL	Cultrosa	Preservada	Reserva com presença de serrapilheira que foi retirada no momento da coleta	02/09/2024	13°56'39''	39°11'4,106''
BD046SL	Cultrosa	SAF recente	SAF recente com menos de dois anos de implantação, não está totalmente formado	02/09/2024	13°56'18,149''	39°10'41,892''
BD047SL	Cultrosa	SAF recente	SAF recente com menos de dois anos de implantação, não está totalmente formado	02/09/2024	13°56'38,973''	39°10'51,737''
BD048SL	Cultrosa	SAF maduro	SAF 2 anos	02/09/2024	13°56'59,034''	39°10'46,805''
BD049SL	Cultrosa	SAF maduro	SAF 3 anos	02/09/2024	13°57'21,038''	39°10'41,339''

Amostra	Fazenda	Classificação da área	Descrição	Data da coleta	Latitude (S)	Longitude (O)
BD050SL	Cultrosa	SAF maduro	SAF 5 anos, com serrapilheira retirada no momento da coleta	02/09/2024	13°57'10,978''	39°10'23,647'
BD051SL	Girassol	SAF maduro	SAF maduro	02/09/2024	13°55'52,739''	39°8'32,654''
BD052SL	Girassol	SAF maduro	SAF maduro	02/09/2024	13°56'2,816''	39°8'35,097''
BD053SL	Girassol	SAF maduro	Área florestal dentro da fazenda.	02/09/2024	13°56'6,455''	39°8'37,57''
BD054SL	Girassol	SAF maduro	SAF maduro	02/09/2024	13°56'0,775''	39°8'26,91''
BD055SL	Sussuarana	Reserva	Área de reserva com serrapilheira retirada no momento da coleta.	03/09/2024	14°0'20,535''S	39°9'38,772''
BD056SL	Sussuarana	SAF recente	SAF recente com menos de dois anos de implantação, não está totalmente formado.	03/09/2024	14°0'16,854''S	39°9'38,114''O
BD057SL	Sussuarana	Degradado	Área degradada	03/09/2024	14°0'39,522''S	39°9'22,252''O
BD058SL	Sussuarana	Degradado	Área degradada com vegetação rasteira	03/09/2024	14°0'20,35''S	39°9'22,892''O
BD059SL	Sussuarana	Degradada	Área degradada	03/09/2024	14°0'24,213''S	39°9'9,785''O

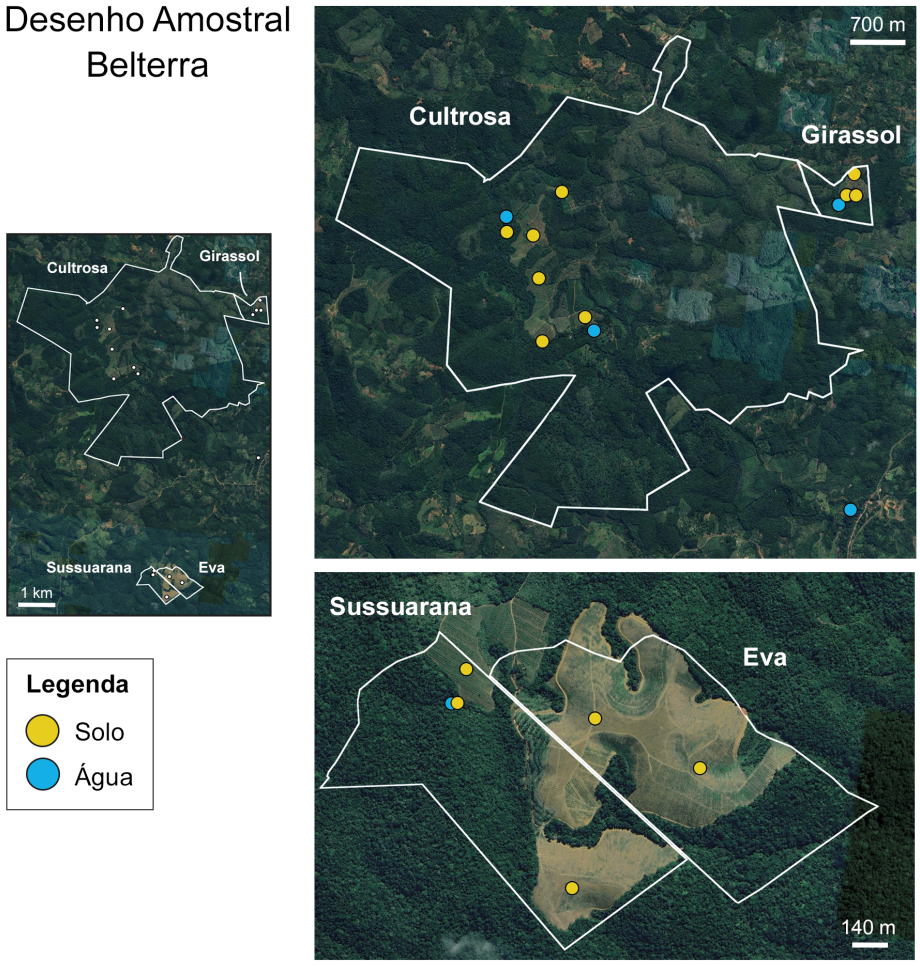


Figura 1. Desenho amostral das coletas realizadas nas fazendas Cultrosa, Girassol, Sussuarana e Eva, além de áreas arredores, para obtenção de amostras de água e solo coletadas para posterior processamento por **metabarcoding**. Para cada ponto de amostragem, as cores indicam o substrato aquoso (círculos azuis) ou de solo (círculos amarelos).

Extração de DNA ambiental

A extração do DNA ambiental foi realizada a partir das amostras de água e solo utilizando kits comerciais adaptados às características de cada substrato.

- Amostras de água: O DNA foi recuperado de filtros Sterivex (Millipore, conforme uma adaptação do protocolo de Miya et al. (2020):
    - Miya M, Gotoh R O & Sado T. (2020). MiFish metabarcoding: a high-throughput approach for simultaneous detection of multiple fish species from environmental DNA and other samples. Fisheries Science, 86(6), 939-970.
- Amostras de solo: foram homogeneizadas e submetidas ao protocolo padrão do kit DNeasy PowerLyzer PowerSoil (Qiagen), que inclui etapas de remoção de impurezas e possíveis inibidores.

Os extratos de DNA resultantes foram purificados e armazenados em condições controladas para garantir sua integridade até as etapas de amplificação e sequenciamento.

PCR, construção de bibliotecas e sequenciamento

As amostras de DNA ambiental extraídas foram enviadas à empresa GoGenetic (Curitiba, PR) para a realização das etapas de amplificação por **PCR**, construção de **bibliotecas** e sequenciamento por **metabarcoding**.

Na etapa de amplificação, foram utilizados sete pares de **primers** específicos para **marcadores genéticos** (COI, 12S rRNA, 16S rRNA, ITS-2 e rbcL), que permitem identificar diferentes grupos taxonômicos, como peixes, vertebrados, mamíferos, plantas e fungos. Destes, três pares de primers foram aplicados exclusivamente às amostras aquáticas para a identificação de peixes. As especificações técnicas completas dos *primers* estão detalhadas na Tabela S1.

Após a amplificação dos marcadores genéticos, os produtos gerados foram purificados, quantificados e utilizados na construção de bibliotecas. No total, foram construídas e sequenciadas 180 bibliotecas, cada uma correspondendo a um dos

marcadores genéticos detectado em uma amostra de DNA ambiental analisada. As bibliotecas foram então sequenciadas na plataforma NextSeq 1000 (Illumina).

Análise Bioinformática: controle de qualidade e classificação taxonômica das sequências

---

Após o sequenciamento, os dados foram analisados para assegurar a qualidade e a confiabilidade dos resultados. Ferramentas específicas (FastQC e MultiQC) foram utilizadas para verificar a integridade e a qualidade das sequências de DNA.

Durante o processamento, **adaptadores** e outras sequências indesejadas geradas no sequenciamento foram removidos, garantindo que apenas dados limpos fossem utilizados nas análises subsequentes. Em seguida, os dados passaram por uma etapa de análise detalhada utilizando o programa QIIME para corrigir possíveis erros de sequenciamento.

As sequências resultantes foram agrupadas em **Unidades Taxonômicas Operacionais (OTUs)**, que reúnem sequências altamente semelhantes (99%) e são amplamente utilizadas em estudos de biodiversidade. As OTUs foram comparadas com bancos de dados genéticos para identificar os organismos presentes e verificar possíveis contaminações (DNA humano e de animais domésticos). Os nomes comuns de cada organismo foram obtidos a partir dos bancos de dados GBIF e iNaturalist, priorizando nomes em português e, quando não encontrados, utilizando os em inglês.

As **OTUs** foram então utilizadas para o cálculo dos índices de diversidade alfa, como riqueza e índice de shannon, e análises de estatística multivariada.

RESULTADOS

Extração de DNA ambiental

O DNA ambiental foi extraído inicialmente das 20 amostras, sendo 15 de solo e 5 de água. A reextração de DNA de 15 amostras de solo foi realizada para o ressequenciamento. No entanto, para as 5 amostras aquáticas originais, todas as amostras (filtros) foram consumidas anteriormente e não foi possível repetir a extração. Todas as amostras apresentaram DNA amplificável.

PCR, construção de bibliotecas e sequenciamento

PCR e Construção de bibliotecas

Foram sequenciadas 55 bibliotecas de solo e água, representando áreas de reserva, SAF maduro, SAF recente e degradadas. A Tabela S2 mostra o resultado do sequenciamento de DNA de todas as bibliotecas, incluindo aquelas nas quais não houve sucesso na amplificação. A ausência de amplificação do DNA persistente após a repetição pode ser devido à ausência de espécies do grupo-alvo na amostra ou inibição da reação de amplificação devido a compontentes do solo não removidos durante a extração de DNA.

