todabio
By Bio Bureau

RELATÓRIO FINAL

Aplicação de eDNA no cálculo de créditos de biodiversidade em sistemas agroflorestais

Cliente:

BELTERRA

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

Versão 2.0

Data:
16/03/2026

SUMÁRIO

SUMÁRIO **2**
APRESENTAÇÃO **3**
MÉTODOS **4**
Coleta de amostras ambientais 4
Extração de DNA ambiental 6
PCR, construção de bibliotecas e sequenciamento 6
Bibliotecas de metabarcoding 6
Bibliotecas de shotgun 7
Análise Bioinformática: controle de qualidade e classificação taxonômica das sequências 7
Cálculo da diversidade total (DT) 8
D-micro 9
D-meso 9
RESULTADOS **11**
Extração de DNA ambiental 11
Construção de bibliotecas e sequenciamento 11
Construção de bibliotecas 11
Sequenciamento 11
Classificação taxonômica 12
Diversidade microbiana (Dmicro) 13
Diversidade de organismos macroscópicos (Dmeso) 17
Diversidade total (DT) 19
Crédito de Biodiversidade 20
Desafios identificados para a aplicação do índice 21
CONCLUSÕES **23**
PESPECTIVAS E RECOMENDAÇÕES **23**
GLOSSÁRIO **25**

APRESENTAÇÃO

O objetivo deste trabalho foi avaliar a aplicação de dados de DNA ambiental no cálculo de créditos de biodiversidade em sistemas agroflorestais (SAFs) de acordo com a metodologia proposta pela Belterra. Para isso, foram utilizadas métricas derivadas do sequenciamento de DNA ambiental para estimar a diversidade total (DT) das áreas avaliadas, considerando componentes associados à biodiversidade microbiana (D_{micro}) e à biodiversidade de organismos macroscópicos (D_{meso}).

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

Os cálculos apresentados neste relatório foram realizados com base em dois conjuntos de amostras ambientais. O primeiro corresponde ao conjunto de amostras de referência já descrito nos relatórios de *baseline*, coletado entre 02 e 03 de setembro de 2024 e processado por **metabarcoding**. O segundo conjunto corresponde a dados de sequenciamento por **shotgun** metagenômico de 8 amostras, selecionadas dentre 12 coletadas entre 30 de setembro e 01 de outubro de 2025 (Tabela 1). Para o sequenciamento foram selecionadas as amostras de áreas sem redundância, ou seja, foram coletadas três amostras de SAF <2 anos e duas de área degradada e uma amostra de cada ambiente foi sequenciado.

Enquanto o *metabarcoding* se baseia na amplificação de marcadores genéticos específicos para grupos-alvo de interesse, o shotgun metagenômico consiste no sequenciamento do DNA total presente na amostra, sem enriquecimento prévio para grupos taxonômicos específicos. Por essa razão, essa abordagem permite uma caracterização mais abrangente da biodiversidade, além da caracterização do repertório gênico e funcional da comunidade.

Tabela 1. Descrição e informações das amostras coletadas para posterior processamento por *shotgun*. Para cada amostra, são apresentados o nome das amostras, a amostra do mesmo local no estudo de metabarcoding, a fazenda onde estão localizadas, classificação da área, data de amostragem, latitude e longitude, controle de qualidade e se foi sequenciada.

Amostra	Amostra metabarcoding	Fazenda	Classificação da área	Data da coleta	Latitude (S)	Longitude (O)	Controle de qualidade	Sequenciada?
BD268SL	BD045SL	Cultrosa	Reserva	01/10/2025	13°56'29,046	39°11'7,335	Sim	Sim
BD269SL	BD046SL	Cultrosa	SAF < 2 anos	01/10/2025	13°56'18,202	39°10'41,711	Sim	Sim
BD270SL	BD047SL	Cultrosa	SAF < 2 anos	30/09/2025	13°56'38,914	39°10'51,984	Sim	Não
BD271SL	BD048SL	Cultrosa	SAF 2 anos	30/09/2025	13°56'59,026	39°10'46,722	Sim	Não
BD272SL	BD049SL	Cultrosa	SAF 3 anos	30/09/2025	13°57'20,815	39°10'41,258	Sim	Sim
BD273SL	BD050SL	Cultrosa	SAF 5 anos	30/09/2025	13°57'10,824	39°10'23,424	Sim	Sim
BD274SL	BD051SL	Girassol	SAF maduro	01/10/2025	13°55'52,763	39°8'32,651	Não	Não
BD275SL	BD055SL	Sussuarana	Reserva	01/10/2025	14°0'21,121	39°9'39,161	Sim	Sim
BD276SL	BD056SL	Sussuarana	SAF < 2 anos	01/10/2025	14°0'16,72	39°9'37,957	Sim	Sim
BD277SL	BD057SL	Sussuarana	Degradada (pasto)	01/10/2025	14°0'39,489	39°9'21,983	Sim	Sim
BD278SL	BD058SL	Sussuarana	Degradada (c/ rasteira)	01/10/2025	14°0'20,54	39°9'22,884	Sim	Sim
BD279SL	BD059SL	Sussuarana	Degradada (pasto)	01/10/2025	14°0'24,548	39°9'10,266	Sim	Não

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 para *metabarcoding*: foram homogeneizadas e submetidas ao protocolo padrão do kit DNeasy PowerLyzer PowerSoil (Qiagen), que inclui etapas de remoção de impurezas e possíveis inibidores.
- Amostras de solo para *shotgun*: foram homogeneizadas e submetidas ao protocolo padrão do kit NucleoSpin Soil Mini Kit (Macherey-Nagel), 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

Bibliotecas de *metabarcoding*

O DNA ambiental foi enviado 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).

