UNIVERSIDADE FEDERAL DE GOIÁS (UFG)
INSTITUTO DE CIÊNCIAS BIOLÓGICAS (ICB)
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR (PPGBM)

LUCAS DONIZETTI VIEIRA

Metagenômica de solos em paisagens agrícolas no Cerrado

Goiânia
2022

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

[ ] Dissertação [x] Tese [ ] Outro*:_____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Lucas Donizetti Vieira

3. Título do trabalho

Metagenômica de solos em paisagens agrícolas no Cerrado

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento [ ] SIM [x] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:
a) consulta ao(à) autor(a) e ao(à) orientador(a);
b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
O documento não será disponibilizado durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.

Documento assinado eletronicamente por **Lucas Donizetti Vieira**, **Discente**, em 01/12/2022, às 14:01, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Rosane Garcia Collevatti**, **Professora do Magistério Superior**, em 06/12/2022, às 11:04, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3368952** e o código CRC **CA61FD29**.

Referência: Processo nº 23070.063891/2022-09 SEI nº 3368952

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

[ ] Dissertação [x] Tese [ ] Outro*:_____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Lucas Donizetti Vieira

3. Título do trabalho

Metagenômica de solos em paisagens agrícolas no Cerrado

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concordа com a liberação total do documento [x] SIM [ ] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:
a) consulta ao(à) autor(a) e ao(à) orientador(a);
b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
O documento não será disponibilizado durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.

Documento assinado eletronicamente por **Lucas Donizetti Vieira**, **Discente**, em 28/07/2025, às 11:54, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5529711** e o código CRC **3A881516**.

Referência: Processo nº 23070.063891/2022-09 SEI nº 5529711

LUCAS DONIZETTI VIEIRA

Metagenômica de solos em paisagens agrícolas no Cerrado

Tese apresentada ao Programa de Pós-graduação em Genética e Biologia Molecular, do Instituto de Ciências Biológicas, da Universidade Federal de Goiás, como requisito para obtenção do título de Doutor em Genética e Biologia Molecular.
Área de concentração: Genética e Biologia Molecular.
Linha de pesquisa: Genética de Populações e Evolução Molecular

Orientadora: Prof.ª Dr.ª Rosane Garcia Collevatti

Goiânia
2022

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Vieira, Lucas Donizetti
Metagenômica de solos em paisagens agrícolas no Cerrado [manuscrito] / Lucas Donizetti Vieira. - 2022.
70 f.: il.

Orientador: Profa. Dra. Rosane Garcia Collevatti.
Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Ciências Biológicas (ICB), Programa de Pós-Graduação em Genética e Biologia Molecular, Goiânia, 2022.
Bibliografia. Apêndice.

1. Estrutura da paisagem. 2. Crómio. 3. Diversidade taxonômica. 4. Archaea. 5. Bacteria. I. Collevatti, Rosane Garcia, orient. II. Título.

CDU 575

UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

ATA DE DEFESA DE TESE

Ata Nº **44** da sessão de Defesa de Tese de **Lucas Donizetti Vieira** que confere o título de Doutor(a) em **Genética e Biologia Molecular**, na área de concentração em **Genética e Biologia Molecular.**

Ao/s **vinte e oito dias do mês de novembro de dois mil e vinte e dois**, a partir da(s) **9h00**, no(a) **Sala de reuniões do CDIM/CIAMB e https://meet.google.com/ang-hxbt-uww**, realizou-se a sessão pública de Defesa de Tese intitulada **"Metagenômica de solos em paisagens agrícolas no Cerrado**". Os trabalhos foram instalados pelo(a) Orientador(a), Professor(a) Doutor(a) **Rosane Garcia Collevatti (ICB/UFG)]** com a participação dos demais membros da Banca Examinadora: Professor(a) Doutor(a) **Layara Alexandre Bessa (IFGoiano)**, membro titular externo; Professor(a) Doutor(a) **Luciana Cristina Vitorino (IFGoiano)**, membro titular externo; Professor(a) Doutor(a) **Clayton Luiz Borges (ICB/UFG)**, membro titular interno; Professor(a) Doutor(a) **Renata de Oliveira Dias (ICB/UFG)**, membro titular interno. Durante a argüição os membros da banca não fizeram sugestão de alteração do título do **trabalho**. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Tese tendo sido o candidato aprovado pelos seus membros. Proclamados os resultados pela Professora Doutora **Rosane Garcia Collevatti**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, ao(s) **vinte e oito dias do mês de novembro de dois mil e vinte e dois**.

TÍTULO SUGERIDO PELA BANCA

Documento assinado eletronicamente por **Rosane Garcia Collevatti**, **Professora do Magistério Superior**, em 28/11/2022, às 13:05, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Clayton Luiz Borges**, **Professor do Magistério Superior**, em 28/11/2022, às 13:05, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Renata De Oliveira Dias**, **Professora do Magistério Superior**, em 28/11/2022, às 13:05, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Alexandre Melo Bailao**, **Coordenador de Pós-graduação**, em 29/11/2022, às 18:39, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº

10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3363528** e o código CRC **A4F5CCCD**.

Referência: Processo nº 23070.063891/2022-09 SEI nº 3363528

Por todo o orgulho que sempre demonstrou pela minha trajetória acadêmica e por muito antes desse momento sempre me chamar de doutor, dedico este trabalho ao meu irmão Rodrigo Favorito Garcia. “*in Memoriam*”

O Grande Inquisidor está disposto a concordar com cristo – a única vez que o faz! – quando lhe diz que “se [...] alguém [...] dominasse a consciência [do homem] – oh, então ele até jogaria fora teu pão e seguiria aquele que seduzisse sua consciência. Nisso tinhas razão. Por que o segredo da existência humana não consiste em apenas viver, mas na finalidade de viver”.

Os irmãos Karamázov - Fiódor Dostoiévski.

Agradecimentos

Como de costume, meu primeiro agradecimento é para minha mãe Luzia Favoritto, pois foi ela quem mais me incentivou durante esses mais de dez anos de caminhada acadêmica. A sua luta, mesmo diante de tantos desafios, está representada em todos os meus títulos e conquistas e apesar da distância física a trago comigo nos meus pensamentos. Com sentimentos muito semelhantes não poderia deixar de agradecer à minha tia Maria Madalena Favoritto, pois ela também me incentivou desde o início e não mediu esforções para que eu tivesse tudo que fosse preciso para chegar até aqui.

Durante a minha caminhada acadêmica eu encontrei a minha companheira, Drª Mariane Brom Sobreiro, a quem me faltam palavras para agradecer. O amor que sinto por você é imensurável e seu companheirismo e ajuda são grandes responsáveis pela minha chegada até esse momento. Agradeço também pela construção do nosso lar junto aos nossos filhos pet Helena e Kyuubi.

Devo muitos agradecimentos aos meus amigos, eles são parte fundamental das minhas conquistas e me ajudaram das mais diversas formas durante os últimos anos. Dentre aqueles que fiz na universidade gostaria de dar um destaque para a Natácia Evangelista de Lima. Agradeço por cada conselho, sugestão e colaboração eles foram fundamentais para meu crescimento com cientista e professor. Agradeço também a minha amiga de mais longa data nessa caminhada Luiza Gabriela Fulgêncio de Lima, mesmo sendo tão espaçados nossos encontros sua amizade é muito importante para mim. Dos amigos que trago a mais tempo, gostaria de agradecer ao Wadson Vinicius Silva, a sua presença desde os momentos mais difíceis até os momentos de maior alegria foi sempre muito especial. Agradeço também a sua companheira Michelle

Santos por ter trazido tanta alegria a nosso grupo e por ter se tornado uma amiga tão querida.

A qualificação é uma parte muito importante no processo do doutorado e para mim foi um divisor de águas, pois tive a oportunidade de aprender muito com duas grandes pesquisadoras. Por isso gostaria de agradecer à Juliana Silveira dos Santos e à Renata de Oliveira Dias

Outro divisor de águas no meu processo de formação foi o doutorado sanduíche e seria difícil nomear e agradecer cada uma das pessoas que me auxiliou nesse processo, por isso gostaria de deixar meu agradecimento geral a todos. Em especial gostaria de agradecer à: minha supervisora na Universidade do Colorado em Boulder, Erin A. Manzitto-Tripp que me auxiliou em todas as etapas desde o primeiro momento, além de me acolher em seu laboratório e me apresentar diversas oportunidades. Aproveito para agradecer a Universidade do Colorado em Boulder por ter disponibilizado toda a estrutura necessária para minha formação durante esses seis meses. Dentre os diversos professores que tive a sorte de conhecer na universidade do Colorado não poderia deixar de agradecer a dois. Professor Nolan C. Kane e professora Stacey D. Smith por terem me acolhido de forma tão carinhosa em seus grupos de pesquisa e por terem me oferecido a oportunidade de colaborar, meu muito obrigado. Da minha estadia no Colorado gostaria de agradecer primeiramente a pessoa que me acolheu logo na minha chegada e que tanto me ajudou na conturbada primeira semana, Jennifer Nieves. Agradeço também a duas grandes amigas que fiz nesse lugar que por vezes foi tão hostil, Danish minha amiga durante os retornos para Thornton e Agnes minha amiga na universidade. Agradeço também aos meus amigos brasileiros Jacqueline e Lorenzo que me acolheram e me incluíram em todas as comemorações de festividades.

Gostaria de agradecer a todos os professores do programa de Pós-Graduação em Genética e Biologia Molecular em especial à professora Daniela de Melo e Silva que me acompanhou durante toda a pós-graduação e tanto me ajudou. Para a minha orientadora Rosane Garcia Collevatti devo tantos agradecimentos que precisaria de páginas para descrever cada situação em que me auxiliou e todas as oportunidades que me garantiu. Gostaria de agradecer por cada reunião e até pelas conversas durante um café, pude aprender tanto com você e levarei esses ensinamentos durante toda minha vida pessoal e profissional.

Agradeço também à Universidade Federal de Goiás, na qual estou desde 2016 e que me disponibilizou um enorme apoio logístico, possibilitando minha formação. Agradeço também ao programa de Pós Graduação em Genética e Biologia Molecular (PGBM), do qual sou aluno também desde 2016. Agradeço a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão tanto da bolsa no Brasil quanto da bolsa concedida para a realização do Programa de Doutorado-sanduíche no Exterior (PDSE). Agradeço também ao INCT (Institutos Nacionais de Ciência e Tecnologia) em Ecologia, Evolução e Conservação da Biodiversidade por gentilmente disponibilizar recursos computacionais para realização de análises.

Escrever o agradecimento de um doutorado se mostrou uma tarefa desafiadora para mim, foram quase cinco anos nessa etapa da minha formação acadêmica e muitas pessoas especiais passaram pela minha vida durante esse tempo. Aproveito essa introdução para agradecer aqueles que porventura esqueci, mas que estiveram presentes e de alguma forma contribuíram para esse trabalho, meu sincero muito obrigado.