Para as amostras de solo, o desenho experimental contou com amostras do mesmo tipo de ambiente, por exemplo, duas amostras de solo de reserva de uma mesma fazenda. Além disso, para *primers* com conhecida dificuldade de amplificação do DNA, houve redundância de *primers*, como ITS e rbcl que amplificam plantas. Portanto, para construção da lista final de espécies, não houve repetição caso houvesse redundância amostral ou de *primers*.

Plantas e fungos

Das 20 bibliotecas amplificadas com *primers* para o marcador **ITS2**, para plantas e fungos, 16 tiveram sequenciamento bem-sucedido e para o marcador **rbcL**, também direcionados a plantas, 13 tiveram sequenciamento bem-sucedido.

Para 3 amostras, 2 de SAF maduro e 1 área degrada, houve redundância de grupo amostral, para 3 amostras, 2 de reserva e 1 de área degradada, houve redundância de primers (**ITS** e **rbcl**) e para 1 amostra, de área urbana, não houve sucesso na amplificação após a repetição.

Peixes

Todas as bibliotecas construídas com marcadores genéticos direcionados para peixes (**12Smifish**, **16sUniversal** e **COI**), executadas apenas nas amostras de água, tiveram o sequenciamento bem-sucedido.

Vertebrados e Mamíferos

A amplificação com os *primers* para os marcadores 12S Vertebrados e 16S para mamíferos foi repetida para as 16 amostras de solo e não houve sucesso. Foram usadas na análise final OTUs das amostras originais. São dados de 2 amostras de Reserva, 1 de SAF recente, 3 de SAF maduro e 2 de área degrada (amostras destacadas em verde na tabela S2).

Sequenciamento e bioinformática

Foram geradas 21.709.637 de leituras de DNA brutas e 7.420.489 limpas. A tabela S3 apresenta as principais métricas de sequenciamento, incluindo o número total de leituras, leituras retidas após a filtragem de qualidade, porcentagem de leituras que passaram pelo filtro, leituras com redução de ruído, leituras mescladas e sequências não quiméricas. A porcentagem de sequências não quiméricas reflete a proporção de leituras de alta qualidade disponíveis para classificação taxonômica. Devido aos requisitos de input do programa de classificação de OTUs, foram excluídas das análises bibliotecas com menos de 20.000 leituras após filtragem de qualidade.

Composição taxonômica

Após processamento dos dados, 304 OTUs foram detectadas, pertecentes a 10 filos, 26 classes, 86 ordens, 140 famílias, 180 gêneros e 97 OTUs classificadas a nível de espécie. Essas OTUs foram organizadas em duas tabelas complementares. A Tabela S4 apresenta a matriz de **frequência** de **OTUs** por amostra, utilizada para todas as análises de diversidade, composição e agrupamento. Já a Tabela S5 traz informações de distribuição, *status* de conservação segundo a IUCN, e classificação quanto à origem (nativa, exótica, invasora) e ocorrência no Brasil.

Um total de 131 OTUs possuem classificação na Lista Vermelha da União Internacional para a Conservação da Natureza e dos Recursos Naturais (**IUCN**) das espécies ameaçadas:

- Vulnerável: 01 (Anta, *Tapirus* sp.)
- Pouco Preocupante: 102
- Não Avaliado: 27

Em relação à origem, entre as 97 OTUs classificadas até espécies, 56 são nativas, 16 são exóticas, 4 são domésticas como boi (*Bos taurus*), porco (*Sus scrofa*), cachorro (*Canis lupus*) e para as demais a informação não está disponível. As OTUs identificadas como domésticas ou de origem claramente antrópica, foram classificadas como contaminações ambientais ou domésticas e excluídas das análises ecológicas principais.

A Figura 2 resume a ocorrência das espécies consideradas mais relevantes, incluindo espécies carismáticas, nativas da fauna e flora regional e *status* de ameaçadas ou vulneráveis segundo a IUCN.

No grupo de vertebrados terrestres, foram detectadas espécies emblemáticas como *Tapirus* sp. (anta), *Dasypus* (tatu), *Bradypus tridactylus* (preguiça) e *Rhinella hoogmoedi* (anfíbio endêmico), refletindo boa cobertura de DNA ambiental em áreas preservadas e SAFs. Entre as plantas, destacam-se espécies nativas como *Cecropia*

pachystachya e *Acalypha*, além de cultivos associados ao uso agroflorestal, como *Musa* (banana), *Manihot esculenta* (mandioca) e *Elaeis guineensis* (dendê).

Nas amostras de água, houve, conforme esperado, predominância de espécies de peixesnativas e de interesse ecológico ou econômico, como *Hoplias malabaricus* (traira), *Astyanax bimaculatus* (lambari), *Gymnotus carapo* (enguia), *Electrophorus electricus* (enguia) e *Potamotrygon* (arraia de água doce).

Essa visualização destaca o valor do método para detectar espécies de importância ecológica, de conservação e de monitoramento, reforçando seu uso como ferramenta de diagnóstico ambiental em paisagens rurais.

Os *heatmaps* completos por grupo taxonômico estão disponíveis no material suplementar e reúnem as **OTUs** detectadas classificadas por grandes grupos biológicos. São apresentados oito *heatmaps*, organizados de acordo com a adequação ecológica e a origem dos organismos detectados. Para facilitar a visualização:

As **OTUs** de algas, anfíbios, bactérias, invertebrados e mamíferos foram agrupadas em uma única figura por grupo, combinando amostras de solo e água.

As **OTUs** de peixes foram apresentadas apenas para as amostras de água, já que esse grupo não foi alvo dos *primers* utilizados em solo.

As plantas foram divididas em dois *heatmaps* — um com amostras de solo e outro com amostras de água — visando melhorar a legibilidade devido à maior quantidade de táxons detectados.

Essa visualização detalhada permite identificar a distribuição e recorrência de cada **OTU** por ambiente e uso do solo, complementando as análises principais e oferecendo um panorama mais completo da composição biológica detectada por DNA ambiental.

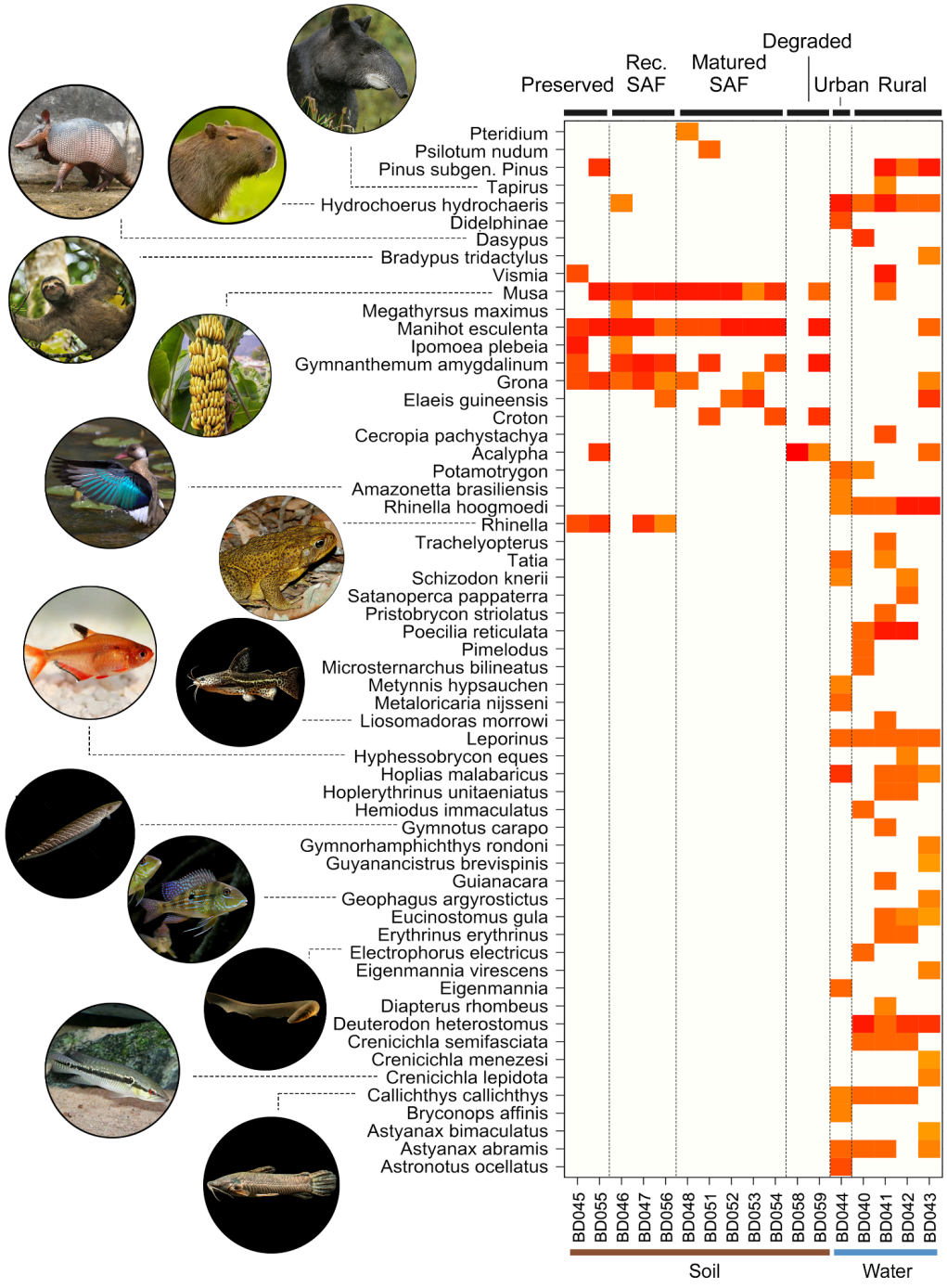


Figura 2. *Heatmap* das espécies mais relevantes detectadas por DNA ambiental nas amostras de solo e água, incluindo espécies nativas, carismáticas, cultivadas e/ou com status de ameaça. A presença é indicada em tons de vermelho, quanto mais vermelho, maior a frequência do DNA. Amostras estão organizadas por tipo de uso do solo.