Bibliotecas de *shotgun*

O DNA ambiental foi enviado à empresa (Macrogen, Hong Kong). Oito bibliotecas foram construídas utilizando o kit TruSeq Nano DNA Library, de acordo com manual do fabricante e sequenciadas na plataforma Novaseq X (Illumina). Diferentemente do metabarcoding, esse método não utiliza amplificação por PCR de marcadores genéticos específicos e todo o DNA presente na amostra é analisado diretamente, sem a seleção prévia de genes ou organismos.

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.

Os dados de **metabarcoding** foram processados com o programa QIIME2, incluindo correção de erros de sequenciamento e agrupamento das sequências em **Unidades Taxonômicas Operacionais (OTUs)**, que reúnem sequências altamente semelhantes (99%) e são amplamente utilizadas em estudos de biodiversidade. As OTUs resultantes foram comparadas com bancos de dados genéticos para identificação dos organismos presentes nas amostras e verificação de possíveis contaminações.

Os dados de **shotgun** metagenômico foram analisados por duas abordagens complementares. A primeira utilizou o programa Kraken2, empregado para a **classificação taxonômica** das sequências e geração de uma lista geral dos táxons detectados nas amostras. A segunda utilizou o programa HUMAnN4, voltado principalmente à caracterização funcional de comunidades microbianas, com ênfase em bactérias. Essa análise permitiu identificar famílias gênicas, reações bioquímicas e vias metabólicas representadas nas amostras, fornecendo uma descrição do potencial funcional associado à fração microbiana.

Os resultados dessas análises foram então utilizados nas etapas subsequentes de cálculo da diversidade microbiana, da diversidade de organismos macroscópicos e da diversidade total, integrando informações taxonômicas e funcionais derivadas das abordagens de metabarcoding e shotgun metagenômico.

Cálculo da diversidade total (DT)

Na metodologia proposta pela Belterra, a diversidade total ($DT = \alpha D_{meso} + \beta D_{micro}$) é definida a partir de dois componentes complementares: a diversidade de organismos macroscópicos (D_{meso}) e a diversidade microbiana (D_{micro}). Dessa forma, a abordagem adotada busca integrar, em uma única métrica, informações associadas tanto à fração visível da biodiversidade quanto à fração microbiana inferida a partir dos dados de sequenciamento.

D-micro

Na metodologia, a diversidade microbiana (D_{micro}) é calculada como o produto entre a diversidade taxonômica e uma proxy de diversidade funcional, segundo a relação $D_{micro} = H' \times DivFun$.

Para esse estudo, estimamos a diversidade taxonômica e funcional da comunidade microbiana a partir dos dados de sequenciamento shotgun das amostras da 2a coleta no campo.

A diversidade taxonômica foi estimada pelo índice de Shannon (H') calculado como $H' = \sum_i p_i \, lnp_i$, onde p_i representa a abundância relativa do táxon *i* na amostra.

A diversidade funcional foi estimada, utilizando três métricas obtidas (via pipeline HUMAnN4) para cada amostra como proxies de diversidade funcional: o número de famílias gênicas ($N_{GeneFamilies}$), o número de reações bioquímicas ($N_{Reactions}$) e o número de vias metabólicas ($N_{Pathways}$). Como as três informações estavam disponíveis a partir da mesma análise funcional, elas foram utilizadas de forma exploratória para avaliar se diferentes níveis de descrição funcional produziriam diferenças relevantes na estimativa da diversidade microbiana. Com isso, foram geradas três estimativas complementares: $D_{GeneFamilies} = H' \times N_{GeneFamilies}$, $D_{Reactions} = H' \times N_{Reactions}$ e $D_{Pathways} = H' \times N_{Pathways}$.

D-meso

Neste estudo, estimamos a diversidade de organismos macroscópicos (D_{meso}) a partir dos dados de sequenciamento metabarcoding das amostras da 1a coleta no campo. D_{meso} foi calculada a partir da combinação entre diversidade taxonômica e dispersão funcional, considerando separadamente os compartimentos de metazoários e plantas. Essa separação foi adotada porque esses dois grupos, que dominaram os dados, dependem de bancos de traços ecológicos distintos.

Para cada compartimento, a diversidade taxonômica foi estimada como o produto entre o índice de Shannon (H') e a dispersão funcional (FDis). O índice de Shannon foi calculado como $H' = \sum_i p_i \, lnp_i$, onde p_i representa a abundância relativa do táxon *i* na amostra. A dispersão funcional (FDis) foi calculada a partir de matrizes de traços ecológicos específicas para metazoários e plantas, utilizando distâncias funcionais entre táxons e sua posição em um espaço funcional reduzido. Para cada grupo, FDis foi obtido como $FDis = \frac{\sum_i p_i d_i}{\sum_i p_i}$, onde d_i corresponde à distância do táxon *i* em relação ao centroide funcional da comunidade. Desta forma, foram calculadas as métricas de diversidade: $D_{metazoa} = H'_{metazoa} \times Fdis_{metazoa}$ e $D_{plantas} = H'_{plantas} \times Fdis_{plantas}$.