Sumário

Resumo

VIEIRA, L. D. **Metagenômica de solos em paisagens agrícolas no Cerrado.** 2022. 70 f. Tese (Doutorado em Genética e Biologia Molecular), Instituto de Ciências Biológicas, Universidade Federal de Goiás, 2022.

A comunidade microbiana no solo é altamente diversa e é composta principalmente de bactérias e fungos, sendo, a maioria dessas espécies ainda não descrita. Nesse trabalho nós objetivamos compreender os efeitos da composição da paisagem na diversidade da microbiota do solo. Especificamente, nós analisámos os efeitos da paisagem e de variáveis do solo na diversidade taxonômica e funcional da microbiota em paisagens agrícolas no Cerrado. Com o programa Kraken 2 nós conseguimos classificar uma média de 7.600 espécies com mais de 10 mil OTU (unidade taxonômica operacional, no inglês) classificadas por sítio de coleta e o programa Bracken reclassificou essas sequências para uma média de aproximadamente 3.800 espécies por sítio. As sequências classificadas como Archaea foram na sua maioria atribuídas ao filo Euryarchaeota e as sequências bacterianas foram predominantemente atribuídas aos filos Proteobacteria e Actinobactéria. A classificação com o MG-RAST utilizando a base de dados SEED subsystems classificou mais de 8,2 milhões de reads por sítio dentro das 28 categorias funcionais. Quanto aos índices de diversidade alfa, variáveis do solo e de paisagem influenciaram negativamente a diversidade Shannon, assim como o índice de riqueza de Margalef, as diversidades filogenéticas MPD e MNTD, além de riqueza e abundância de Archaea e Bacteria. Para a diversidade beta, encontramos uma interação significativa com variáveis da paisagem de forma do fragmento (APSI, do inglês average patch shape index), porcentagem de pastagem, de habitat e de floresta, e contaminação por crómio (Cr). Considerando todos os nossos resultados, nós encontramos um perfil taxonômico de Archaea e Bacteria de solos alterado pela paisagem agrícola. Além disso, os altos

valores dos índices de diversidade alfa contrastando com os baixos valores dos índices de diversidade beta indicam alterações na riqueza das espécies e a sua heterogeneidade. Em conclusão, APSI e Cr são as duas variáveis preditivas que melhor explicam a diversidade e riqueza de Archaea e Bacteria em paisagens agrícolas no Cerrado. Nós verificámos que o uso intensivo da terra para a agricultura aumenta a riqueza e abundância de espécies de microrganismos do solo, reduzindo simultaneamente a sua diversidade.

Palavras-chave:

Estrutura da paisagem; Crómio; Diversidade taxonômica; Archaea; Bacteria.

Abstract

VIEIRA, L. D. **Soil metagenomics in agricultural landscapes in the Cerrado.** 2022. 70 s. Thesis (PhD in Genetics and Molecular Biology), Instituto de Ciências Biológicas, Universidade Federal de Goiás, 2022.

The microbial community in the soil is highly diverse and is composed mainly of bacteria and fungi with most of these taxa currently undescribed. Here we aimed to understand the effects of landscape composition on soil microbiota diversity. Specifically, we analyzed the effects of landscape and soil variables in the composition, taxonomic, phylogenetic and functional microbial diversity in Cerrado agricultural landscapes. Kraken 2 was able to classify a mean of 7,600 species with more 10 thousand OTU classification per site and Bracken reclassify it to a mean of approximately 3,800 species per site. Archaea sequences were mostly assigned to the highly diverse phylum Euryarchaeota and bacterial sequences were predominancy assigned to the phyla Proteobacteria and Actinobacteria. MG-RAST classification using the SEED subsystems database classified over 8.2 million reads per site within the 28 functional categories. Regarding alpha diversity indexes, soil and landscape variables negatively influenced Archaea and Bacteria species Shannon diversity index, Margalef richness index, MPD and MNTD phylogenetic diversities, richness and abundance. For beta diversity, we found significant interaction with landscape variables of (APSI), percentage of pasture, habitat and forest, and local variable chromium (Cr). Taken together, our results show a taxonomical profile for Archaea and Bacteria of disturbed soils caused by changes in the landscape use. Also, the high values of alpha diversity indexes contrasting with the low values of beta diversity indexes indicate changes in species richness and its heterogeneity. Lastly, APSI and Cr are the two predictive variables that better explained the diversity and richness of Archaea and Bacteria in the

Cerrado agricultural landscapes. We found that intensive land use for agriculture increases the species richness and abundance of soil microorganisms while reducing their diversity.

Keywords:

Landscape structure; Chromium; Taxonomic diversity; Archaea; Bacteria.

Introduction

Natural landscapes have been transformed into agricultural landscapes all over the world, with approximately 11% of the planet's land area covered by crops, and other 30% devoted to pasture (Raven & Wagner, 2021). Expansion and intensification of agriculture in tropical areas, marked by a continued loss and rapid alteration of old rainforests, woodlands and semi-arid environments have caused conflicts between food production and nature conservation (Laurance et al., 2014). Agricultural intensification may threaten biodiversity due to the loss of undisturbed habitats, landscape homogenization, and the use of agricultural chemicals (Medan et al., 2011). Agriculture intensification may also affect the soil microbiota community, altering the composition and abundance of Archaea and Bacteria, and thus affecting key ecosystem functioning (Zhou et al., 2021).

Soil microorganisms may play different roles in the ecosystem, from soil formation to nutrient input, affecting plant recruitment and growth (Schulz et al., 2013). The functional diversity of microbial communities can influence the stability, productivity, and resilience of the ecosystem in relation to stress and disturbance (Torsvik & Øvreås, 2002). Agricultural management might lead to a reduction in microbiome relative abundance (Ohigashi et al., 2021) and taxonomical and functional diversity (Muñoz-Arenas et al., 2020). The loss of microbial diversity in the soil can compromise key specialized functions such as organic matter decomposition (Singh et al., 2014) and reduction of heavy metal contamination (Wu et al., 2006; Kudo et al., 2013).

Vegetative ground cover can strongly influence the diversity and structure of the soil bacterial community (Chu et al., 2011), but land use changes have contrasting effects. For instance, the conversion of grassland to arable fields did not affect diversity

or the abundance of nitrogen-fixing bacteria (French et al., 2017), but the transformation of tropical forests in rubber plantation lead to an increase of Operational Taxonomic Unit (OUT) and Shannon index of bacterial community (Lan et al., 2017). Such effects could be explained by the fact that landscape composition can alters the microbiome composition. In particular, flowering cabbage fields promote the predominance of Gammaproteobacteria class while *Brassica* fields can lead to the predominance of Bacilli class (Kumar et al., 2022). Most studies have looked for changes in microbiome composition given the effect of local factors such as vegetation cover composition (Montagna et al., 2018; Keet et al., 2019), heavy metal contamination (Zhong et al., 2022), and fertilizer addition (Li et al., 2019). In particular, no studies have evaluated the effect of landscape composition on the structure and diversity of Archaea and Bacteria.

Soil nutrients and structure play important roles in microbial community composition and diversity (Wolińska et al., 2017), but this association may be altered due to agriculture intensification (Veldkamp et al., 2020). Nutrients like phosphorus and potassium change at forest patch edge and are influenced by the vegetation type and pedogenesis, respectively (Schröder & Fleig, 2017). In addition, agriculture intensification can also lead to soil heavy metal contamination due to the use of high amounts of agrochemicals (Slimane & El-Hafid, 2021). Heavy metal contamination changes microbial community composition and diversity, and also influences population density and typical activity of soil microbiome (Ashraf & Ali, 2007). For instance, sites exposed to hexavalent chromium (VI) experience a large reduction in bacterial diversity, greater than sites exposed to arsenic (Sheik et al., 2012).

The high heterogeneity in soil and vegetation types in the Cerrado ecoregion implies a diverse profile in microbiome composition and diversity (Procópio & Barreto,

2021). The Cerrado already lost more than half of its natural vegetation cover (Colli et al., 2020), such factor has an important influence on the microbiome diversity and composition. In this ecoregion a few studies show the predominance of the phylum Proteobacteria followed by Actinobacteria in pristine and disturbed soils respectively (Quirino et al., 2009). In Cerrado savannas the main genus found were Bradyrhizobium, Azospirillum and Rhizobium (Catão et al., 2014; Souza et al., 2016).

Here we address the effects of landscape structure and soil physicochemical characteristics on soil microbiota community in agricultural landscapes in the Cerrado ecoregion. We expect to find a soil microbiome (i) mainly composed by the phyla Proteobacteria, and (ii) with a functional profile consisting of a high number of reads assigned the categories of stress response and virulence, disease and defense. Because our sampling sites are predominantly surrounded by agricultural landscapes, and the change in vegetation cover can affect the functioning of the ecosystem (Potthast et al., 2012). We expect that abundance, richness and diversity of Archaea and Bacteria are positively influenced (iii) by soil fertility, regardless of the vegetation type; (iv) by habitat amount (% habitat, % savanna, and % forest); (v) by landscape compositional heterogeneity (i.e., Shannon Diversity Index; SHDI). And negatively affected (vi) by the presence of heavy metal in the soil; (vii) by land cover change (% pasture and % agriculture); (viii) by fragmentation (higher number of patches); (ix) by the increase of patch complexity (higher Average Patch Shape Index; Shape Mean); and (x) by agricultural matrix dominance.

Material and methods

Study area

The study area is part of the Long-Term Ecological Research (LTER) project called COFA-LTER (Functional Connectivity in an Agricultural Landscape), located in

Central-West Brazil in the Cerrado ecoregion (Figure 1). The area of this project is a mosaic composed of crops (mainly soybean and maize) and pasturelands (21.6%) interspersed by small remnants of natural vegetation, such as savannas (9.7%), riparian and seasonally dry forests (17.8%) and wetlands (1.1%, see details in Santos et al. 2021). We mapped the land cover in the COFA-LTER landscape using visual digitalization and manual classification of high-resolution images of Google Earth from 2019, freely available at the Geographic Information Systems QGis 2.4, with validation by field checking. The final map of the area corresponded to a 5 m spatial resolution, comprising 11 different land cover classes (Figure 1B): (i) water courses; (ii) savanna and open savanna; (iii) seasonal and riparian forests; (iv) wetland; (v) pasture; (vi) agriculture (corn or soybean); (vii) rural building; (viii) mining; (ix) urban area; (x) road and train rail, and (xi) Eucalyptus spp. plantation.

Soil sampling design

To survey the microbiome and estimate alpha and beta diversity we randomly selected 32 sampling sites in the COFA-LTER area (Figure 1B, Table S1). The sampling sites were in landscapes composed of different amounts of savanna, forests, agriculture and pasture (Figure 1B, Table S1).

In each sampling site we collected soil for metagenomics and physicochemical analyses. We first cleaned the sampling site by removing the leaf litter on the soil surface. Soil samples were collected using an alcohol-sterilized auger to a depth of approximately 20 centimeters. Any visible root, plant material, or rock was manually removed before sieving the collected soil (2 mm sieve). For metagenomics, at each site we collected the soil and stored it in four sterile centrifuge tubes. For physicochemical analyses, we collected 500 grams of soil and stored it in clean and dry plastic bags. Soil samples were collected in January/2020.