Esforço amostral e profundidade de sequenciamento

Para avaliar o esforço amostral e a profundidade de sequenciamento de cada biblioteca, foram geradas curvas de rarefação por amostra. Essas curvas indicam a quantidade de espécies (OTUs) detectadas em função do número de leituras de DNA, permitindo verificar se a profundidade de sequenciamento foi suficiente para capturar a diversidade presente nas amostras. As amostras atingiram um platô para todos os *primers*, indicando saturação da curva e, portanto, um esforço amostral adequado (Figura Suplementar S1).

Adicionalmente, foi gerada uma curva de acumulação de espécies por número cumulativo de amostras (Figura 3). Essa análise mostra o crescimento da riqueza de OTUs conforme mais amostras são incluídas na análise, e ajuda a estimar se o número de amostras coletadas foi suficiente para representar a diversidade presente nas diferentes áreas. As amostras analisadas (círculos preenchidos) resultaram em pouco mais de 300 OTUs observados, e a modelagem indica que seriam necessárias cerca de 22 amostras para capturar 90% da diversidade estimada (~550 OTUs).Isso demonstra que o esforço amostral atual foi substancial, embora ainda abaixo da saturação completa, sugerindo que um maior número de amostras deveria ser coletado.

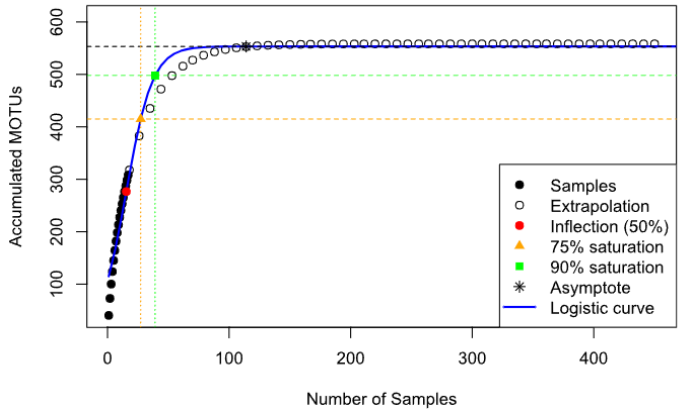


Figura 3. Curva de acumulação de OTUs por número cumulativo de amostras. Círculos preenchidos indicam as amostras reais analisadas; círculos vazados representam extrapolações baseadas em uma curva logística. As amostras atuais resultaram na detecção de pouco mais de 300 OTUs, com

tendência à saturação. A extrapolação indica que a diversidade total estimada poderia atingir cerca de 550–600 OTUs com aumento do esforço amostral.

ESTATÍSTICA ECOLÓGICA

Diversidade alfa

As métricas de diversidade alfa — abundância total de sequências, riqueza (número de OTUs), índice de Shannon e Evenness — foram calculadas individualmente para cada amostra e estão representadas na Figura 3. A média de riqueza e shannon por área estão disponíveis na Tabela S6. Essas métricas permitem caracterizar a estrutura ecológica das comunidades. A abundância representa o total de sequências detectadas em cada amostra e pode estar associada à biomassa ou à atividade biológica local. Quanto maior a abundância, maior o número de registros de DNA. A riqueza indica o número de táxons detectados. Quanto maior a riqueza, mais diversos são os grupos presentes. O índice de Shannon reflete uma comunidade tanto rica quanto equilibrada. Quanto maior o índice de Shannon, mais diversa e equilibrada é a amostra. O evenness mostra o quão uniformemente os táxons estão distribuídos. Quanto maior o evenness, mais homogênea é a distribuição entre os táxons.

As métricas de diversidade alfa apresentam os valores de riqueza e índice de Shannon para cada amostra, organizadas por tipo de uso do solo (Figura 4). Observa-se que as amostras de água provenientes das fazendas apresentam os maiores valores de riqueza e Shannon, refletindo comunidades biologicamente mais diversas, um reflexo do uso de primers específicos para peixes nessas amostras.

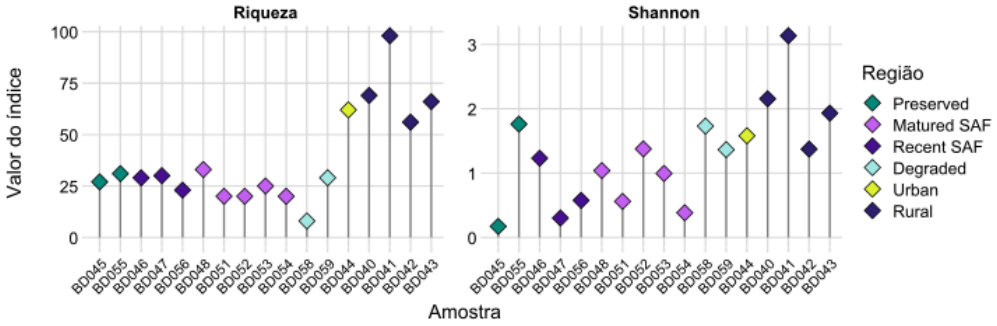


Figura 4. Diversidade alfa por amostra, considerando quatro métricas: Riqueza esquerdo) e Índice de Shannon (direito). Os pontos representam as amostras individuais, coloridas de acordo com a região ou uso do solo.

Entre as amostras de solo, embora as duas amostras da área preservada apresentem riqueza semelhante, seus valores de índice de Shannon são bastante distintos, indicando que a distribuição de abundâncias dos táxons foi mais equilibrada na amostra da Fazenda Cultrosa (BD055) do que na da Fazenda Sussuarana (BD045). Entre os SAFs maduros (BD046 e BD056), há também variação considerável do Shannon, sugerindo heterogeneidade interna nas comunidades. As amostras dos SAFs recentes e área degradada (BD050, BD053, BD058) mostraram valores baixos e consistentes para ambos os índices, refletindo comunidades menos diversas e com dominância de espécies mais evidente.

Esses padrões indicam que a riqueza e a diversidade não são apenas função do número de espécies, mas também da equitatividade, e ambas estão associadas ao uso e histórico do solo. Áreas menos impactadas tendem a sustentar comunidades mais equilibradas, enquanto áreas degradadas favorecem dominância de poucos táxons. Isso reforça o uso do DNA ambiental como ferramenta sensível para avaliar o progresso ecológico de SAFs ao longo do tempo.

Composição de comunidades e separação entre usos do solo

Para entender se diferentes tipos de uso do solo — como áreas preservadas, SAFs e áreas degradadas — abrigam comunidades biológicas distintas, utilizamos três ferramentas estatísticas que ajudam a visualizar e comparar a composição dessas comunidades: **Análise de Componentes Principais (PCA)**, **Análise de Coordenadas Principais (PCoA)** e **Análise Discriminante de Componentes Principais (DAPC)**.

Essas análises são úteis porque, em vez de olhar para cada espécie separadamente, elas resumem toda a informação das espécies presentes em cada amostra e mostram como as amostras se agrupam com base em sua similaridade biológica. Além disso, permitem identificar padrões ecológicos complexos e responder de forma objetiva se os ambientes analisados abrigam comunidades distintas ou estão em transição ecológica.

As análises exploratórias mostraram alta dissimilaridade entre as amostras de solo e de água Ao retirar as amostras de água, foi possível observar maior separação entre as amostras de solo indicando que sua inclusão mascarou as diferenças entre áreas de solo (Figuras Suplementares 3, 4 e 5). Diante desse efeito, as análises subsequentes foram conduzidas separadamente para as amostras de solo, permitindo maior sensibilidade na detecção das diferenças entre áreas de uso (degradada, SAFs, reserva).

A PCoA (Figura 5) mostra uma separação clara entre as amostras de reserva, localizadas em um extremo do gráfico, e os SAFs agrupados, com as áreas degradadas distribuídas no meio, sugerindo pouca diferença entre SAFs e área degradada. A dispersão observada nas amostras de SAF sugere que há espécies detectadas nesses ambientes que não estão presentes nas reservas, mas que ainda compartilham similaridade com áreas degradadas. A significância estatística da

separação entre os grupos foi confirmada pelo teste ANOSIM (R^2 = 0,663; p = 0,006).

A PCA (Figura 6) apresentou um padrão semelhante, com as amostras de reserva e áreas degradadas bem separadas, enquanto os SAFs se posicionaram entre elas, formando um gradiente. Amostras de SAF tiveram tendência a se agrupar mais próximas das reservas, com algumas mais próximas das áreas degradadas. Esse padrão sugere que os SAFs estão em processo de transição ecológica, mas ainda não apresentam sinal suficientemente distinto para serem separados como grupo independente.

Por fim, a DAPC (Figura 7), que maximiza a separação entre grupos definidos, mostrou mostrou uma distinção clara entre as amostras de área de reserva e as demais, mas sobreposição entre SAFs e áreas degradadas. Isso indica que, com base na composição das comunidades biológicas, é possível diferenciar amostras degradadas de preservadas com segurança, mas não há ainda separação consistente entre SAF e área degradada.