Por fim, a diversidade macroscópica foi expressa como uma média ponderada entre os dois componentes, segundo a relação $D_{meso} = w_{metazoa} D_{metazoa} + w_{plantas} D_{plantas}$, em que os pesos refletem a contribuição relativa de metazoários e plantas em cada amostra (número de leituras de sequenciamento para cada um). Dessa forma, a métrica final incorpora simultaneamente a diversidade taxonômica e funcional dos dois principais compartimentos macroscópicos detectados.

RESULTADOS

Extração de DNA ambiental

O DNA ambiental das amostras de solo para shotgun foi extraído em quantidade suficiente para sequenciamento para 9 amostras e um controle negativo da extração[1].

Construção de bibliotecas e sequenciamento

Construção de bibliotecas

Das 9 amostras de solo sequenciadas por shotgun, 8 amostras passaram no controle de qualidade e tiveram bibliotecas construídas e sequenciadas. A amostra BD274SL (Girassol, SAF maduro) não passou no controle de qualidade e não seguiu para o sequenciamento[2].

Sequenciamento

Para os dados de metabarcoding, foram analisadas 54 bibliotecas, cada uma correspondente ao conjunto de sequências gerado para um marcador genético (COI, 12S, 16S, ITS2 e/ou rbcL) em uma amostra ambiental (Tabela S2). Considerando a soma das bibliotecas por amostra, cada amostra gerou em média cerca de 1,2 milhão de leituras brutas e 412 mil leituras não-quiméricas. As **leituras não-quiméricas**

[1] A descrição dos resultados detalhados para extração de DNA do metabarcoding foram descritos no relatório anterior.

[2] O detalhamento do resultado da construção de bibliotecas para metabarcoding foi descrito detalhadamente no relatório anterior.

correspondem ao conjunto final obtido após as etapas de processamento bioinformático e refletem a fração de sequências de alta qualidade efetivamente disponíveis para classificação taxonômica, indicando quantidade adequada para as análises subsequentes.

As amostras de shotgun geraram, em média, cerca de 82,6 milhões de leituras por amostra, com comprimento mediano de 151 pb e aproximadamente 12 Gbp por biblioteca (Tabela S3). O maior volume de sequenciamento na comparação com o metabarcoding é necessário porque o shotgun sequencia o DNA total presente na amostra, exigindo maior profundidade para recuperar adequadamente a diversidade taxonômica e funcional. Os relatórios de qualidade indicaram bom desempenho das amostras nos principais parâmetros avaliados, incluindo estatísticas básicas, qualidade por base e conteúdo de GC, sem evidência de problemas sistemáticos relevantes. Esses resultados indicam que os dados apresentaram qualidade adequada para as etapas subsequentes de classificação taxonômica e análise funcional.

Classificação taxonômica

O processamento dos dados de shotgun com a ferramenta Kraken2 mostrou grande abundância de sequências bacterianas, que dominaram as classificações (Tabela S5). Isso deve ser consequência da maior proporção de DNA bacteriano nas amostras, já que o shotgun sequencia o DNA total, sem enriquecer sequências de um grupo taxonômico específico. Dos 1895 táxons identificados, 1636 são bactérias, com apenas 50 plantas e 46 metazoários.

Uma ressalva importante é que Kraken2 utiliza uma forma diferente de busca por correspondência em banco de espécies, utilizando perfis de pequenos fragmentos da sequência (**Kmers**) ao invés do alinhamento da sequência toda (**BLAST**). Enquanto o BLAST é a metodologia padrão para comparação de dados de metabarcoding, que produz muito menos leituras de sequenciamento para comparar, Kraken2, que seria computacionalmente impraticável com os milhões de leituras de sequenciamento produzidas, é o padrão para dados de shotgun. A comparação



desses dois métodos não fazia parte do escopo deste estudo (e, computacionalmente, impraticável). Porém, Kraken2, favorece a identificação de microorganismos e limita a identificação de metazoários em função do grande número de genomas completos de microorganismos em base de dados comparado ao número limitado de genomas de metazoários.

De fato, a baixa riqueza de táxons macroscópicos recuperada pelos dados de shotgun indicou que esse conjunto não era o mais adequado para sustentar o cálculo da **diversidade macroscópia** (D_{meso}). Para remediar esse resultado, utilizamos os dados de metabarcoding produzidos na primeira coleta[3].

Para o cálculo da **diversidade microscópica** (D_{micro}), por sua vez, foram utilizados os resultados do programa HUMAnN4, que além de fornecer a composição taxonômica (tal como o Kraken2, Tabela S6), fornece também métricas funcionais empregadas como proxy de diversidade funcional microbiana, como por exemplo, número de famílias gênicas (a seguir).

Diversidade microbiana (D_{micro})

A Tabela 2 apresenta, para cada amostra, os valores das três métricas utilizadas neste estudo como proxy de diversidade funcional microbiana: número de famílias gênicas, número de reações bioquímicas e número de vias metabólicas inferidas a partir dos dados de sequenciamento. De modo geral, o padrão de variação entre amostras foi semelhante para as três métricas: amostras com maior número de famílias gênicas tendem também a apresentar maior número de reações e de vias metabólicas, enquanto amostras com menor valor em uma dessas métricas também tendem a apresentar valores menores nas demais.