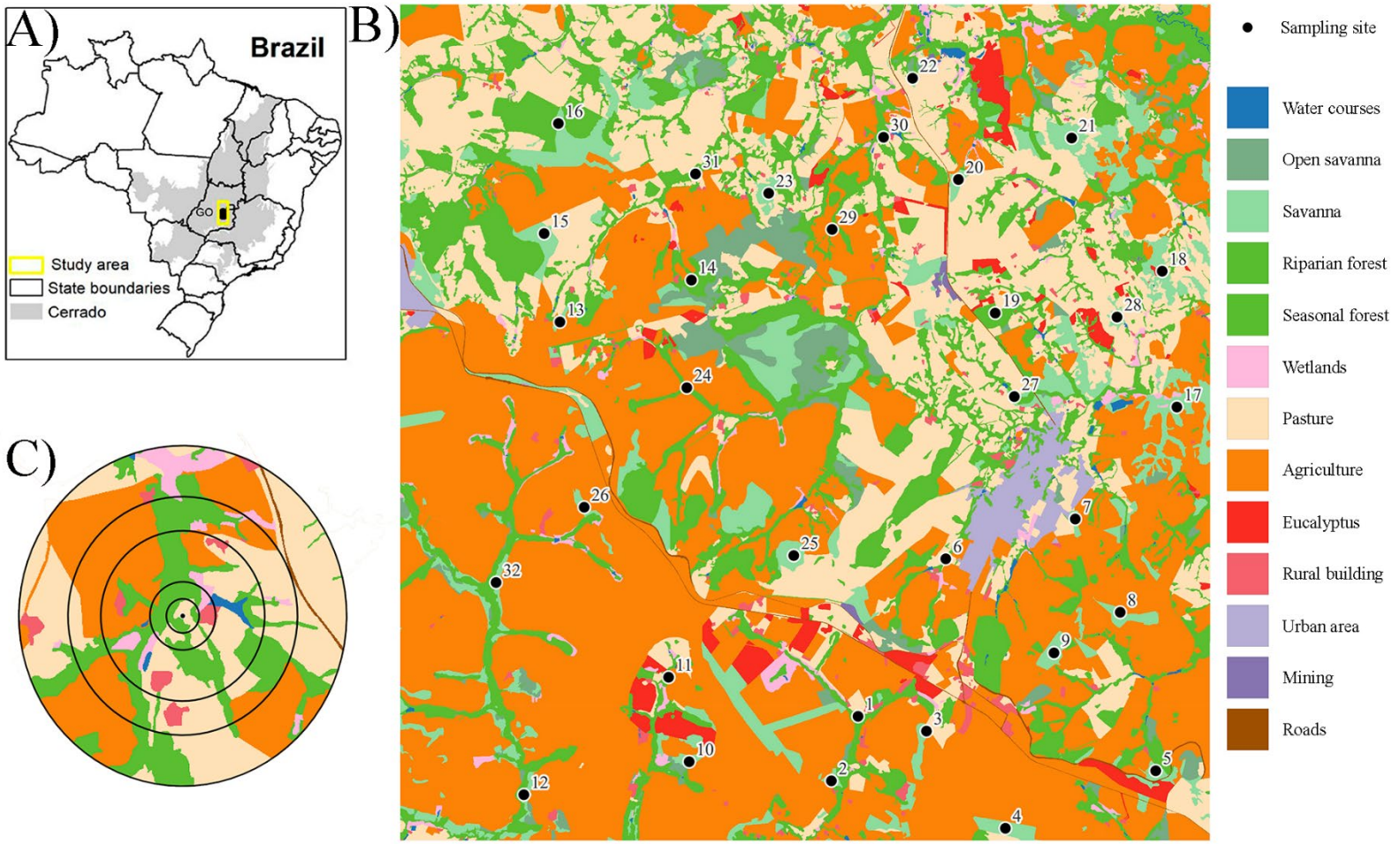

Figure 1. Geographical distribution of sampling sites in the COFA-LTER (Long-Term Ecological Research) study area. (a) The landscape in Central Brazil, in Goiás State (GO), within the Cerrado ecoregion (in gray). (b) Land cover composition. Black dots are the 32 sampling sites. (c) Example of multiscale approach used to calculate the landscape metrics, buffers of 100, 200, 500, 700, and 1000m.

Soil physicochemical analyses

For physicochemical analyses, we use the facility provided by SOLOCRIA Laboratory (Goiânia, BR). Soil physical composition was accessed by the amount of organic matter (g/dm3), percentage of clay, sand, and silt (g/Kg), and also the pH (CaCl2). For chemical composition, we obtained soil concentrations of cadmium (Cd, ppm), chromium (Cr, ppm), nickel (Ni, ppm), lead (Pb, ppm), exchangeable cations calcium (Ca2+, cmolc/dm3), magnesium (Mg2, cmolc/dm3), aluminum (Al3+, cmolc/dm3), and hydrogen (H+, cmolc/dm3), phosphorus (P, ppm), potassium (K, ppm), and zinc (Zn, ppm). In addition, we calculated cation exchange capacity (CEC, meq/100 ml soil) and percent base saturation (BS = [(Ca2+ + Mg2+ + K+)/CEC] × 100).

High-throughput DNA sequencing, raw data processing, and metagenome assembly

Soil DNA was extracted using the DNeasy PowerSoil Pro Kit (Qiagen, DK), following the manufacturer's instructions. DNA quality was checked and measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific, US) at A260/280 nm. DNA sequencing was carried out in Novogene Corporation Inc. (Sacramento, CA), using the Nextera XT DNA Library Preparation Kit (Illumina, US) and Illumina NovaSeq 6000 to generate paired-end 2x150-bp reads. All reads were trimmed for adapters using Trimmomatic v. 0.39 (Bolger et al., 2014) with the following parameters and instructions: -phred33; base quality encoding; the “TruSeq3-PE-2” fasta file with adapters to be removed; removing nine bases, regardless of quality, from the beginning of the read; scanning the read with a 4-base wide sliding window, cutting when the average quality per base drops below 15; and dropping reads which are less than 36 bases long after the above steps. Reads were checked before and after filtering with the

software FastQC v. 0.11.9 for quality control. With the filtered data we performed de novo shotgun assembly for each sampling site, using metaSPAdes with default parameters (Nurk et al., 2017).

Taxonomy and functional assignment

To assign the taxonomic profile to the sequences we analyzed the aligned metagenomes using Kraken 2 (Wood & Salzberg, 2014), with RefSeq Archaea and Bacteria database (accessed in April 2021). Then, we estimated taxa abundance for each sampling site with Bracken v. 2.7 (Lu et al., 2017), exclusively accepting taxa with more than 10 assigned sequences. Bracken files were generated for different taxa levels for further analysis. For results visualization we generated graphs using Pavian v.1.0 (Breitwieser & Salzberg, 2020). The results from all sampling sites were concatenated using the combine_reports tool implemented in the suite KrakenTools v. 1.2 (https://github.com/jenniferlu717/KrakenTools), and data visualization was done with Krona v. 2.8.1 (Ondov et al., 2011).

Sequence annotation was performed in the Metagenomics Rapid Annotation with Subsystems Technology online server (MG-RAST) v. 4.0.3 (Meyer et al., 2008), accessed in May 2022, using the reads filtered and free from adapters. We used default settings and removed all possible host-specific sequences (e.g., human and mouse). For the functional assignment we used the following databases: the Clusters of Orthologous Genes (COGs, Tatusov et al., 1997); the evolutionary genealogy of genes: Non-supervised Orthologous Groups eggNOG, Jensen et al., 2008); the KEGG ORTHOLOGY (KO, Kanehisa et al., 2016); and the SEED Subsystems (Overbeek et al., 2014). We also performed a taxonomical classification using the MG-RAST to compare with Kraken 2 classification.

The International Committee on Systematics of Prokaryotes (ICSP) changed bacteria and archaea phyla in the second half of 2021. Since we built our database before the changes, we used the previous phyla classification in our study (for more information see Oren et al., 2021; Oren & Garrity, 2021).

Analyses of microbial community alpha and beta diversities

Using the complete Archaea and Bacteria dataset, which were assembled using the tool combine_reports (see above), we calculated the richness as the number of species present in each site. Meanwhile, abundance was calculated as the number of sequences assigned for each species in each site. Alpha diversity (the mean diversity of Archaea and Bacteria species in each site) was calculated as both indexes of Shannon and Simpson that were estimated using the function diversity, included in the R package *vegan* v. 2.6-2 (Dixon, 2003). Then, we estimated species evenness (describes the abundance distribution among species in a community), using the evenness function implemented in the R package *tabula* v. 1.8.0. We also estimated Margalef, and Menhinick richness indexes using the functions margalef and menhinick from the R package *abdiv* v. 2.2.0 (https://github.com/kylebittinger/abdiv).

Next, we built a phylogenetic tree for further analysis using SED text editor and the KrakenSummary, a function inside of the python package *krakenplot* (https://github.com/clintval/krakenplot). We assessed phylogenetic alpha diversity for each site using the Archaea and Bacteria occurrence and the phylogenetic tree. We calculated the Mean Phylogenetic Distance (MPD) and Mean Nearest Taxon Distance (MNTD) using the *picante* v. 1.8.2 (Kembel et al., 2010) R package considering cophenetic phylogenetic tree and weighted abundance. We used phylogenetic and taxonomic diversity indexes as response variables related to the diversity and richness of Archaea and Bacteria.

To quantify the differences in microbiome composition between the pairs of sites, we estimate beta diversity using the Sørensen dissimilarity index (β_{SOR}) (Bryant et al., 2008). We also calculated beta diversity considering spatial turnover (β_{SIM}) and nestedness (β_{SNE}) components (Baselga, 2010). To perform these analyses, we first converted the abundance dataset into a presence-absence dataset, and used the *betapart* v. 1.5.6 R package (Baselga & Orme, 2012).

Multi-scale landscape variables

Because the scale-of-effect of landscape metrics in the soil microbiota community is unknown, we obtained multi-scale landscape metrics using concentric buffers around each sampling site (Jackson and Fahrig, 2015). We generated buffers of 100, 200, 500, 700, and 1000 m radii considering each sampling site as the central point of each landscape (Figure 1C). We calculated the percentage of habitat (as the total vegetation cover in each landscape, i.e., % of savanna + % forest), the percentage of savanna and the percentage of forests as proxies of habitat amount. To account for the effect of land cover changing, we calculated the percentage of pasture and the percentage of agriculture (mainly soybean and maize), the two major agricultural cover in the landscapes, and the dominant matrix corresponding to the matrix with a higher proportion in the landscape. The number of patches (NP) and average patch shape index (APSI) were calculated as proxies of fragmentation. APSI measures the complexity of the patch shape compared to a standard shape (square) of the same size, and increases without limit as the patch shape becomes more irregular (McGarigal et al., 2012). We calculated the shape index, considering all natural vegetation types (forests + savannas) as a unique class, and calculated the mean of these values for each landscape, ignoring the 0 values. When the landscape was entirely occupied by a single patch, then shape index = 1, assuming the patch is regular. To account for the effects of the diversity of

different land cover classes we calculated the landscape compositional heterogeneity using Shannon Diversity Index (SHDI) (Fahrig et al., 2011). All landscape metrics were calculated using Fragstats (McGarigal et al., 2012).