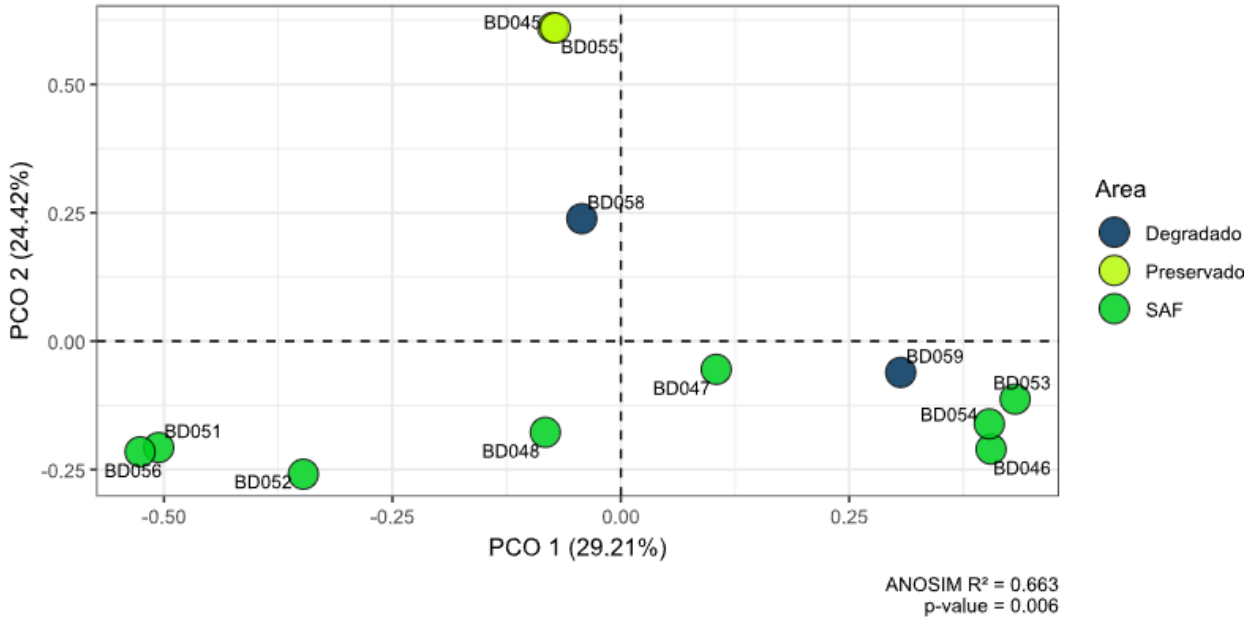


Figura 5. Análise multivariada aplicada às amostras de solo para avaliar a separação entre áreas degradadas, SAFs (maduros e recentes agrupados) e áreas preservadas. PCoA baseada na distância

de Jaccard (presença/ausência) mostra separação significativa entre os grupos, confirmada por ANOSIM (R^2 = 0,663; p = 0,006).

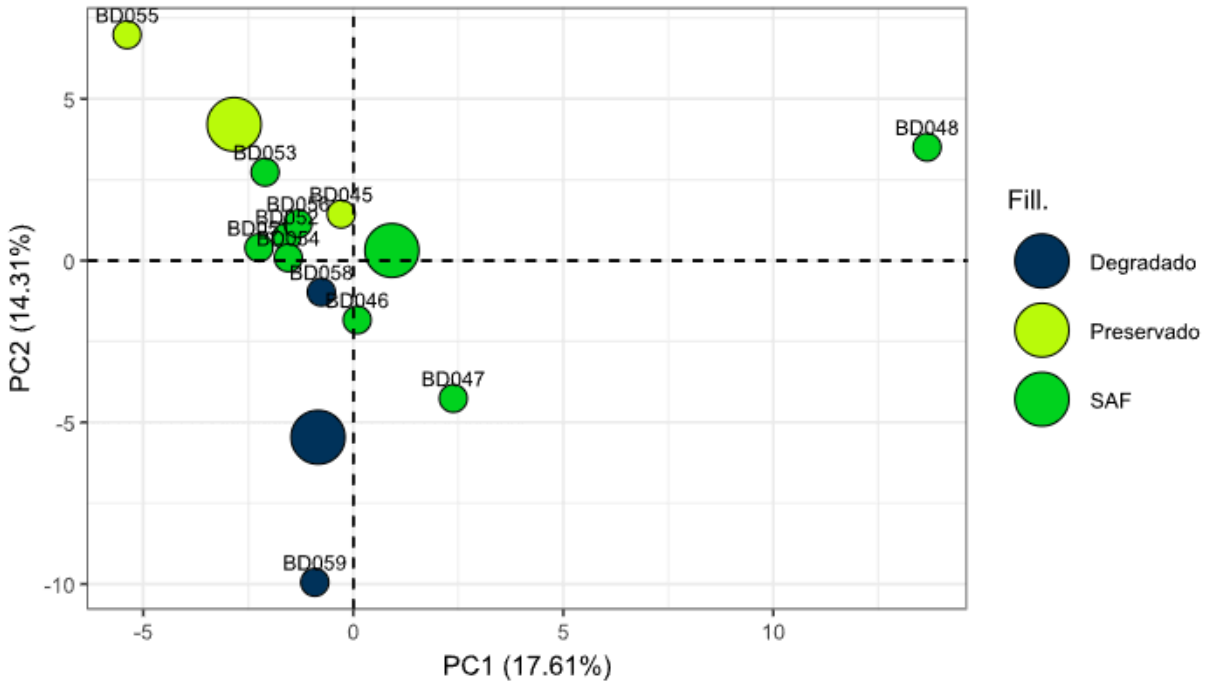


Figura 6. Análise multivariada aplicada às amostras de solo para avaliar a separação entre áreas degradadas, SAFs (maduros e recentes agrupados) e áreas preservadas. PCA com base na abundância relativa das OTUs revela um gradiente de variação entre áreas degradadas e preservadas, com SAFs posicionados entre esses extremos.

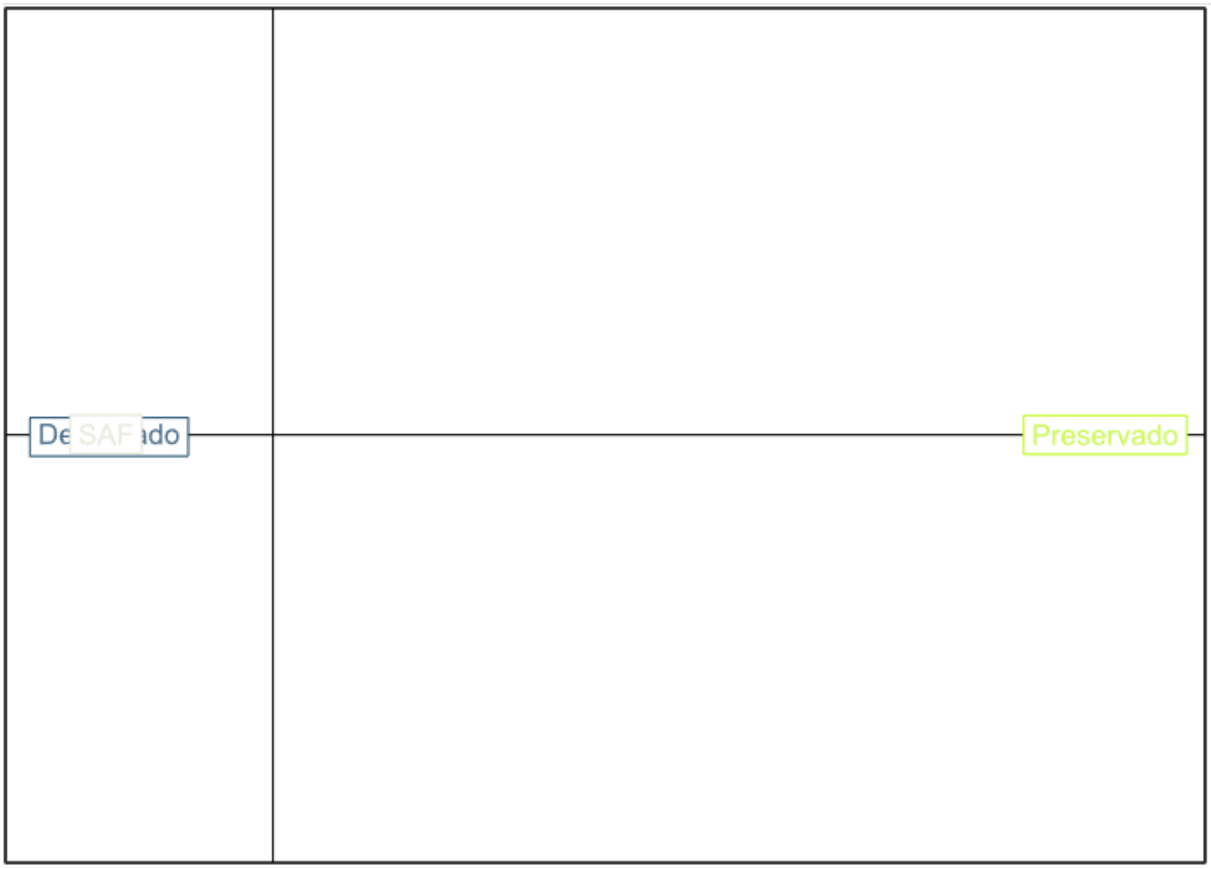


Figura 7. Análise multivariada aplicada às amostras de solo para avaliar a separação entre áreas degradadas, SAFs (maduros e recentes agrupados) e áreas preservadas. DAPC evidencia separação clara entre área preservada e as demais áreas com sobreposição dos SAFs e áreas degradadas, reforçando que ainda não se diferenciam como grupo independente.

As análises com SAF maduro e SAF recente separados estão no material suplementar (Figura Suplementar 5, 6 e 7).

Foram também realizadas as análises excluindo a área preservada, conforme solicitado (Figura Suplementar 8, 9 e 10). A exclusão da área preservada foi proposta para uma possível ampliação da resolução entre SAFs e degradado. Não foi observada diferença estatística entre os grupos (Figura Suplementar 8), houve sobreposição entres os SAFs (Figura Suplementar 9 e 10), sugerindo que a diferenciação entre os SAFs ainda não é possível.

CONCLUSÕES

O levantamento da biodiversidade por DNA ambiental permitiu caracterizar a diversidade de metazoários e plantas em amostras ambientais de água e solo com diferentes usos da terra.

As análises demonstraram que o DNA ambiental, além de uma ferramenta para identificar uma ampla gama de organismos — incluindo espécies nativas, cultivadas, carismáticas e com status de ameaça — é útil para distinguir os diferentes usos do solo com base na composição biológica das comunidades.