[3] O processamento dos dados de metabarcoding identificou 304 taxons, dos quais 10 classificados a nível de filo, 26 a classes, 86 a ordens, 140 a famílias, 180 a gêneros e 97 classificados a nível de espécie (Tabela S4). A maior parte das OTUs foi classificada como metazoário (181) ou planta (116), com algumas poucas bactérias (4). O enriquecimento por metazoários e plantas não é ao acaso, mas consequência dos marcadores selecionados para o processamento, que amplificam e enriquecem o DNA desses grupos taxonômicos.

Tabela 2. Número de táxons detectados e valores das três métricas utilizadas como proxy de diversidade funcional microbiana em cada amostra: número de famílias gênicas, número de reações bioquímicas e número de vias metabólicas inferidas a partir dos dados de sequenciamento. As amostras estão identificadas por tipo de uso do solo e fazenda de origem.

Amostra	Tipo	Fazenda	Nº de táxons	Nº de famílias gênicas	Nº de reações	Nº de vias metabólicas
BD268D	Reserva	Cultrosa	187	185800	1937	692
BD269D	SAF < 2 anos	Cultrosa	259	239067	1970	701
BD272D	SAF 3 anos	Cultrosa	323	239817	1974	696
BD273D	SAF 5 anos	Cultrosa	291	217990	1969	707
BD275D	Reserva	Sussuarana	145	190845	1937	678
BD276D	SAF < 2 anos	Sussuarana	332	254320	1988	710
BD277D	Área degradada	Sussuarana	233	195125	1950	697
BD278D	Área degradada	Sussuarana	438	261295	1985	697

A Tabela 3 reúne os valores de $D_{GeneFamilies}$, $D_{Reactions}$ e $D_{Pathways}$ calculados para cada amostra. Em geral, as amostras de SAF apresentaram valores de D_{micro} superiores aos das reservas. As duas reservas exibiram os menores valores do conjunto, especialmente da Reserva de Sussuarana (BD275D) , enquanto as amostras de SAF mantiveram valores consistentemente mais altos, com destaque para de SAF de 5 anos de Cultrosa (BD273D) e SAF de < 2 anos de Sussuarana (BD276D) . As amostras de área degradada mostraram maior heterogeneidade: pasto de Sussuarana (BD277D) apresentou valores próximos aos das reservas, enquanto a área degradada com rasteira de Sussuarana (BD278D) exibiu os maiores valores de todo o conjunto, indicando grande variação interna dentro dessa categoria.

Tabela 3. Valores do índice de Shannon (H') e das três estimativas de diversidade microbiana calculadas para cada amostra: DGeneFamilies, DReactions e DPathways. As métricas de diversidade microbiana foram obtidas pelo produto entre H′ e, respectivamente, o número de famílias gênicas, o número de reações bioquímicas e o número de vias metabólicas inferidas a partir dos dados de sequenciamento.

Amostra	Tipo	Fazenda	Shannon (H')	$D_{GeneFamilies}$	$D_{Reactions}$	$D_{Pathways}$
BD268D	Reserva	Cultrosa	2,08	386919,09	4033,70	1441,05
BD269D	SAF < 2 anos	Cultrosa	2,79	667965,17	5504,28	1958,63
BD272D	SAF 3 anos	Cultrosa	2,52	603963,63	4971,39	1752,83
BD273D	SAF 5 anos	Cultrosa	2,90	632119,19	5709,63	2050,13
BD275D	Reserva	Sussuarana	1,86	354337,94	3596,39	1258,83
BD276D	SAF < 2 anos	Sussuarana	2,86	727231,32	5684,71	2030,25
BD277D	Área degradada	Sussuarana	1,97	385122,78	3848,76	1375,69
BD278D	Área degradada	Sussuarana	3,73	975010,68	7406,94	2600,82

A Figura 1 resume esses resultados em três painéis: $D_{GeneFamilies}(Figura\ 1A)$, $D_{Reactions}(Figura\ 1B)$ e $D_{Pathways}(Figura\ 1C)$. Embora os valores absolutos variem entre as três métricas, o padrão geral foi semelhante em todas elas: reservas com valores mais baixos, SAFs com valores intermediários a altos e área degradada com rasteira (BD278D) como valor máximo. Assim, a escolha entre famílias gênicas, reações bioquímicas ou vias metabólicas teve pouca influência sobre o padrão comparativo geral observado para D_{micro}.