Statistical analysis

For alpha diversity analyses we first identified the scale-of-effect of the landscape variables as the scale where the model had the highest coefficient of determination (R2; Jackson and Fahrig, 2015). We performed a multiscale test of independence for multivariate vectors using the R multifit function (Huais 2018). The buffer sizes of 100 and 200 m radii had all zero inflated categories removed and then were tested for all landscape metrics in the downstream analysis. The remaining categories were SHDI, APSI, and percentage of habitat. Next, we verified the multicollinearity among the explanatory variables. For landscape variables we ran a full linear model for the selected variables in the scale-of-effect analysis and then estimated the variance inflation factor (VIF). We used a stepwise approach to eliminate models with VIFs > 3.0 (Zuur et al., 2010). For soil variables we calculated Pearson's correlation coefficient between all variables using the *corrplot* v. 0.92 R package (https://github.com/taiyun/corrplot), and eliminated variables with values > 0.5. We also removed the soil variables lead (Pb), nickel (Ni), and cadmium (Ca) due to the lack of variation among sampling sites. We also analyzed the spatial autocorrelation in response variables using Moran's I test implemented in the *ape* v. 5.6-2 R package (Paradis et al., 2004). We found no significant autocorrelation in any alpha diversity index (Table S2). We tested the error distribution of each response variable using the function descdist implemented in the package *fitdistrplus* v. 1.1-8.

Finally, we used Gaussian error distribution for diversity and richness indexes and negative binomial distribution for richness and abundance coupled with model

selection based on AIC (Akaike's information criterion) to select the most parsimonious models that explain alpha diversity using landscape metrics selected in the previous steps (the scale-of-effect and collinearity) and soil metrics separately. We expected different responses between savanna and forests due to differences in soil type and fertility, abbreviated hereafter as VT. Thus, for soil physicochemical variables we ran the model selection with interaction terms between soil variables and vegetation type. We also ran a null model (y ~ 1) composed only of the intercept of the response variable, representing a null effect of the predictors on the response variable (Burnham and Anderson, 2002). We estimated AIC corrected for small sample sizes (AICc), the difference between each model and the best model (ΔAICc), and the relative contribution of each model to explain the observed pattern (wAIC). Models with ΔAICc < 2 were considered as equally likely to explain our data (Zuur et al., 2009). Whether the null model was included as a plausible model, we consider no effect of landscape or soil variable on the response variable, since it is the model with the lower number of parameters (Crawley, 2012). After choosing the best models for landscape metrics and soil variables, we perform a joined analysis to select the predictive variable that best explains each response variable. For this, we ran AIC selection comparing only the best models for landscape and soil variables considering each alpha diversity index.

We analyzed the relationship of beta diversity, turnover and nestedness components coupled with landscape structure and soil variables using Multiple Matrix Regression with Randomization (MMRR) (Wang, 2013) implemented in the lgrMMRR function, part of the *PopGenReport* v. 2.2.2 R package (http://www.popgenreport.org/). To obtain the matrices of landscape and soil variables, we calculated the difference in landscape metrics at 100m spatial scale between each pair of sampling sites, and the difference in soil variables. We also used a geographical distance matrix to account for

spatial autocorrelation. We performed 10,000 random permutations to obtain the statistical significance of matrix correlations. All the steps from the section material and methods are summarized in the figure 2.

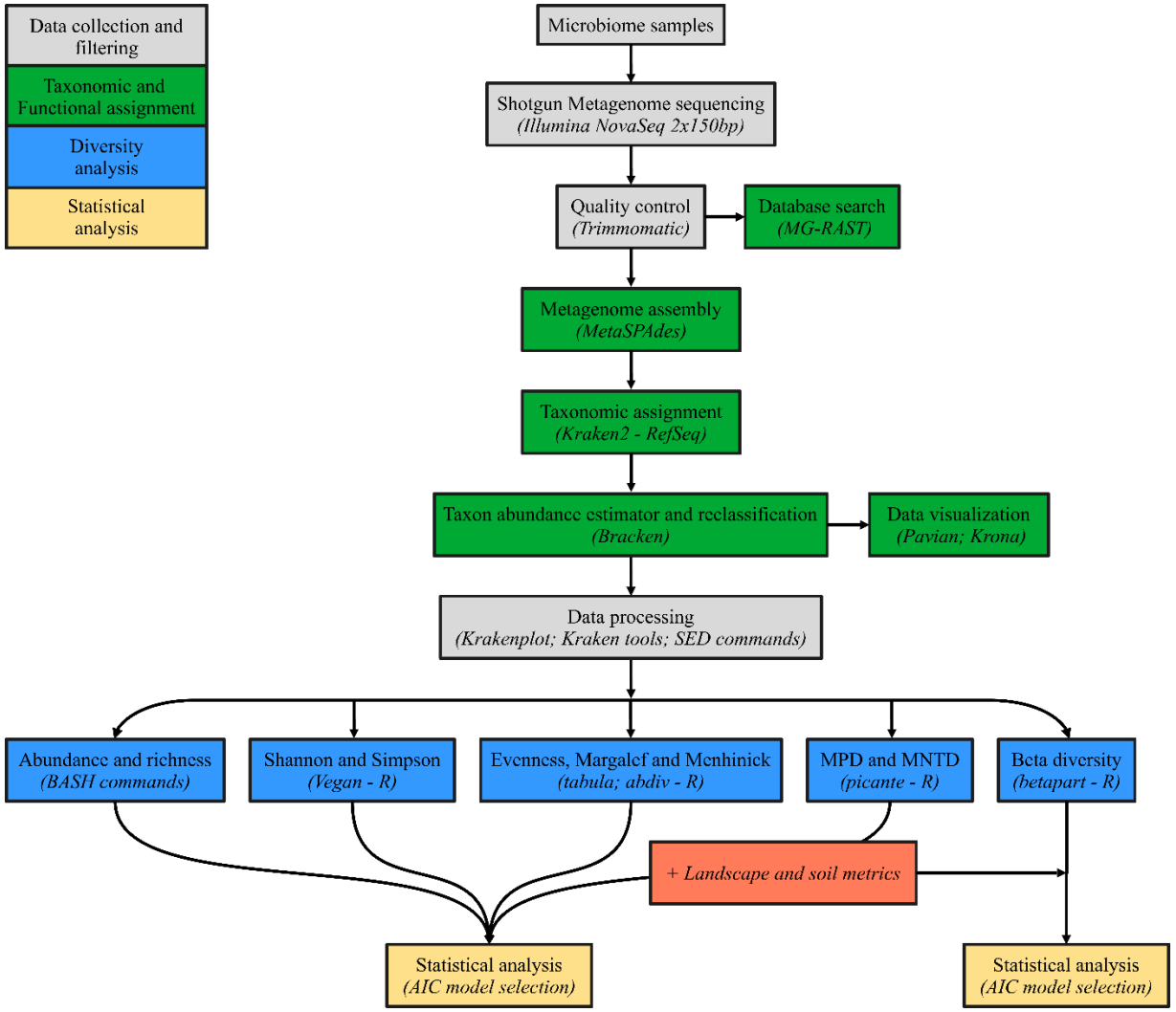

Figure 2. Workflow of analyses, each color indicates an analysis category.

Results

Taxonomy assignment

Whole genome sequencing of the 32 samples of Cerrado soil generated > 960 million paired reads or > 288 Gbp of sequence data (Table S3). After trimming and filtering > 946 million paired reads remained in the data set, a mean of > 29 million paired reads per sampling site (Table S3). Further on quality control, we obtained high quality raw data, with phred score ranging from 34 to 36, and the FastQC software mainly showed changes caused by the cut of adapters.

Kraken 2 revealed high percentage of novelty and high diversity in the soil microbiome, with c. 70% of sequences in each sample classified as unassigned, i.e., from unknown organisms. Kraken was able to classify a mean of 7,600 species per sampling site with more 10 thousand OTU (Operational Taxonomic Unit) classified per site. Bracken reclassified the species to a mean of c. 3,800 species per site (Figure S1 and Table S4). Bracken reclassification represents a species loss of 50%, following the threshold of 10 sequences assigned via Kraken 2. Assigned sequences represented two domains, Archaea and Bacteria, due to the construction of the database from RefSeq. Bacteria predominated in the assigned sequences, accounting for over 99% of the classified sequences in any sampling site (Table S4).

Most Archaea sequences were assigned to the highly diverse phylum Euryarchaeota, followed by the phylum Thaumarchaeota (syn. Nitrososphaerota) (Figure 3). The Natrialbaceae family was the most frequent among archaea families in the dataset along with Halorubraceae and Haloarculaceae. *Halorubrum* was the most abundant Archaea genus, and *Halobacterium hubeiense* the most abundant species.

Bacteria sequences were predominantly assigned to the phylum Proteobacteria (syn. Pseudomonadota) in 28 of the 32 sampling sites. Followed by Actinobacteria (syn.

Actinomycetota) that were the most assigned phylum in 04 remaining sites (Figure S2). At family level, most Bacteria sequences belonged to Bradyrhizobiaceae and Streptomycetaceae. Most sequences belonged to five bacteria species: *Bradyrhizobium diazoefficiens* (syn. *Bradyrhizobium japonicum*), followed by *Bradyrhizobium erythrophlei*, *Burkholderia ambifaria*, *Cupriavidus metallidurans*, and lastly *Delftia acidovorans* (Figure 4 and Table S4).

MG-RAST taxonomic classification was similar to Kraken 2 for domain, phylum, class, order, and family, but not at the genus level. At domain level, Archaea had lower percentage of reads classification in MG-RAST than in Kraken 2 (c. 0.3% in each sampling site). For Bacteria, although we found *Bradyrhizobioum* genus among the most common with MG-RAST classification, we found a higher number of genera classified in our samples with MG-RAST than with Kraken 2 (Figure 5). Contrasting with others sites, that had *Conexibacter* or *Bradyrhizobioum* as the most abundant genus in the sample, site 30 had the *Anaeromyxobacter* genus as the most abundant (Figure 5).