As análises multivariadas mostraram uma diferenciação clara entre área degrada e reserva e um gradiente ecológico, com as amostras de SAF posicionadas entre elas. Em especial, observou-se que SAFs maduros tendem a se agrupar mais próximos das áreas preservadas, enquanto SAFs recentes se aproximam das degradadas, indicando que o avanço no estágio de implantação agroflorestal pode ser detectado por meio da composição de DNA ambiental. O padrão de distribuição espacial das amostras reforça a existência de uma transição ecológica mensurável, que pode ser acompanhada ao longo do tempo.

Portanto, o DNA ambiental não apenas permite o levantamento da biodiversidade, mas também fornece suporte técnico e ecológico para monitorar a recuperação de áreas produtivas, como os SAFs, podendo ser integrado a estratégias de restauração ambiental, certificação e planejamento do uso sustentável da terra.

GLOSSÁRIO

Bibliotecas
Bibliotecas são coleções de fragmentos de DNA preparados para sequenciamento em experimentos de nova geração (NGS). Cada biblioteca é composta por produtos amplificados de DNA, correspondentes a um marcador genético, que são purificados, quantificados e organizados para garantir a geração de dados confiáveis durante o processo de sequenciamento

Classificação Taxonômica
A classificação taxonômica consiste na identificação e organização dos organismos em categorias hierárquicas (reino, filo, classe, ordem, família, gênero e espécie) com base em sequências de DNA. Em estudos de DNA ambiental, a comparação das sequências com bancos de dados genéticos permite atribuir as OTUs a um nível taxonômico específico.

DNA Ambiental (eDNA)
O DNA ambiental é o material genético coletado diretamente de ambientes naturais, como água, solo ou ar, sem a necessidade de capturar os organismos presentes. Esse DNA é liberado no ambiente por meio de células, secreções ou fragmentos de organismos, permitindo identificar e monitorar a biodiversidade de forma não invasiva.

Marcadores Genéticos
Marcadores genéticos são sequências específicas de DNA utilizadas para identificar organismos e analisar a diversidade biológica. Em técnicas como o metabarcoding, os marcadores são amplificados por PCR e sequenciados, possibilitando a classificação taxonômica dos organismos presentes nas amostras.

Metabarcoding

O metabarcoding é uma técnica de sequenciamento de DNA que permite identificar múltiplas espécies em uma única amostra ambiental (como solo ou água). A técnica combina a amplificação de marcadores genéticos por PCR com tecnologias de sequenciamento de nova geração (NGS), possibilitando análises rápidas e precisas da diversidade biológica em amostras complexas.

OTU (Unidade Taxonômica Operacional)

OTU é uma unidade utilizada para agrupar organismos semelhantes em análises de dados de sequenciamento. Ela representa uma agrupação de sequências de DNA com alto grau de similaridade (geralmente 97-99%) e serve como um substituto para a identificação de espécies, especialmente em estudos que utilizam DNA ambiental.

PCR (Reação em Cadeia da Polimerase)

PCR é uma técnica laboratorial utilizada para **amplificar segmentos específicos de DNA** a partir de pequenas quantidades de material genético. A amplificação ocorre por meio de **primers** que direcionem a amplificação para regiões-alvo do DNA (marcadores genéticos). É uma etapa essencial para gerar material suficiente para o sequenciamento.

Adaptadores

Adaptadores são sequências artificiais de DNA adicionadas durante a preparação das bibliotecas para o sequenciamento. A remoção adequada dos adaptadores é fundamental para evitar interferências no processamento bioinformático.

Primers

Primers são pequenas sequências de DNA sintético utilizadas como "sondas" no processo de **PCR**. Eles se ligam a regiões-alvo específicas do DNA, em nosso caso, em marcadores genéticos específicos, permitindo a amplificação de segmentos específicos para análise genética.

Leituras (Reads)

Reads são fragmentos de DNA sequenciados que representam as unidades básicas dos dados gerados em experimentos de sequenciamento de nova geração (NGS). A quantidade e a qualidade dos reads são fatores críticos para análises robustas e confiáveis da biodiversidade.

Frequência de OTUs/espécies

Proporção de *reads* de uma OTUs em relação ao total de *reads* de uma amostra.

PCA (Análise de Componentes Principais)

É uma técnica que resume muitas variáveis em dois ou três eixos principais (PCs), que mostram os maiores padrões de variação entre as amostras. Ela é útil para visualizar tendências ou gradientes ecológicos. O objetivo do PCA é revelar padrões naturais nos dados de forma simplificada, sem considerar previamente a qual grupo cada amostra pertence.

PCoA (Análise de Coordenadas Principais)

É uma técnica parecida com o PCA, mas que parte diretamente das diferenças entre as amostras, como as diferenças na presença ou ausência de espécies. Isso a torna especialmente útil em estudos ecológicos, pois permite visualizar o quão diferentes são as comunidades biológicas entre locais. O PCoA gera um gráfico que mostra como as amostras se agrupam com base na composição das espécies detectadas.

DAPC (Análise Discriminante de Componentes Principais)

É uma técnica que combina o PCA com uma etapa de classificação supervisionada. Isso significa que, além de resumir os dados em eixos principais, a DAPC tenta separar grupos previamente definidos (por exemplo: “área preservada” vs. “degradada”) com base na composição das espécies. Ela mostra até que ponto é possível distinguir os ambientes, sendo útil para identificar padrões claros de separação ou sobreposição entre grupos.

MATERIAL SUPLEMENTAR

Curvas de rarefação por amostra e marcador genético

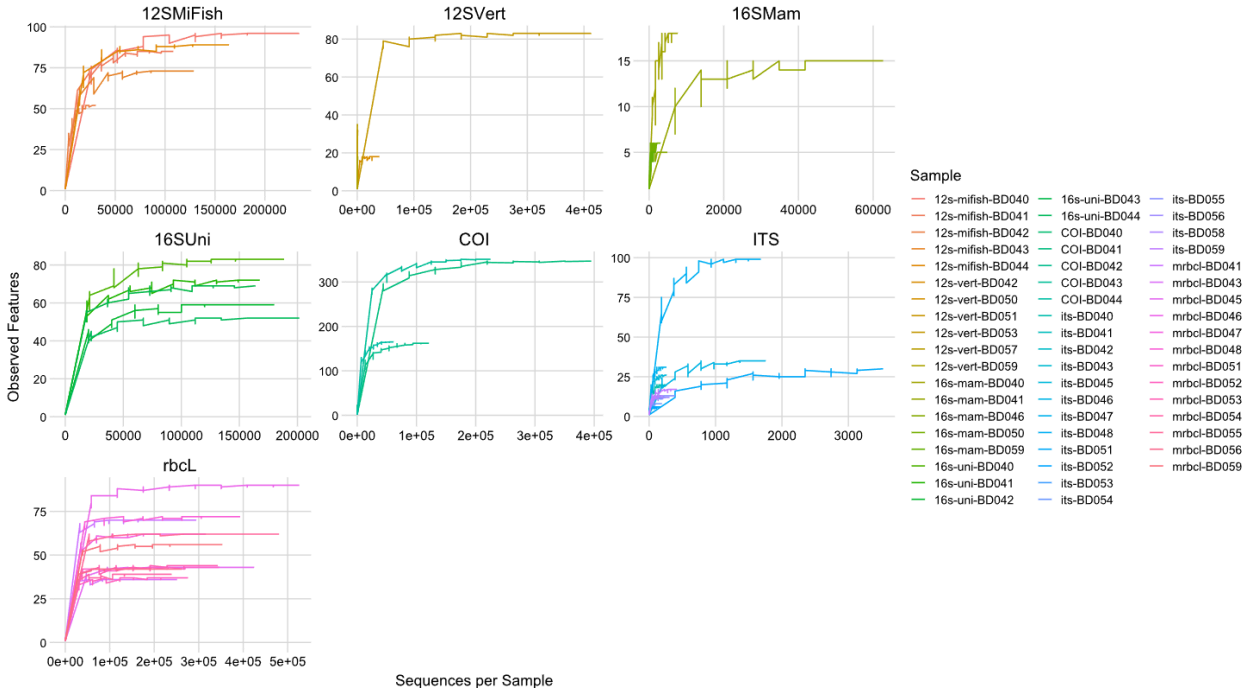


Figura suplementar 1. Curvas de rarefação por amostra e marcador genético, mostrando o número de OTUs acumuladas em função do número de sequências obtidas. As curvas apresentam tendência à saturação, atingido o platô,, indicando suficiência do esforço de sequenciamento.