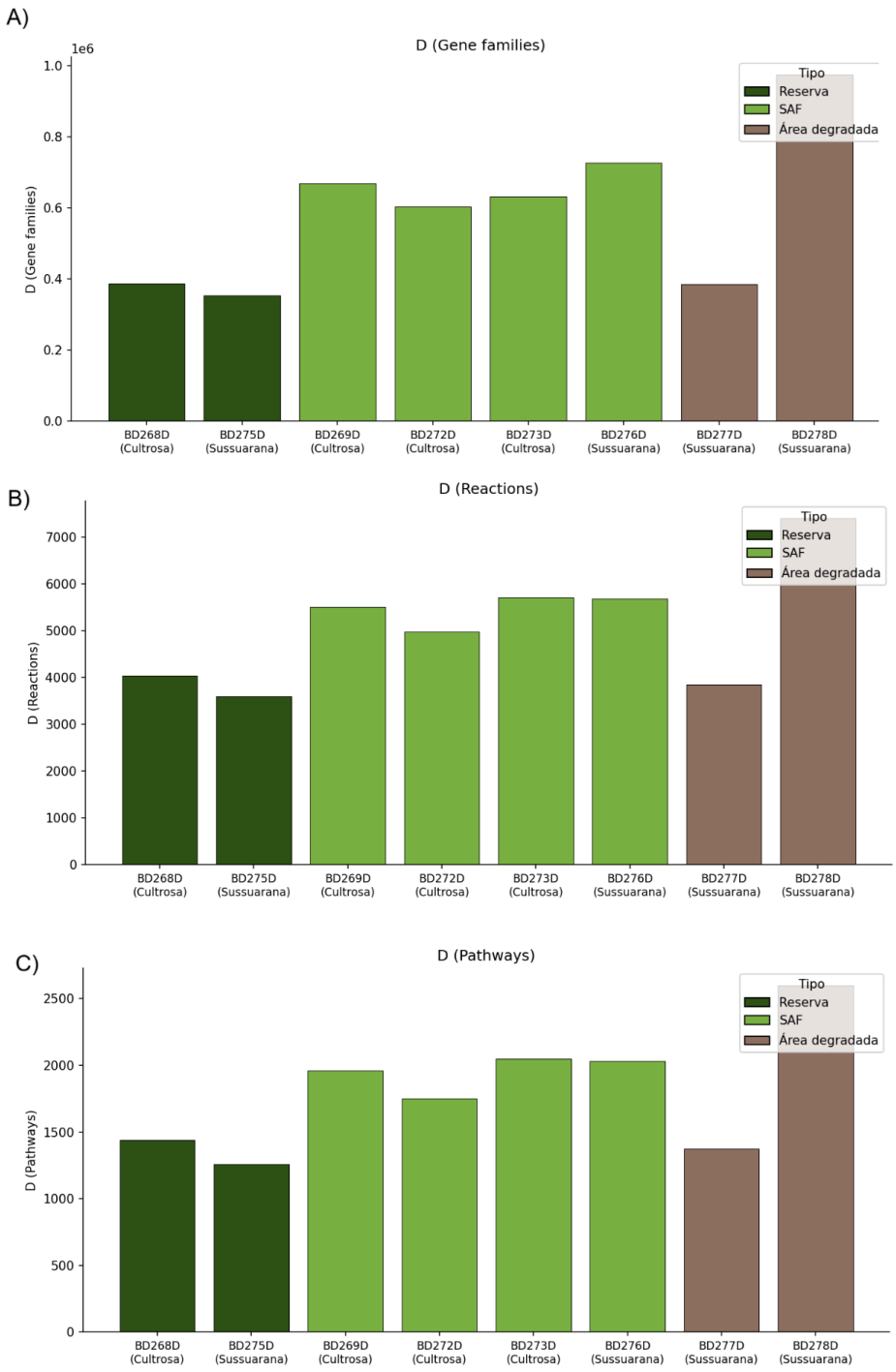


Figura 1. Valores das três estimativas de diversidade microbiana calculadas para cada amostra, considerando como proxy de diversidade funcional o número de famílias gênicas (A), o número de reações bioquímicas (B) e o número de vias metabólicas (C). Em todos os painéis, as amostras estão organizadas por tipo de uso do solo e fazenda de origem.

Diversidade de organismos macroscópicos (D_{meso})

Embora o conjunto de dados de shotgun tenha sido utilizado para caracterizar a fração microbiana (D_{micro}), essas amostras apresentaram predominância de DNA bacteriano, com baixa recuperação de táxons macroscópicos. Por essa razão, o cálculo de D_{meso} foi realizado principalmente com base no conjunto anterior de amostras (2024), processadas com a abordagem de metabarcoding, que permitiu melhor representação de metazoários e plantas.

A estimativa de D_{meso} exigiu, como etapa prévia, a consulta a bancos de traços ecológicos para o cálculo da dispersão funcional (FDis) dos táxons detectados. Nesse processo, observou-se baixa sobreposição entre os traços disponíveis para metazoários e para plantas, de modo que o cálculo precisou ser realizado separadamente para esses dois compartimentos. Assim, foi estimado um valor de FDis — e, por consequência, um componente de diversidade — para metazoários e outro para plantas, os quais foram posteriormente combinados em D_{meso} por meio de pesos proporcionais à abundância relativa de cada grupo em cada amostra. Esse procedimento permitiu operacionalizar o cálculo, mas também evidenciou um desafio técnico importante: a dependência de bases de traços ecológicos com cobertura e estrutura compatíveis entre diferentes grupos biológicos.