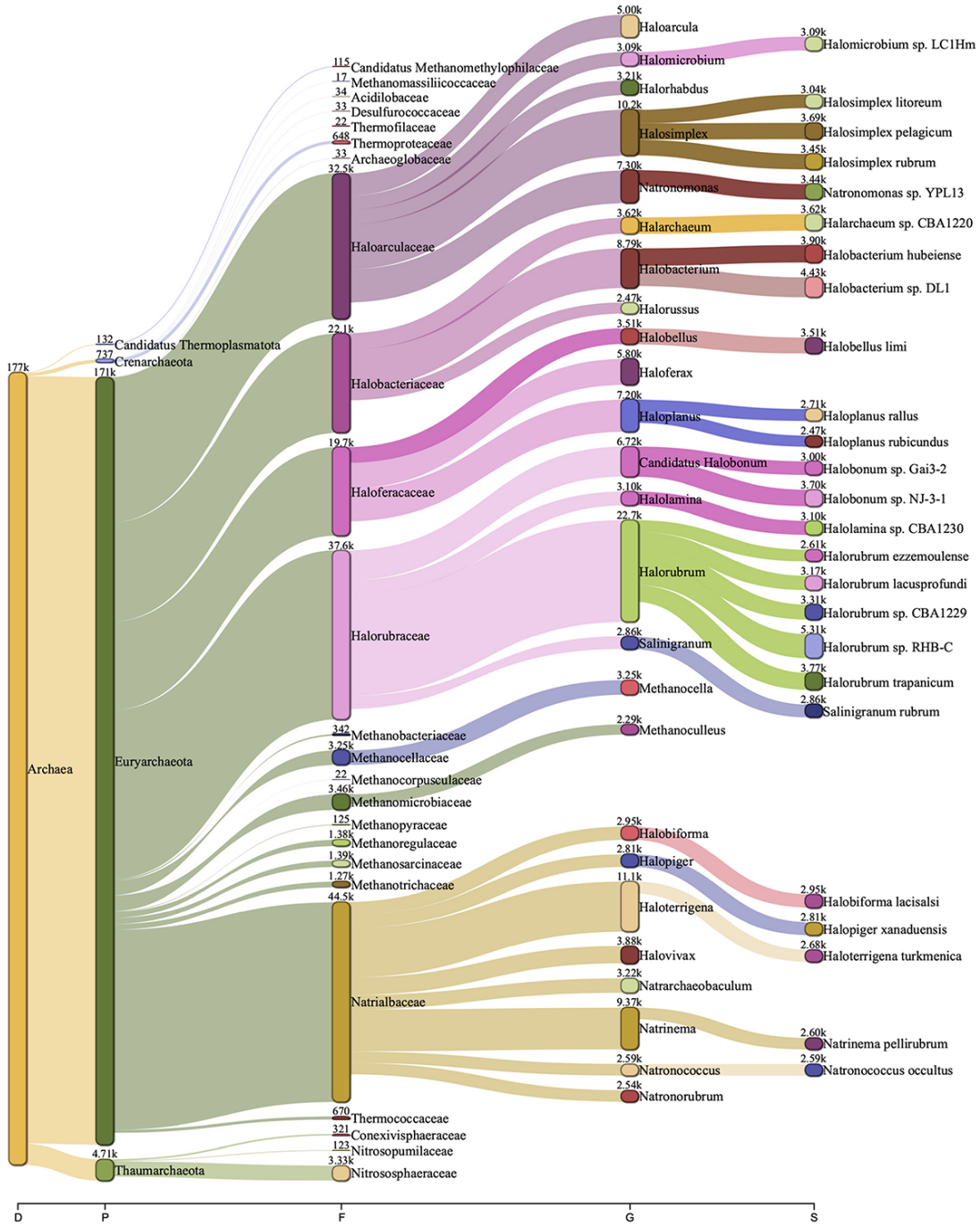

Figure 3. Classification of the most abundant representants at each taxonomy level for the domain Archaea. for the combined dataset of the 32 sampling sites. Above each bar is the number of sequences assigned to that taxonomic level.

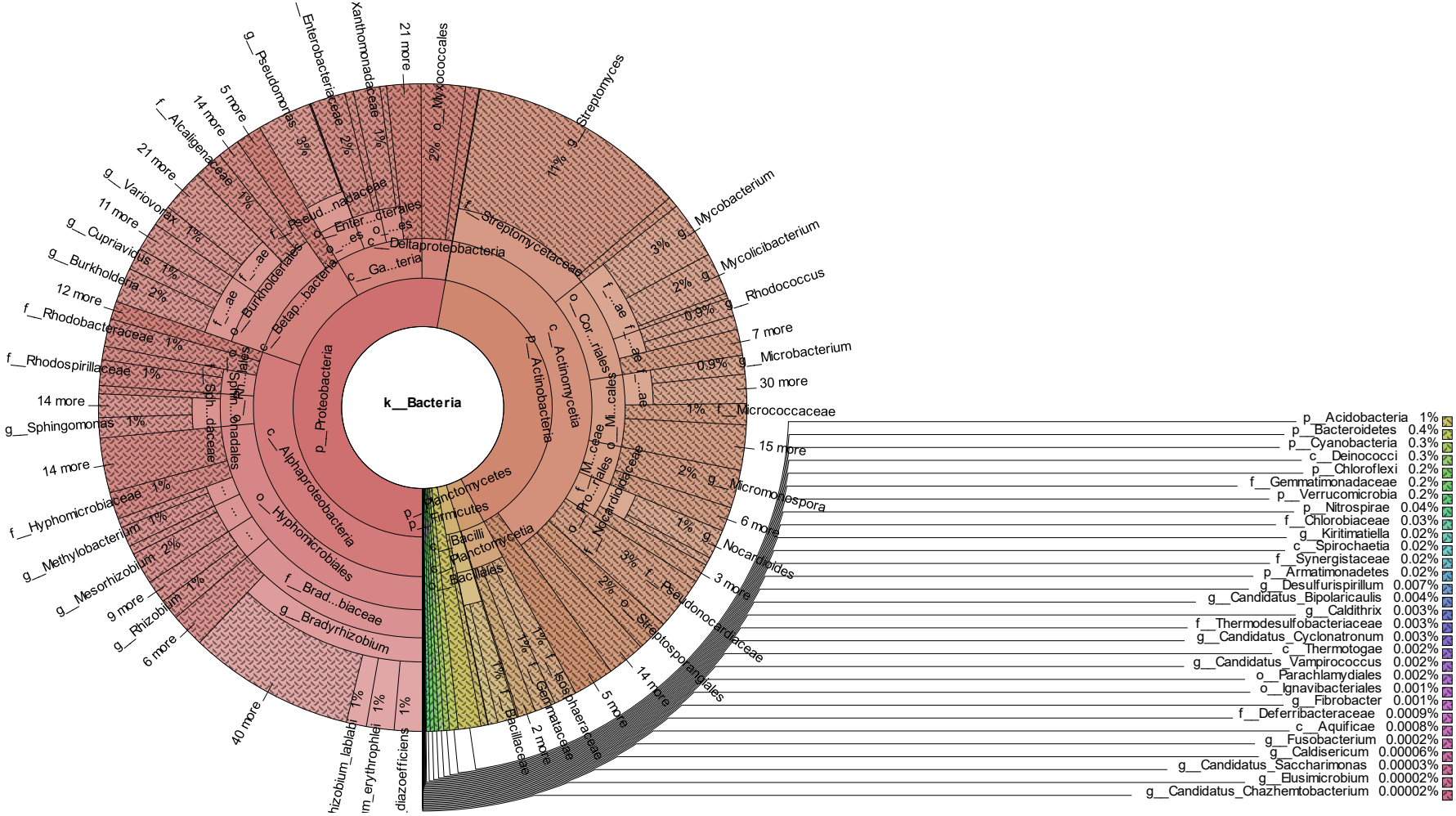

Figure 4. Bacteria classification using Kraken 2 and reclassification with Bracken for the combined dataset of the 32 sampling sites. An interactive version of this image is attached in html format.

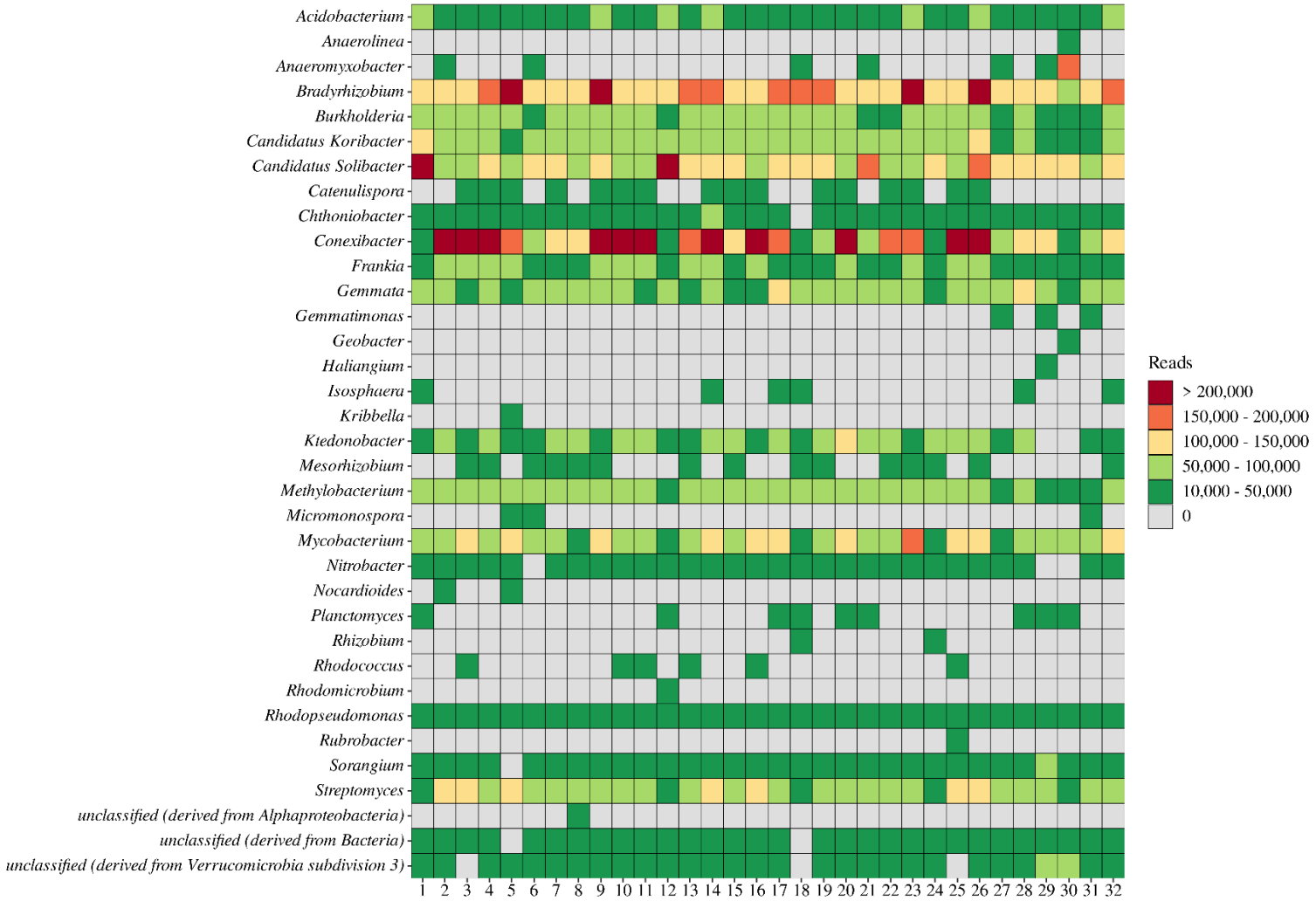

Figure 5. Number of sequences obtained for each genus in each sampling site. Classification was obtained MG-RAST.