Análises multivariadas da composição de comunidades em amostras de solo e água

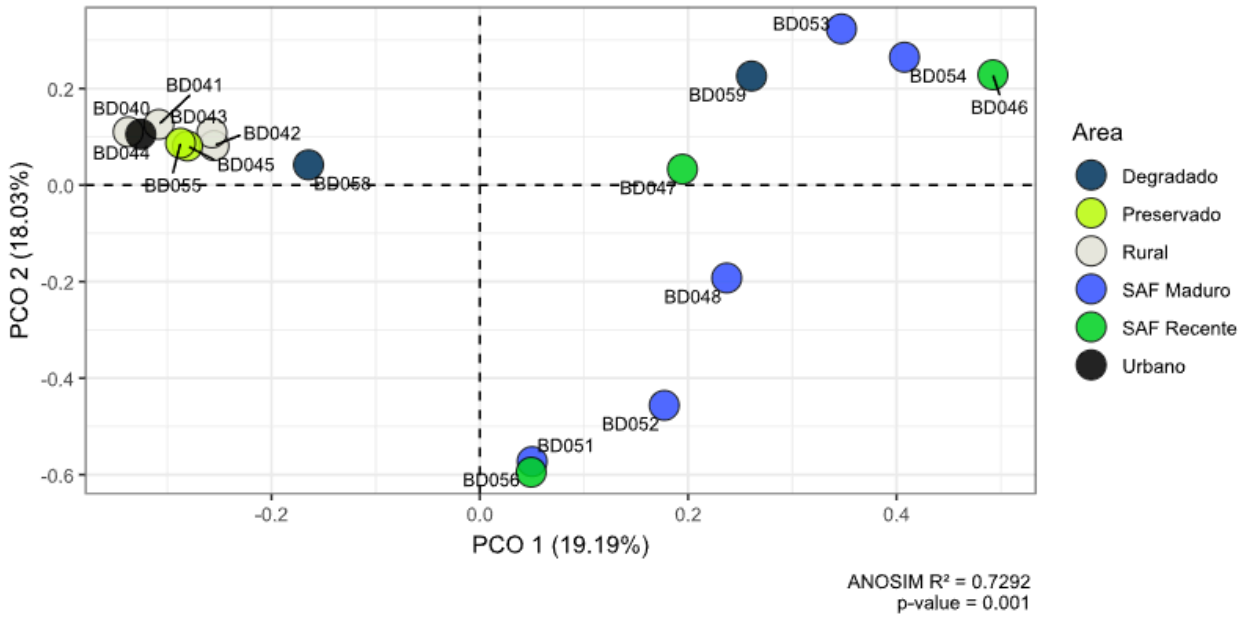


Figura Suplementar 2. Análise multivariada da composição de comunidades em amostras de solo e água, incluindo diferentes usos da terra (área preservada, SAFs, degradada) e água das fazendas (rural) e cachoeira Rio Acarai (urbana). PCoA baseada na distância de Jaccard, considerando presença/ausência de OTUs. As amostras de água (rural e urbana) formam grupos compactos e distintos com alta similaridade interna, enquanto as amostras de solo apresentam maior dispersão, indicando maior heterogeneidade entre os ambientes terrestres. A análise ANOSIM confirma a diferença significativa entre os grupos ($R^2 = 0.7292$, $p = 0.001$).

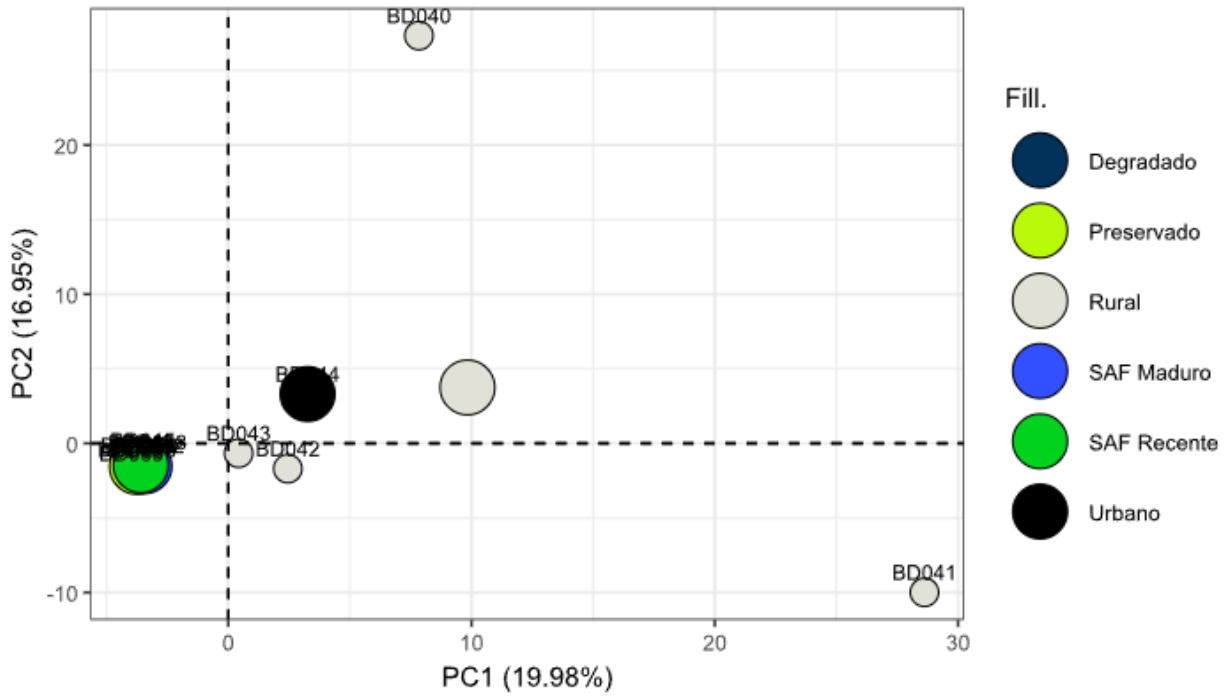


Figura Suplementar 3. Análise multivariada da composição de comunidades em amostras de solo e água, incluindo diferentes usos da terra (área preservada, SAFs, degradada) e água das fazendas (rural) e cachoeira Rio Acarai (urbana). PCA com dados de abundância relativa de OTUs. As solo exibem se sobrepõe e as amostras de água são mais dispersas.

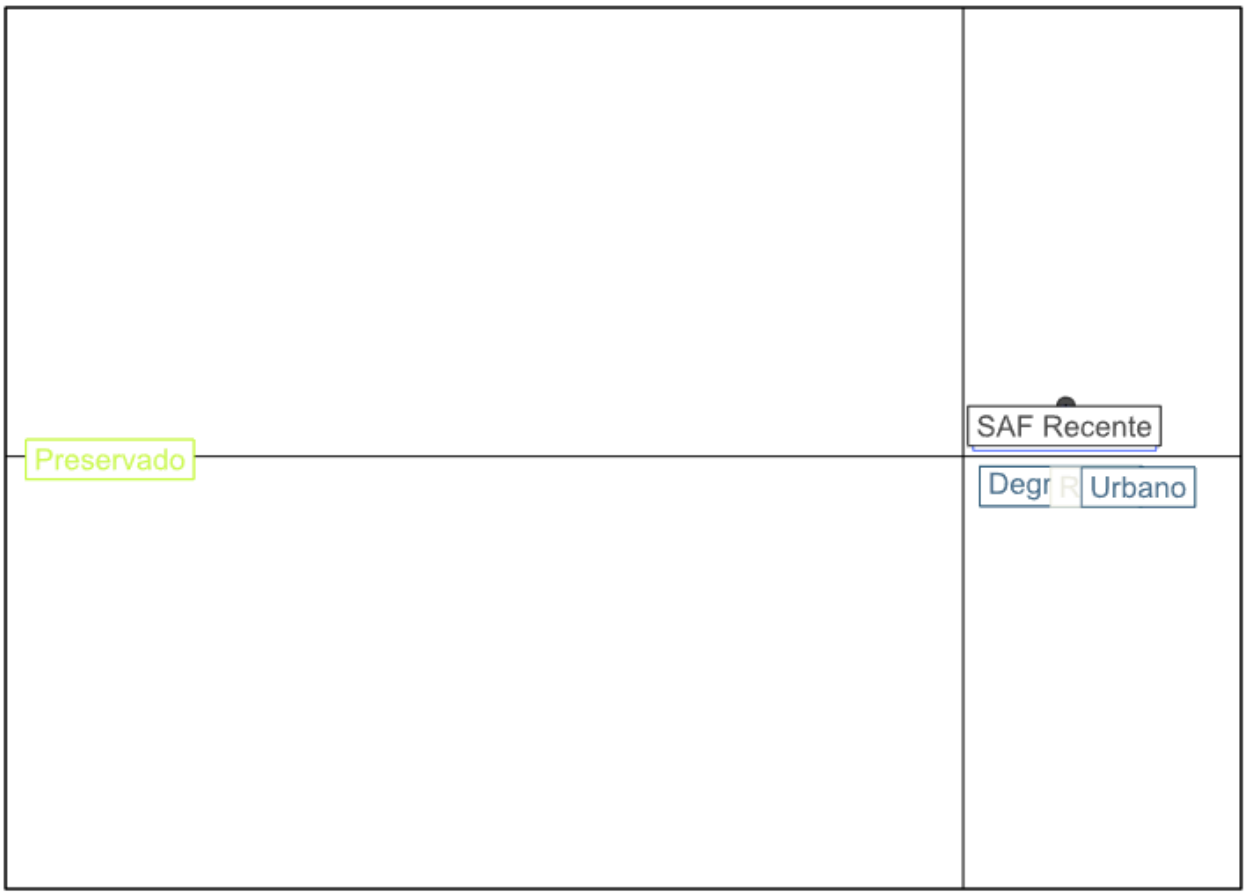


Figura Suplementar 4. Análises multivariadas da composição de comunidades em amostras de solo e água, incluindo diferentes usos da terra (área preservada, SAFs, degradada) e água das fazendas (rural) e cachoeira Rio Acarai (urbana). DAPC baseada em classificação prévia dos grupos de uso do solo e água. A análise mostra sobreposição entre os grupos de solo (SAF e degradado) e água tanto urbana quanto rural. Somente a área de solo preservada manteve-se separada . Conclui-se que a inclusão de amostras aquáticas mascara as diferenças internas entre os usos do solo e, por isso, as demais análises foram realizadas somente com as amostras de solo.