A Tabela 4 apresenta os valores de D_{meso} calculados para cada amostra. Como essa métrica resulta da combinação entre os componentes de metazoários e plantas, também são mostrados os valores de $D_{metazoa}$, $D_{plantas}$ e os pesos relativos atribuídos a cada compartimento. Considerando as comparações dentro de uma mesma fazenda, observou-se na Cultrosa um aumento de D_{meso} da área preservada para os sistemas agroflorestais: a amostra preservada (BD045) apresentou valor inferior, enquanto as amostras de SAF recente (BD046 e BD047) exibiram valores mais altos, e a amostra de SAF maduro (BD048) apresentou o maior valor desse conjunto. Na Sussuarana, por outro lado, a amostra de reserva (BD055) e a amostra de SAF recente (BD056) apresentaram valores igualmente baixos de D_{meso}, enquanto as

amostras de área degradada mostraram maior variação, incluindo uma amostra com valor nulo (BD058) e outra com valor mais elevado (BD059).

Tabela 4. Valores da diversidade de organismos macroscópicos (D_{meso}) calculados para cada amostra, com discriminação de seus dois componentes: diversidade associada a metazoários ($D_{metazoa}$) e diversidade associada a plantas ($D_{plantas}$). A métrica D_{meso} corresponde à combinação ponderada desses dois componentes, considerando a contribuição relativa de metazoários e plantas em cada amostra. Células vazias indicam ausência de táxons de um dos grupos para a amostra.

Amostra	Tipo	Fazenda	$D_{metazoa}$	$D_{plantas}$	D_{meso}
BD040	Rural	Cultrosa	0.1257	0.1462	0.1257
BD041	Rural	Cultrosa	0.1131	0.5983	0.3826
BD042	Rural	Girassol	0.0969		0.0969
BD043	Rural	Sussuarana	0.0542	0.2173	0.1093
BD044	Urbano	Cachoeira Acaraí	0.0844	0.0	0.0844
BD045	Preservada	Cultrosa		0.0018	0.0018
BD046	SAF recente	Cultrosa	0.0427	0.0532	0.0532
BD047	SAF recente	Cultrosa		0.0419	0.0419
BD048	SAF maduro	Cultrosa	0.0756	0.2015	0.2015
BD051	SAF maduro	Girassol		0.0075	0.0075
BD052	SAF maduro	Girassol	0.0	0.2732	0.2732
BD053	SAF maduro	Girassol		0.0465	0.0465
BD054	SAF maduro	Girassol		0.0177	0.0177
BD055	Reserva	Sussuarana		0.0012	0.0012
BD056	SAF recente	Sussuarana		0.0007	0.0007
BD058	Degradado	Sussuarana		0.0	0.0
BD059	Degradado	Sussuarana		0.0473	0.0473

É importante notar uma grande discrepância (superior a 10x), entre os valores de D_{meso} em diferentes amostras, por exemplo 0.0532 para “SAF recente - Cultrosa” (BD046) e 0.0018 para “Reserva - Cultrosa” (BD045). Apesar de podermos eventualmente encontrar explicações biológicas para um SAF ter uma biodiversidade 10x superior a uma área de reserva, essa discrepância é melhor explicada por 1) variação na eficiência da extração de DNA 2) variação na microlocalização da

amostra. A limitação no número de amostras com $D_{metazoa}$ está ligada aos problemas com sequenciamento da primeira coleta que já foram relatados no primeiro relatório.

Adicionalmente, a utilização do índice de Shannon para o cálculo de D_{meso} pode não ser a abordagem mais adequada neste contexto, já que ainda há controvérsia sobre o quanto dados de metabarcoding refletem as abundâncias reais dos táxons. Como esse índice é sensível às abundâncias relativas, parte da variação observada pode refletir vieses metodológicos, e não diferenças biológicas reais. Isso pode ser observado na comparação entre BD045 e BD046: ambas apresentam 24 táxons de plantas detectados, mas diferem em cerca de sete vezes no índice de Shannon, indicando que essa discrepância decorre principalmente das abundâncias estimadas. Assim, diferenças nas abundâncias relativas parecem explicar boa parte da variação em Shannon e, consequentemente, em D_{meso}, embora valores de abundância em metabarcoding devessem ser considerados com cautela.

Diversidade total (DT)

A Tabela 5 apresenta os valores de diversidade total (DT) calculados para as amostras com estimativas disponíveis tanto de D_{meso} quanto de D_{micro}, segundo a relação $DT = \alpha D_{meso} + \beta D_{micro}$, com α=1.000.000 e β=1. Esses valores foram adotados para tornar as contribuições de D_{meso} e D_{micro} mais comparáveis em ordem de grandeza, uma vez que os valores brutos de D_{meso} eram muito menores que os de D_{micro}. Com essa ponderação, evitou-se que a componente microbiana dominasse a métrica final apenas por diferença de escala numérica. Ainda assim, essa etapa evidencia que a construção de um índice integrado depende de escolhas de normalização e ponderação que influenciam diretamente a interpretação ecológica do resultado.