Functional assignment

MG-RAST functional annotation defined 50% of the reads as unknow protein (~4.7 million reads), c. 49% were defined as annotated protein (~4.5 million reads), and 0.1% were defined as ribosomal RNA (~8 thousand reads).

For the COG database, MG-RAST classified a mean of ~4.3 million sequences per sampling site in the categories Metabolism (~1.9 million sequences), Cellular Processes and Signaling (~865,144 sequences), Information Storage and Processing (~694,555 sequences), and Poorly Characterized (~885,019) (Table S5 and Figure S3). The NOG database, which have the same categories as COG, had less sequences classified (approximately 257 thousand sequences for each site) and the majority of sequences was in the poorly characterized category (a mean of ~198,710 sequences) (Table S6 and Figure S4). Next, with the KEGG database, MG-RAST assigned a category for ~ 3 million sequences per sampling site, mostly belonging to metabolism category (a mean of ~1,8 million sequences) followed by environmental information processing (a mean of ~550 thousand sequences) (Table S7 and Figure S5). With 28 categories, the SEED subsystems database was the one with most sequences classified (a mean of ~ 8.2 million sequences). Most sequences were assigned to housekeeping functions (Table S8 and Figure 6).

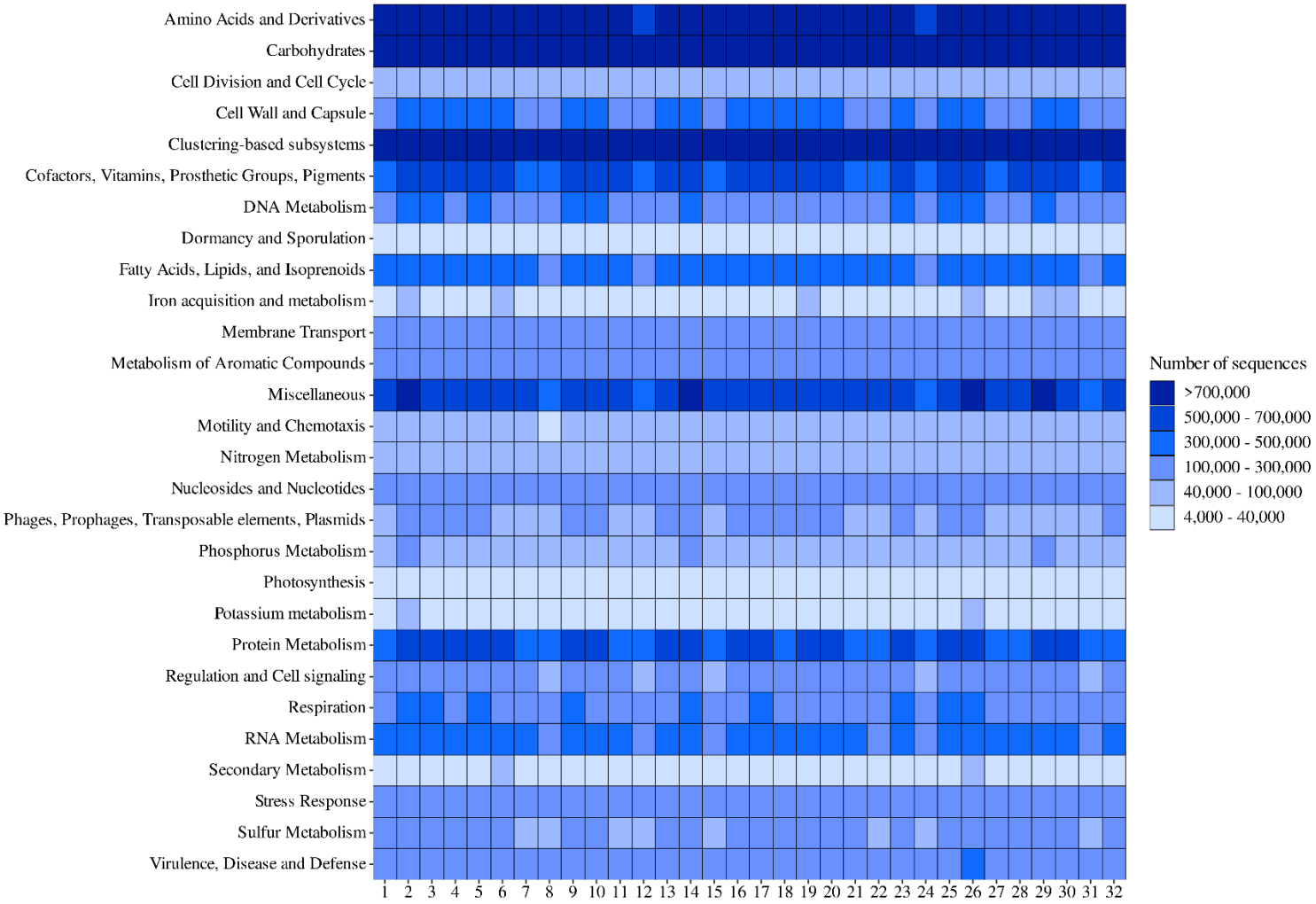

Figure 6. Subsystem classification with MG-RAST software.

Effects of landscape structure and soil on alpha diversity

Richness of Archaea and Bacteria was represented by a total of 4,676 species with a mean 3,757 species per sampling site (Table S9). Abundance, calculated as the number of sequences assigned for each species in each site, had a mean of approximately 1.3 million sequences assigned per sample. Shannon and Simpson diversities, measured using species presence and its abundance, had values ranging from 5.851 to 6.075 and from 0.993 to 0.995 respectively (Table S9). The Margalef richness index showed more variation compared to the other indexes measured, with values ranging from 75.644 to 92.850 (Table S9).

We found standard levels of lead, cadmium and nickel (heavy metals) in the sampling sites, based on the threshold determined by the Brazilian National Environment Council (CONAMA, resolution Nº 430 from 2011). Hexavalent chromium [Cr (VI)] was the exception, appearing with higher values than the recommended for agricultural sites in three different samples (Figure S6). As for the exchangeable cations capacity (CEC) and micronutrients, values were similar between sites whereas variation was only clear in calcium (Ca), and in a smaller scale for Zinc (Zn). Regarding CEC and base saturation (BS), the values varied similarly, with the higher values on both metrics found in collection sites of riparian forest (Figure S6). The soil texture, measured by composition (clay, silt, and sand), varied greatly for each collection site, with silt as the component less present (Figure S6). Soil pH ranged from 4.2 to 5.7, showing an acidic characteristic.

Overall, individual models showed that the landscape variables that affect microbiome community are average patch shape index (APSI), landscape compositional heterogeneity (SHDI), and percentage of forest (Table S10). Shannon diversity index were affected by APSI at 500m, Menhinick by percentage of forest at 700m, evenness,

Margalef and richness by APSI at 200m, and mean phylogenetic distance (MPD) by SHDI at 100m. Abundance and mean nearest taxon distance (MNTD) were equally affected by two different predictive landscape variables, APSI at 200m and 700m (abundance) and SHDI at 100m and APSI at 1000m (MNTD).

After removing soil variables with results higher than 0.5 in the multicollinearity matrix and variables with no variation, only seven from the original 18 variables remain (Figure S7). Between the most parsimonious models for local variables chromium (Cr) was the one that affects more response variables. For Simpson, Menhinick, and species evenness no model of soil variable was selected, since for these three the null model had results of ΔAICc <2 (Table S11). Four models from soil variables showed significant interaction with natural vegetation cover (ΔAICc < 2), for richness and Margalef we found small difference in the slope for forest and savanna. For abundance and MNTD the effects of the soil variable, hydrogen (H) and potassium (K) respectively, differ between natural vegetation cover (Figure S8 and Table S11).

Together, soil and landscape variables significantly influenced Archaea and Bacteria species Shannon diversity index, Margalef richness index, MPD and MNTD phylogenetic diversities, richness and abundance (Table 1). Specifically, APSI at 200m spatial scale ($b = 0.056$; $p < 0.001$) had a positive effect, and Cr, in savannas (slope = -0.031; $p = 0.017$), had a negative effect in Archaea and Bacteria richness (Figure 7a), the same was found for Margalef richness index (APSI slope = 0.927; $p < 0.001$ and Cr slope = -0.102.) (Figure 7d). APSI at 200m spatial scale (slope = 0.135; $p = 0.016$) and Cr (slope = 0.016; $p = 0.04$) had positive effects in the abundance of Archaea and Bacteria (Figure 7b). Shape Mean at 500m spatial scale (slope = 0.051; $p = 0.002$) and cation exchange capacity (CEC) (slope = 0.006; $p = 0.023$) had positive effects in Shannon diversity index (Figure 7c). SHDI at 100m spatial scale (slope = -0.172; p =

0.045) and phosphorus (P) (slope = -0.035; p = 0.049) negatively affected MPD (Figure 7e). Lastly, SHDI at 100m spatial scale (slope = -0.104; p = 0.049) and potassium K (slope = -0.350; p = 0.046) had negative effects in MNTD phylogenetic diversity (Figure 7f).

Table 1. Model selection for joined models of soil and landscape variables. MPD mean phylogenetic distance, MNTD mean nearest taxon distance, Cr Chromium, VT vegetation type cover, APSI Average Patch Shape Index, CEC cation exchange capacity, P phosphorus, K potassium, df degrees of freedom, ΔAICc difference between the AICc value of a given model and the model with the lowest ΔAICc value in each set.

Response variables	Model	df	ΔAICc	Model weight
Richness	~Cr*VT + APSI (200m)	9	0	0.995
Abundance	~Cr + APSI (700m)	4	0	0.935
Shannon	~CEC + APSI (500m)	4	0	0.992
Margalef	~Cr*VT + APSI (200m)	9	0	0.993
MPD	~P + SHDI (100m)	4	0	0.999
MNTD	~K + SHDI (100m)	4	0	0.967

Effects of landscape structure and soil on beta diversity

The extent of change in the composition of Archaea and Bacteria communities throughout the sampling sites, estimated by the Sørensen-based multiple-site dissimilarity index, ranged between 0.037 and 0.133 and showed higher differentiation among the 06, 27, 29, and 30 sampling sites (β_{SOR} = 0.518; Figure 8a). The turnover component accounted for the majority of the overall β diversity (β_{SIM} = 0.377; Figure 8b), compared to the nestedness component (β_{SNE} = 0.141; Figure 8c), meaning that variation in diversity between community was mainly due to the species replacement. The turnover component variation ranged from 0.010 to 0.042 and the nestedness component variation ranged from 0.001 to 0.061.