Análises multivariadas aplicadas às amostras de solo com SAFs separados em maduro e recente

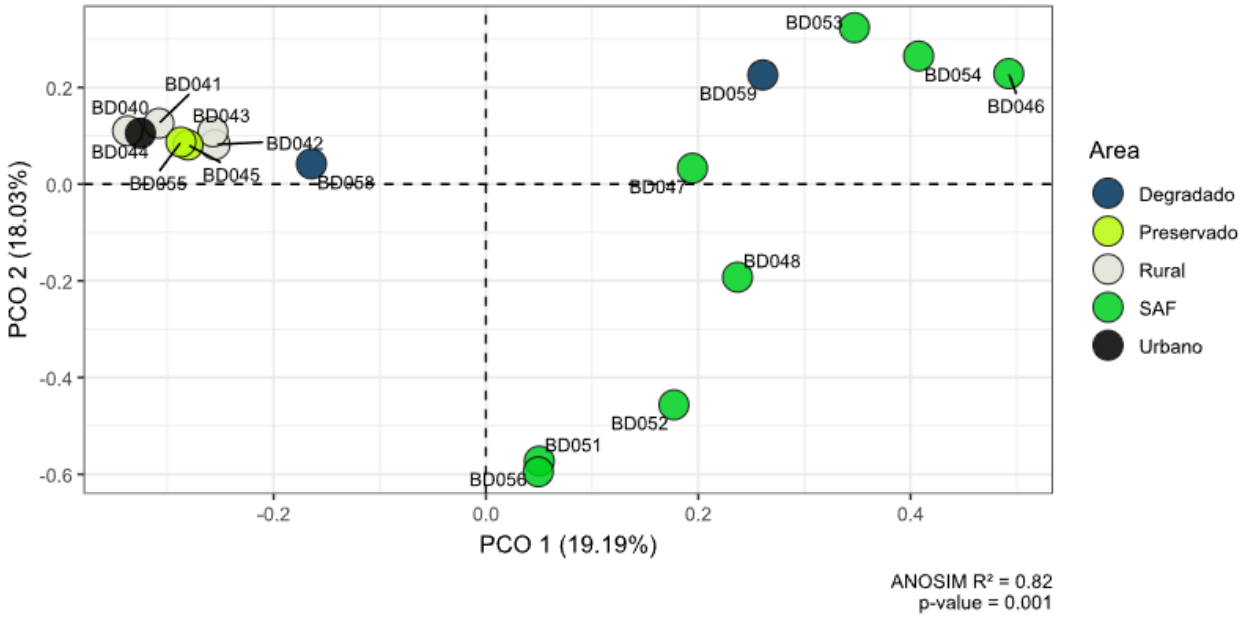


Figura Suplementar 5. Análise multivariada aplicada às amostras de solo, comparando áreas preservadas, degradadas e SAFs (juntos maduro e recente). PCoA com distância de Jaccard baseada em dados de presença/ausência de OTUs. Há separação entre áreas preservadas e SAFs e com áreas degradadas ocupando posição intermediária. A análise ANOSIM indica diferença significativa entre os grupos (R^2 = 0.663, p = 0.006).

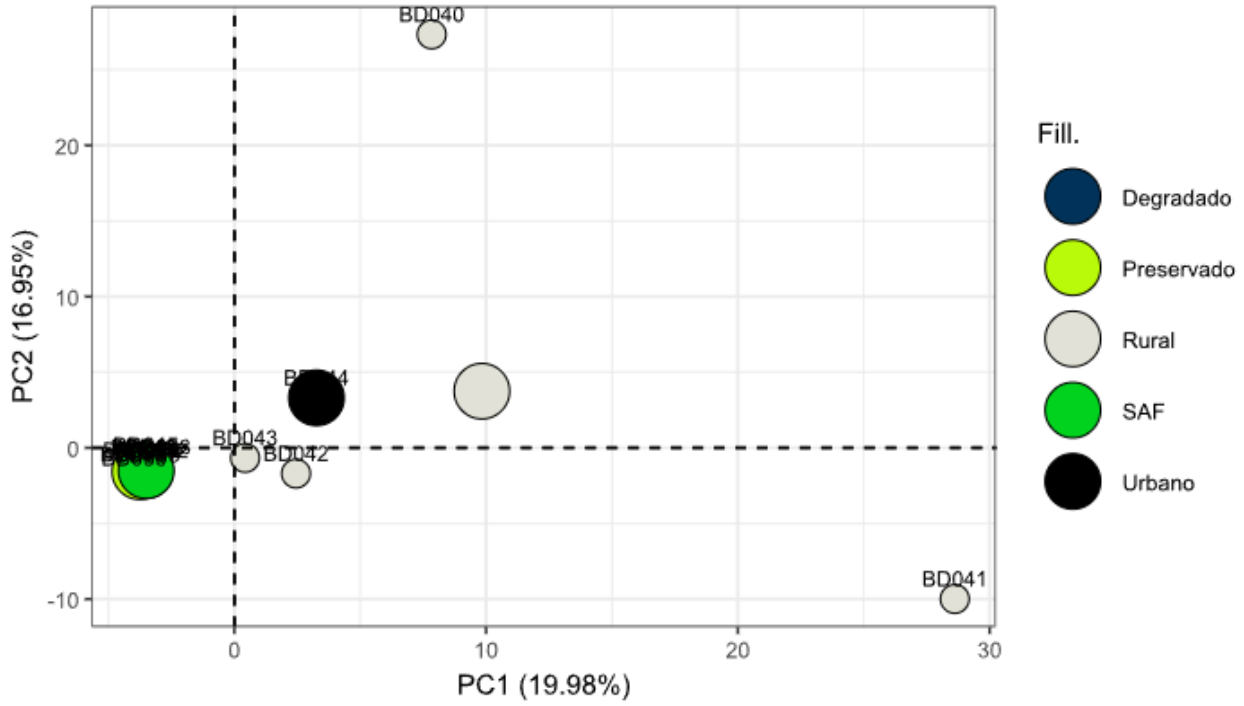


Figura Suplementar 6. Análise multivariada aplicada às amostras de solo, comparando áreas preservadas, degradadas e SAFs (juntos maduro e recente). PCA baseada em abundância relativa de OTUs, revelando um gradiente de transição entre áreas degradadas, SAFs e preservadas. SAF maduro tende a se agrupar mais próximo das áreas preservadas, enquanto SAF recente se aproxima das áreas degradadas.

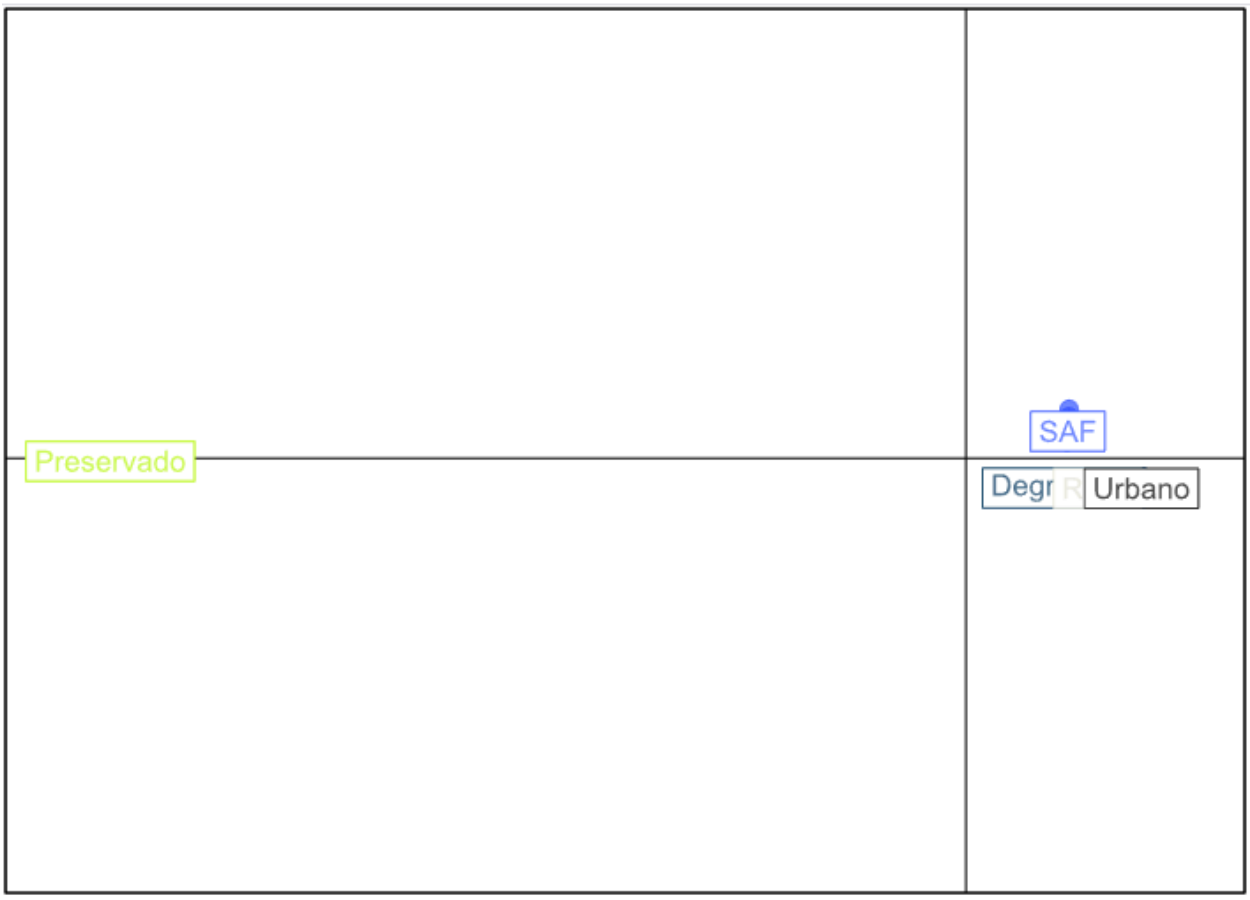


Figura Suplementar 7. Análise multivariada aplicada às amostras de solo, comparando áreas preservadas, degradadas e SAFs (juntos maduro e recente). DAPC utilizando classificação prévia dos grupos. A separação entre área preservada e demais áreas é evidente. Os SAFs e a área degradada se sobrepõem, indicando que ainda não há distinção clara entre SAFs, seja maduro ou recente e área degradada com base na composição biológica detectada por DNA ambiental.