Tabela 5. Valores de diversidade total (DT) calculados para as amostras com estimativas disponíveis tanto de Dmeso quanto de Dmicro. A tabela apresenta, para cada área, os códigos correspondentes das amostras de metabarcoding e shotgun, bem como os valores de Dmso, Dmicro e da diversidade total calculada segundo a relação DT=1.000.000×Dmeso+DmicroDT,

Amostra metabarcoding	Amostra shotgun	Fazenda	Tipo solo	D_{meso}	D_{micro}	DT (1e6*Dmeso+1*Dmicro)
BD045	BD268D	Cultrosa	Reserva	0,0018	1441,0549	3197,6681
BD046	BD269D	Cultrosa	SAF < 2 anos	0,0532	1958,6291	55206,7032
BD055	BD275D	Sussuarana	Reserva	0,0012	1258,8285	2438,1544
BD056	BD276D	Sussuarana	SAF < 2 anos	0,0007	2030,2542	2757,6990
BD058	BD278D	Sussuarana	Área degradada c/ rasteira	0,0000	2600,8245	2600,8245

O valor discrepante de D_{meso} BD046 (SAF<2 anos - Cultrosa = 0,0532), reportado anteriormente, gerou também um DT discrepante para SAF<2 anos - Cultrosa (aproximadamente 22 vezes superior ao da **Reserva)**, e por isso optamos por não utilizar esse dado para estimar o Delta DT (crédito de biodiversidade)

Na **Sussuarana**, as diferenças foram mais sutis: a amostra de **SAF < 2 anos** apresentou o maior valor de DT, seguida pela **área degradada com vegetação rasteira**, enquanto a **Reserva** apresentou o menor valor.

Crédito de Biodiversidade

Assim, pudemos aplicar a fórmula proposta pela Belterra para estimar créditos de biodiversidade na Fazenda Sussuarana:

SAF em relação à Reserva: **ΔDT = 319**

SAF em relação à Área degradada: **ΔDT = 157**

O resultado indica a viabilidade técnica de estimar o crédito de biodiversidade com a metodologia da Belterra a partir de dados de DNA ambiental, utilizando a metodologia shotgun, ou uma combinação shotgun/metabarcoding.

Esses resultados não devem, ainda, serem interpretados biologicamente, em função dos problemas de sequenciamento da primeira coleta, do número limitado de amostras da segunda coleta mas, principalmente, as limitações nas bases de dados de funcionalidades para microrganismos e metazoários.

Com as análises estatísticas ecológicas baseadas em riqueza e abundância dos dados de metabarcoding apresentados no primeiro relatório, demonstramos que SAF tinha ganho ecológico com relação à área degradada, ainda que em algumas (PCA) ficasse mais próximo a reserva e em outras (PCoA e DAPC) mais próximas á área degradada.

Dessa forma, o valor positivo do SAF com relação a Reserva quanto com relação à área degradada pode representar isso: um sistema em transição que, na ausência de um número representativo de amostras, pode se agrupar com um extremo ou outro.

Análises de DNA ambiental ganham robustez e poder de resolução com o acúmulo de informação histórica sobre o ambiente.

Desafios identificados para a aplicação do índice

Os resultados obtidos neste estudo devem ser interpretados como um exercício exploratório de aplicação do índice, útil sobretudo para evidenciar desafios metodológicos que ainda precisam ser enfrentados antes de sua utilização em contextos operacionais de cálculo de créditos de biodiversidade.

O primeiro desafio diz respeito à **representatividade das amostras de solo**. As comparações realizadas mostraram heterogeneidade importante entre amostras classificadas dentro de uma mesma categoria de uso do solo, indicando que o número e a distribuição espacial das amostras ainda podem ser insuficientes para representar de forma robusta a variação ecológica de cada tipo de área. Isso limita a

interpretação dos valores obtidos para D_{meso}, D_{micro} e DT como atributos generalizáveis de categorias como Reserva, SAF ou Área degradada.

O segundo desafio está relacionado à integração entre os dados de shotgun e metabarcoding. Na prática, os dados de shotgun puderam ser utilizados para o cálculo de D_{micro}, mas não forneceram recuperação suficiente de táxons macroscópicos para sustentar o cálculo de D_{meso}. Por essa razão, D_{meso} precisou ser estimado a partir do conjunto de metabarcoding gerado anteriormente. Isso introduz uma inconsistência importante no índice integrado, pois os componentes microbiano e macroscópico passaram a ser derivados de amostras diferentes, ainda que parcialmente coletadas nas mesmas áreas, e em momentos distintos. Assim, a diversidade total (DT) não reflete dois componentes medidos exatamente sobre a mesma amostra ambiental, mas sim uma combinação entre conjuntos que diferem em metodologia e também em tempo de coleta, o que limita a comparabilidade direta entre D_{micro} e D_{meso}.

O terceiro desafio envolve a construção de **uma matriz de traços ecológicos adequada** para o cálculo da dispersão funcional (Fdis). Na prática, nem todos os táxons identificados nas amostras possuíam traços disponíveis nas bases consultadas, o que reduziu a cobertura funcional do conjunto detectado. Além disso, foi necessário combinar diferentes bases de dados para aumentar o número de táxons com informação, mas essas bases nem sempre utilizam os mesmos atributos, o que dificulta a comparação direta entre organismos no cálculo do FDis. Outro ponto é que parte das variáveis disponíveis pode refletir condições ambientais associadas à ocorrência da espécie, e não características intrínsecas do organismo, enquanto alguns traços apresentam cobertura muito limitada entre os táxons. Dessa forma, o cálculo da dispersão funcional não depende apenas da existência de bases abertas, mas também de um processo criterioso de seleção, harmonização e filtragem dos traços, de modo a gerar uma matriz ecologicamente coerente para subsidiar a estimativa de D_{meso}.