Average patch shape index ($p < 0.001$), % pasture (p = 0.022), and % vegetation cover (0.035) significantly explained the variation in beta diversity (R2 = 0.687) among sites (Table S12). For soil variables, only chromium (p = 0.027) significantly explained variation in beta diversity (R2 = 0.273) (Table S12). We found no effects of landscape

variables in the turnover component (Table S13). Percentage of sand ($p = 0.042$) and clay ($p = 0.047$) had significant effects in the turnover component ($R2 = 0.051$) (Table S13). However, APSI ($p < 0.001$, $R2 = 0.546$) and of Cr ($p = 0.049$, $R2 = 0.154$) significantly affected the nestedness component (Table S14).

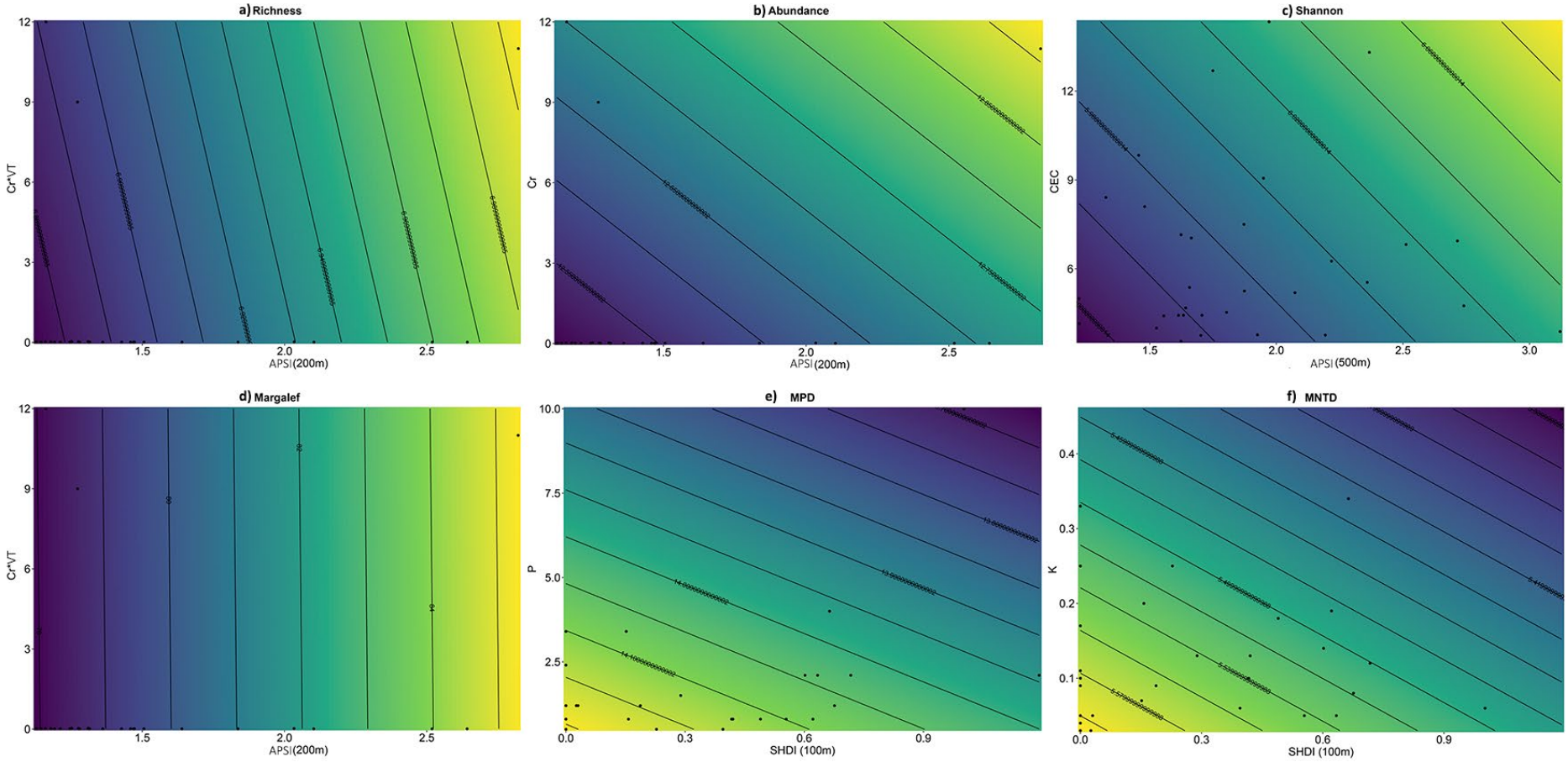

Figure 7. Relationship between soil and landscape variables and Archaea and Bacteria diversity. The dark blue-green-yellow color gradient in the response surfaces correspond to response variables values from low to high, respectively. Cr Chromium; Cr*VT Chromium accounting for different natural vegetation cover; APSI average patch shape index; SHDI compositional heterogeneity; MPD mean phylogenetic distance, MNTD mean nearest taxon distance.

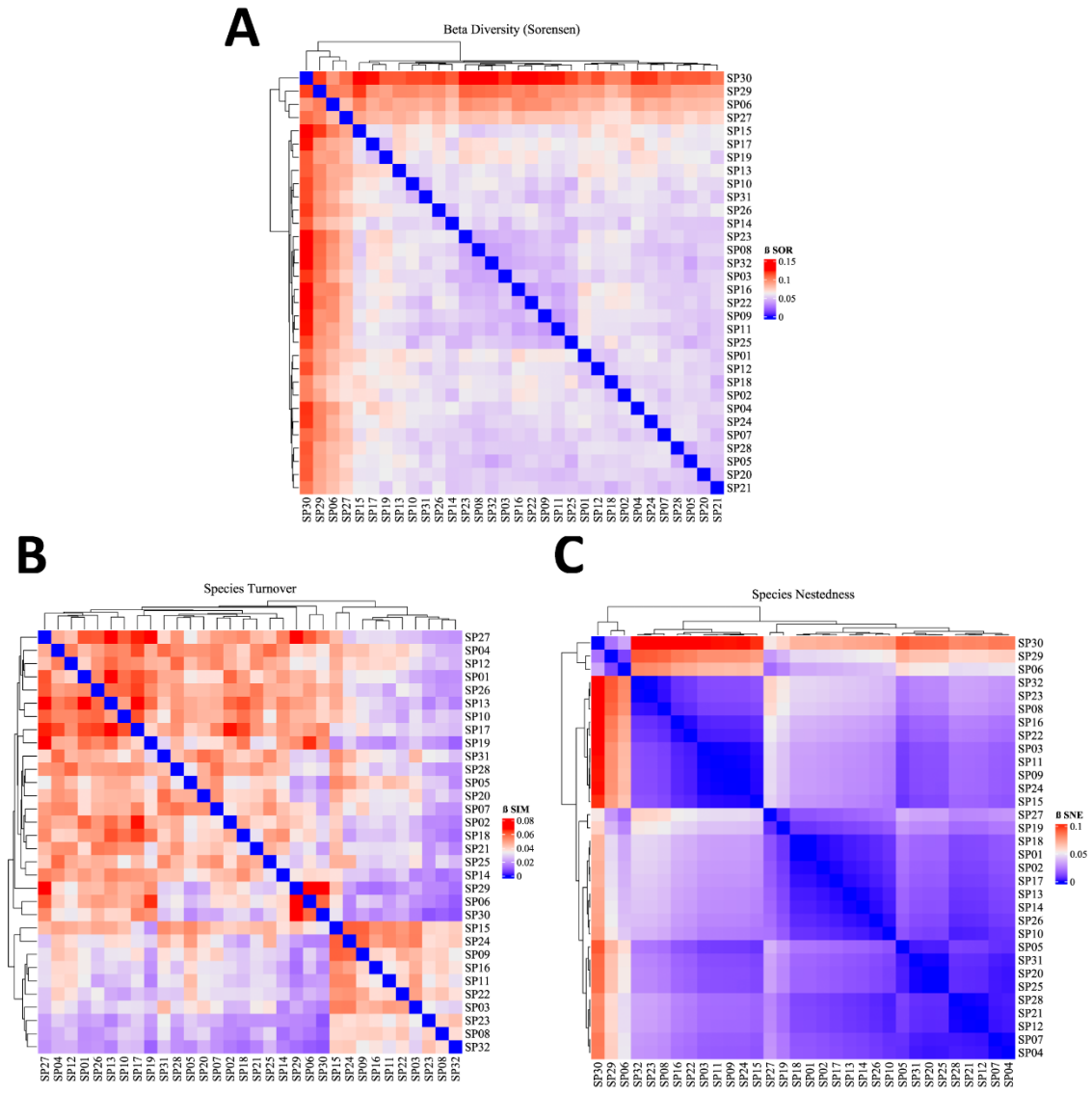

Figure 8. Beta diversity of Archaea and Bacteria between pairs of sampling sites in agricultural landscapes. A beta diversity; B turnover component; C nestedness component.

Discussion

Our results show a high percentage of novelty in taxonomical and functional classification, with c. 70% of unknow species and 50% of unknow protein or ribosomal RNA. Our results also indicate high alpha diversity and low beta diversity. This contrast between alpha and beta diversity might be an early sign of homogenization by altering the relationship among the average species richness and its heterogeneity (Hewitt et al., 2010). Our findings show that Archaea and Bacteria community structure and diversity (alpha and beta) are shaped by landscape structure and by heavy metal contamination.

Average patch shape index (APSI) positively affected abundance, richness, evenness, and alpha phylogenetic diversity (MNTD), and negatively affected Margalef's richness index and Shannon diversity index. Irregular patch shape reduces plant functional variability (Arellano-Rivas et al., 2018) and plant richness (Santos et al., 2022). Chromium concentration in soil also affected Archaea and Bacteria community structure, positively affecting abundance and richness and negatively affecting Margalef's richness index. The higher amount of Cr found in our results can be traced back to anthropogenic activities in the surrounds of the sampling sites. Contamination by Cr is widely present in soil due to its use as a pesticide to preserve wood. The presence of high levels of Cr can have disadvantageous effects on soil, affecting nitrification, aerobic nitrite oxidation, and the metabolic function of microbial communities (Li et al., 2020), also affecting their diversity and abundance (He et al., 2016).

Taxonomy assignment

The percentage of taxa classification using Kraken 2 (~ 30%) are within the expected, once taxonomical classification in metagenomics studies is determined mainly by the composition of the database (Nasko et al., 2018; Ye et al., 2019).

Although RefSeq database is in constant growth, with over 1,000 Archaea species and almost 100,000 Bacteria species, shotgun metagenomic studies usually classify ~ 30 to 60% of the sequences (Cissell & McCoy, 2021; Seitz et al., 2021). Approximately half of the species classified by Kraken 2 were lost in Bracken's reclassification because species with less than 10 sequences assigned were excluded. The Bayesian method implemented by Bracken reassign the results of Kraken 2 to mitigate the problem of database continuous update, resulting in a more accurate species-level assignment (Lu et al., 2017).