Análises multivariadas aplicadas às amostras de solo excluindo áreas preservadas

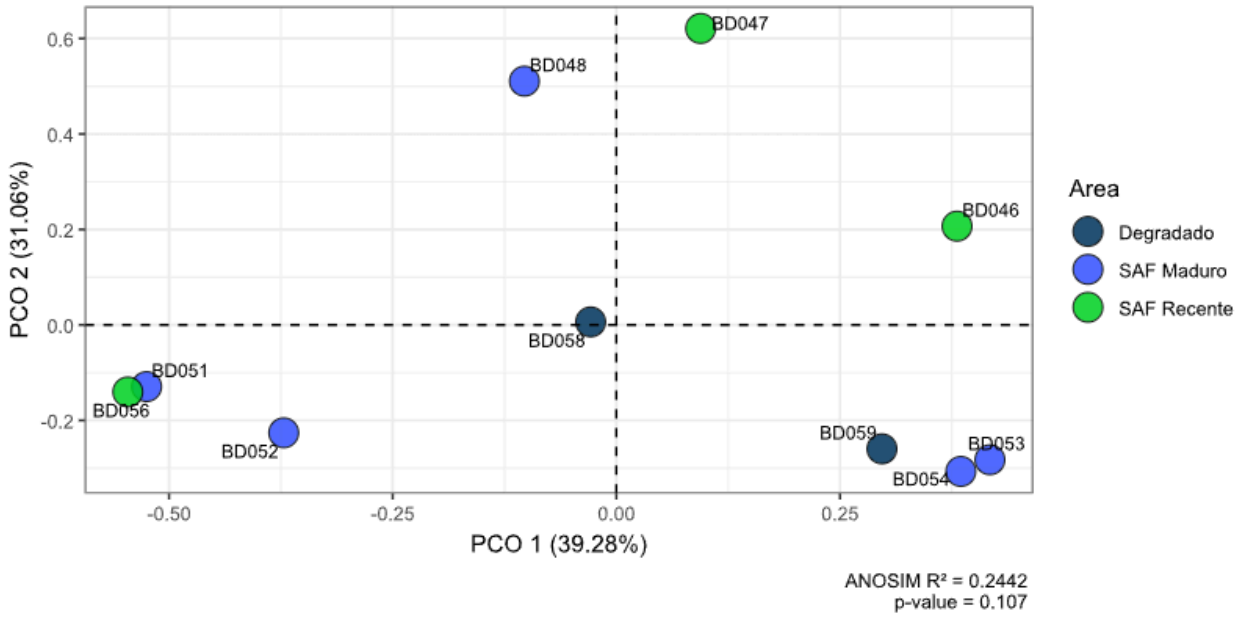


Figura Suplementar 8. Análises multivariadas conduzidas sem a inclusão das amostras de área preservada. PCoA com distância de Jaccard mostra a distribuição das amostras de SAFs (maduros e recentes) e áreas degradadas. Não foi observada separação estatisticamente significativa entre os grupos (ANOSIM $R^2 = 0.2442$, $p = 0.144$).

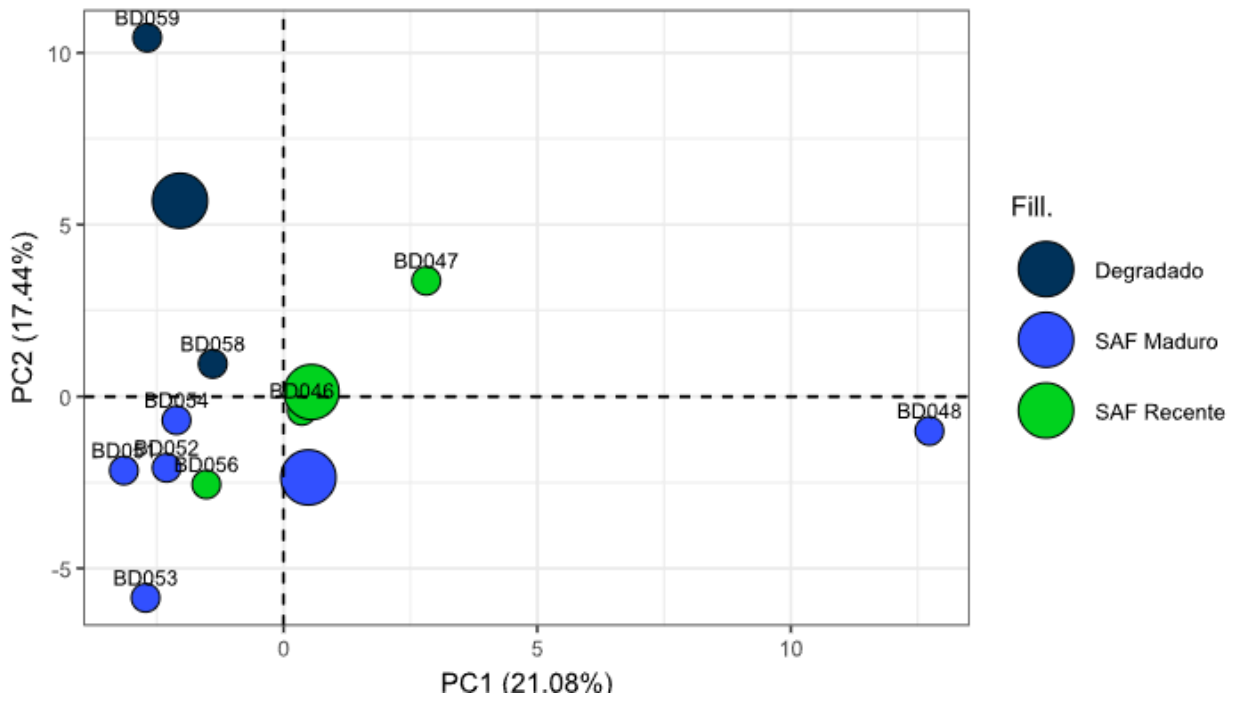


Figura Suplementar 9. Análise multivariada conduzidas sem a inclusão das amostras de área preservada. PCA baseada em abundância relativa de OTUs mostra sobreposição entre os grupos com somente uma amostra de SAF maduro com maior distância entre as outras. No entanto, o baixo poder de explicação dos PC1 e PC2 (menor que 50%), sugere que essa diferença não é biológica.

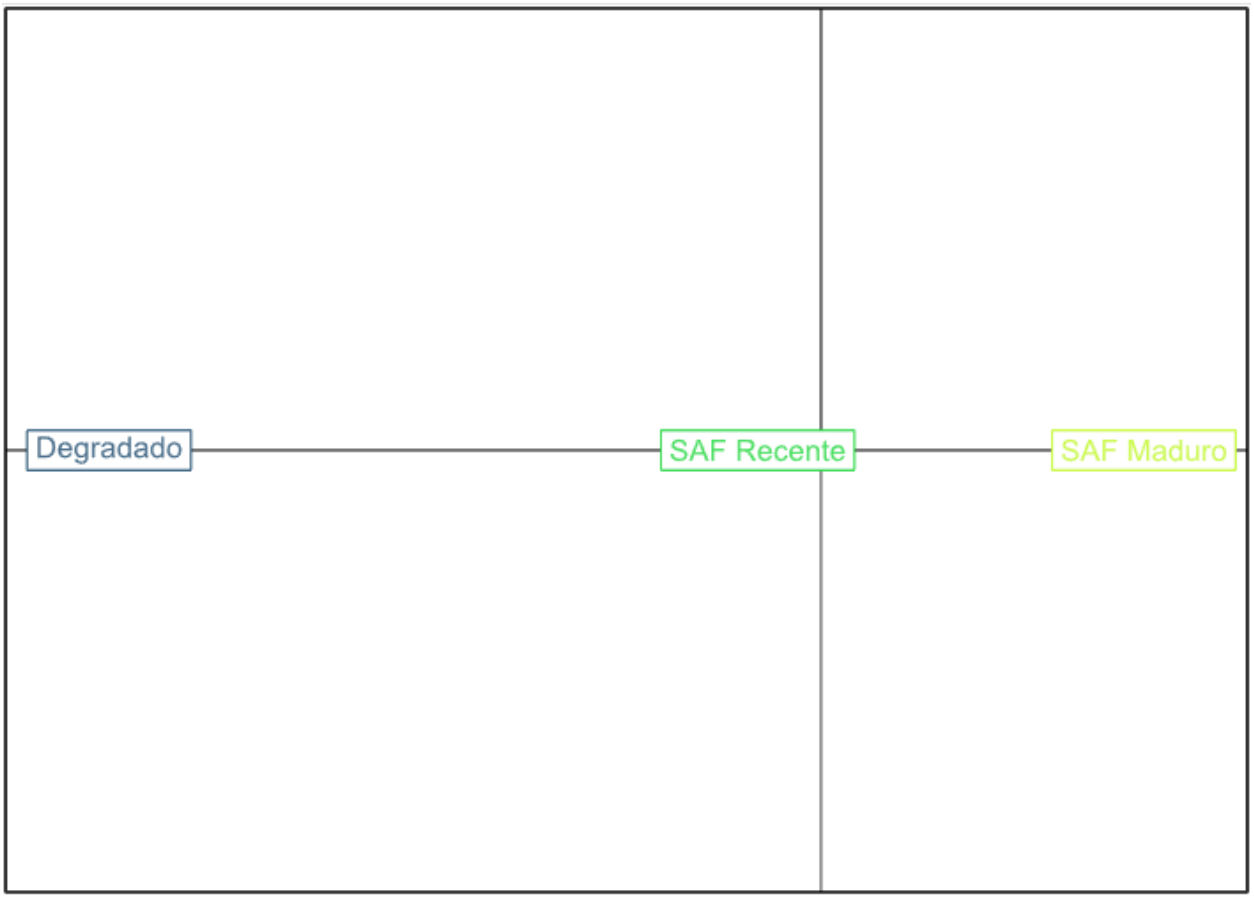


Figura Suplementar 10. Análises multivariadas conduzidas sem a inclusão das amostras de área preservada. DAPC apresenta elipses parcialmente separadas entre SAF maduro e SAF recente, mas com ampla sobreposição entre ambos e com as áreas degradadas.