CONCLUSÕES

A aplicação de dados de DNA ambiental ao cálculo exploratório de créditos de biodiversidade permitiu construir um índice integrado, combinando a diversidade microbiana (D_{micro}) e a diversidade de organismos macroscópicos (D_{meso}) em uma métrica única de diversidade total (DT).

O componente microbiano pôde ser estimado de forma consistente a partir dos dados de shotgun, com padrões semelhantes entre as diferentes proxies funcionais testadas. Já o componente macroscópico dependeu do uso de dados de metabarcoding e da consulta a bases de traços ecológicos, o que exigiu adaptações metodológicas para integrar metazoários e plantas em uma mesma métrica.

Os resultados também evidenciaram desafios importantes para a aplicação direta do índice, incluindo a heterogeneidade entre amostras de uma mesma categoria de uso do solo, a integração entre dados obtidos por metodologias e tempos de coleta distintos, e a cobertura ainda incompleta das bases abertas de traços ecológicos.

PESPECTIVAS E RECOMENDAÇÕES

Como próximos passos, sugerimos adotar uma estratégia de **eDNA first** para o monitoramento ambiental das propriedades, isto é, fazer do DNA ambiental o ponto de partida para a caracterização e o acompanhamento da biodiversidade. Na prática, isso significa priorizar **amostras compostas (*pool*) de solo** e **amostras de corpos d'água**, combinando metabarcoding e shotgun para construir uma base molecular comparável ao longo do tempo. Esse desenho tende a gerar dados mais consistentes, escaláveis e úteis para diferentes finalidades, substituindo a lógica de campanhas isoladas por uma rotina contínua de geração de evidência.

Também recomendamos **não utilizar o índice de Shannon como normalizador principal** para o cálculo de diversidade em dados moleculares. Para esse tipo de dado, faz mais sentido trabalhar com métricas ligadas à **riqueza** e à composição observada, preservando melhor o valor informacional da detecção por DNA ambiental. Na mesma lógica, sugerimos reduzir o número de pontos de coleta e aumentar a frequência amostral, pois, nesse contexto, séries temporais mais densas tendem a produzir mais valor do que campanhas extensas e esparsas. Esse monitoramento deve ser orientado por perguntas específicas que o DNA ambiental possa responder de forma objetiva, por exemplo detecção de mudança ambiental, comparação entre áreas, identificação de impacto ou acompanhamento de recuperação.

Por fim, entendemos que um futuro **crédito de biodiversidade** deve emergir da robustez dos dados produzidos por esse monitoramento frequente por DNA ambiental, e não de levantamentos pontuais. No médio prazo, sugerimos avaliar a instalação de **amostradores autônomos** nos locais onde haja interesse em monitoramento recorrente. Isso inverte a lógica atual, em que é preciso ir a campo a cada nova demanda, e cria uma infraestrutura permanente de geração de dados capaz de responder com mais eficiência a necessidades futuras, como certificação, gestão operacional e resolução de conflitos com stakeholders.

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

BLAST
Ferramenta de comparação de sequências de DNA ou proteína com bancos de dados genéticos, baseada no alinhamento entre a sequência analisada e sequências de referência. Em estudos de DNA ambiental, o BLAST é amplamente utilizado para identificar organismos presentes nas amostras, pois permite comparar sequências inteiras ou trechos longos com alta sensibilidade e detalhamento.

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.

Famílias gênicas
Conjuntos de genes semelhantes entre si, geralmente associados a funções biológicas relacionadas. Sua quantificação ajuda a descrever a diversidade funcional presente em uma comunidade microbiana.

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.

HUMAnN4

Programa de bioinformática utilizado para caracterizar o potencial funcional de comunidades microbianas a partir de dados de shotgun metagenômico. Ele permite identificar famílias gênicas, reações bioquímicas e vias metabólicas representadas nas amostras.

Kmers

Pequenos fragmentos de DNA de tamanho fixo obtidos a partir da divisão de uma sequência maior em partes menores e sobrepostas. Em análises de bioinformática, esses fragmentos são usados como assinaturas da sequência original, permitindo comparações rápidas com grandes bancos de dados genéticos, especialmente em abordagens de classificação taxonômica de dados de shotgun.

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.

Kraken2

Programa de bioinformática utilizado para classificar sequências de DNA em grupos taxonômicos, comparando-as com bancos de dados de referência. No contexto deste relatório, foi utilizado para gerar uma visão geral dos táxons detectados nos dados de shotgun.

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.

Leituras não-quiméricas

São as sequências que permanecem após a remoção de artefatos gerados durante a amplificação e o processamento bioinformático. Elas representam a fração de sequências de alta qualidade efetivamente disponível para classificação taxonômica.

QIIME2

Plataforma de bioinformática utilizada para processar dados de metabarcoding. Ela permite realizar etapas como controle de qualidade, correção de erros, agrupamento de sequências e classificação taxonômica dos organismos detectados.

Shotgun metagenômico

Abordagem de sequenciamento em que todo o DNA presente em uma amostra é analisado diretamente, sem amplificação prévia de marcadores específicos. Isso permite detectar uma ampla variedade de organismos e também inferir informações sobre o repertório gênico e funcional da comunidade.