The Proteobacteria was the largest and most diverse phylum in our samples. Characterized by gram-negative bacteria, Proteobacteria has high morphological diversity, and its community structures can be used as a bacterial indicator for changes in soil environmental factors associated with land-use change (Kim et al., 2021). Proteobacteria tend to be dominant in the Cerrado soil, even in areas changed by anthropogenic activities (see more in Procópio & Barreto, 2021). Actinobacteria, another abundant phylum in the Cerrado soil (Pessoa-Filho et al., 2015; Souza et al., 2016), was also highly diverse in our samples. Interestingly, Actinobacteria is usually more abundant in tillage and pasture areas (Quirino et al., 2009; Souza et al., 2016).

We found a prevalence of *Bradyrhizobium* genus in our samples, following Kraken 2 classification. Although *Bradyrhizobium* is found in both native and disturbed environments the genus seems to be better adapted to agriculture areas (Procópio & Barreto, 2021). *Bradyrhizobium diazoefficiens* the most common species in our samples is a soybean N2-fixing endosymbiont (Itakura et al., 2008; Mathis et al., 1997). The predominance of this species was expected once our sampling sites are surrounding by soybean plantation and *B. diazoefficiens* is the most abundant rhizobial species in the soybean rhizosphere (Liu et al., 2018). *Bradyrhizobium erythrophlei*, the second most

abundant species, is also highly present in soil samples associated with plant roots (Michel et al., 2017; Yao et al., 2015), and together with *B. diazoefficiens* can be found in association with soybean rhizosphere (Liu et al., 2021).

The high concentration of heavy metal in our sampling sites may have affected the community structure leading to the high abundance of heavy metal tolerant species. For instance, *Cupriavidus metallidurans* the most abundant species in sites 15 and 23, is a heavy metal resistance species with several genetic adaptations and a complex transcriptional network related to heavy metal metabolism (Janssen et al., 2010; Monsieurs et al., 2011). Furthermore, *Anaeromyxobacter* was found in all sampling sites, but was the most abundant genus in the site 32. Composed by *Anaeromyxobacter dehalogenans*, *Anaeromyxobacter* has prominent role in reducing metals such as uranium and iron in contaminated soils (Sanford et al., 2002; Wu et al., 2006).

Functional assignment

The functional assignment for Bacteria and Archaea in the 32 sites resulted in different profiles relaying on the database, both for category and for number of sequences classified. Although we found significant differences in Bacterial composition between sites, especially the sites with higher levels of Cr, the same was not observed for most functional categories. These functions represent core resources such as cell growth, signal transduction, and material metabolism, so they might remain unchanged even under stress of soil Cr contamination (Liu et al., 2022).

Motility and chemotaxis categories were highly abundant in sites 31 and 32, that were close to streams in a riparian forest. Motility and chemotaxis genes tend to be more abundant at less structured soils close to water courses and high organic matter (Dini-Andreote et al., 2018), and play a role in multiple functions in microbes physiology, such as nutrient acquisition, expansion of population range and biofilm

formation (Colin et al., 2021). Most genes in chemotaxis and motility categories decreases in abundance with the land use conversion (Berkelmann et al., 2018, 2020).

Alpha diversity

To our best knowledge no study was conducted in order to search for correlations between soil microbiome and landscape structure. Our main goal with diversity and richness indexes was to look for the effects of soil and landscape variables in the taxonomic composition of Archaea and Bacteria in the soil of Cerrado agricultural landscapes. Alpha diversity was mainly affected by the vegetation shape (average patch shape index) at medium spatial scale (200 and 500 m), followed by the landscape compositional heterogeneity (SHDI).

We found that Archaea and Bacteria Shannon diversity decreases with the increase in irregular patches in the landscape. Irregular shapes, with angular edges (higher complexity), present smaller core habitat areas (Laurance & Yensen, 1991) and thereby, are less effective in protecting and preserving native species (Collinge, 1996). Phylogenetic diversity increased with the increase of compositional heterogeneity, supporting the mosaic concept proposed by Duelli 1997 and revised by Fahrig 2017. In fact, for the Cerrado ecoregion a more diversified mosaic leads to higher local plant diversity (Santos et al., 2022). Percentage of forest had a positive effect in the Menhinick richness index. Landscape composition, including higher proportions of forest, is found promoting the assemblage of Proteobacteria and Actinobacteria (Kumar et al., 2022).

We found that chromium concentration was the principal soil variable associated with richness and diversity of microbiome diversity in the agricultural landscapes. Our results also indicated that CEC, phosphorus, potassium and hydrogen have an influence on alpha diversity. These results corroborate previous studies that suggest CEC as a

main abiotic factor correlated with Protist community structure (Araujo et al., 2018), and that phosphorus and potassium are the two principal biogenic macronutrients and play important roles in microbial physiology (Chamberlain et al., 2020). Soil variables were associated with different vegetation cover for phylogenetic diversity and abundance. The type of vegetation cover alters soil properties and it is able to modify the composition and diversity of the soil bacterial community (Dinakaran et al. 2019). Changes in land use and land management also influence changes in bacterial community composition (Bevivino et al., 2014).

Beta diversity

The values of beta diversity based on Sørensen multiple-site dissimilarity index among sampling sites was similar to results found along an Alpine succession gradient and it is low compared to plant and fungal communities (Montagna et al., 2018). The low beta diversity may indicate a homogenization, with a low number of different species between sites. Similar results were found in land use change at the Amazon rainforest, with increase in alpha diversity and a decrease in beta diversity as the forest is converted in agricultural systems (Rodrigues et al., 2013). The sampling site 30 contributed disproportionately to the beta diversity most likely due to the higher concentration of chromium.

The turnover component accounted for 72.7% of the beta diversity, while the nestedness component accounted for 27.3%. The results implies that the changes between sites in Archaea and Bacteria are mainly driven by the replacement of some species by others (Baselga, 2010). Several studies have reported the changes in bacterial community via species turnover such as, in response to nitrogen enrichment (Liu et al., 2022), and under humic acid amendment (Li et al., 2019). Also, species turnover is

higher in heavy metal contaminated soils than in the non-contaminated soils (Zhong et al., 2022).

Conclusion

We found that intensive land use for agriculture increases the species richness and abundance of soil microorganisms while reducing their diversity. Changes in fragment shape, compositional heterogeneity of the landscape, and chromium contamination were the main factors that led to the loss of microbial diversity. Taking together our results show that rather than eliminating Archaea and Bacteria species number, land use changes are modifying community's structure through species turnover. Our results reinforce that fragmentation and habitat loss in the Cerrado ecosystem caused by land use change has altered the composition of the soil microbiota. The change in structure and diversity of Archaea and Bacteria leads to loss and/or change of fundamental soil functions such as nutrient cycling, thus affecting plant establishment and dispersal.

References

Arellano-Rivas, A., De-Nova, J. A., and Munguía-Rosas, M. A. (2018). Patch isolation and shape predict plant functional diversity in a naturally fragmented forest. Journal of Plant Ecology 11, 136–146. doi: 10.1093/JPE/RTW119.

Ashraf, R., and Ali, T. A. (2007). Effect of heavy metals on soil microbial community and mung beans seed germination. Pak J Bot 39, 629–636.

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. Global Ecology and Biogeography 19, 134–143. doi: 10.1111/J.1466-8238.2009.00490.X.

Baselga, A., and Orme, C. D. L. (2012). betapart: an R package for the study of beta diversity. Methods Ecol Evol 3, 808–812. doi: 10.1111/J.2041-210X.2012.00224.X.

Berkelmann, D., Schneider, D., Engelhaupt, M., Heinemann, M., Christel, S., Wijayanti, M., et al. (2018). How rainforest conversion to agricultural systems in Sumatra (Indonesia) affects active soil bacterial communities. Front Microbiol 9, 2381. doi: 10.3389/FMICB.2018.02381/BIBTEX.

Berkelmann, D., Schneider, D., Meryandini, A., and Daniel, R. (2020). Unravelling the effects of tropical land use conversion on the soil microbiome. Environmental Microbiomes 15, 1–18. doi: 10.1186/S40793-020-0353-3/FIGURES/9.

Bevivino, A., Paganin, P., Bacci, G., Florio, A., Pellicer, M. S., Papaleo, M. C., et al. (2014). Soil Bacterial Community Response to Differences in Agricultural Management along with Seasonal Changes in a Mediterranean Region. PLoS One 9, e105515. doi: 10.1371/JOURNAL.PONE.0105515.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. doi: 10.1093/bioinformatics/btu170.

Breitwieser, F. P., and Salzberg, S. L. (2020). Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics 36, 1303–1304. doi: 10.1093/BIOINFORMATICS/BTZ715.

Bryant, J. A., Lamanna, C., Morlon, H., Kerkhoff, A. J., Enquist, B. J., and Green, J. L. (2008). Microbes on mountainsides: Contrasting elevational patterns of bacterial and plant diversity. Proceedings of the National Academy of Sciences 105, 11505–11511. doi: 10.1073/PNAS.0801920105.

Catão, E. C. P., Lopes, F. A. C., Araújo, J. F., de Castro, A. P., Barreto, C. C., Bustamante, M. M. C., et al. (2014). Soil acidobacterial 16s rrna gene sequences reveal subgroup level differences between savanna-like cerrado and atlantic forest brazilian biomes. Int J Microbiol 2014. doi: 10.1155/2014/156341.

Chamberlain, L. A., Bolton, M. L., Cox, M. S., Suen, G., Conley, S. P., and Ané, J. M. (2020). Crop rotation, but not cover crops, influenced soil bacterial community composition in a corn-soybean system in southern Wisconsin. Applied Soil Ecology 154, 103603. doi: 10.1016/J.APSOIL.2020.103603.

Chu, H., Neufeld, J. D., Walker, V. K., and Grogan, P. (2011). The Influence of Vegetation Type on the Dominant Soil Bacteria, Archaea, and Fungi in a Low Arctic Tundra Landscape. Soil Science Society of America Journal 75, 1756–1765. doi: 10.2136/SSSAJ2011.0057.

Cissell, E. C., and McCoy, S. J. (2021). Shotgun metagenomic sequencing reveals the full taxonomic, trophic, and functional diversity of a coral reef benthic cyanobacterial mat from Bonaire, Caribbean Netherlands. Science of The Total Environment 755, 142719. doi: 10.1016/J.SCITOTENV.2020.142719.

Colin, R., Ni, B., Laganenka, L., and Sourjik, V. (2021). Multiple functions of flagellar motility and chemotaxis in bacterial physiology. FEMS Microbiol Rev 45, 1–19. doi: 10.1093/FEMSRE/FUAB038.

Colli, G. R., Vieira, C. R., and Dianese, J. C. (2020). Biodiversity and conservation of the Cerrado: recent advances and old challenges. Biodivers Conserv 29, 1465–1475. doi: 10.1007/S10531-020-01967-X/FIGURES/3.

Collinge, S. K. (1996). Ecological consequences of habitat fragmentation: implications for landscape architecture and planning. Landsc Urban Plan 36, 59–77. doi: 10.1016/S0169-2046(96)00341-6.

Dinakaran, J., Vikram, K., Hanief, M., Bidalia, A., Tambat, S., and Rao, K. S. (2019). Changes in vegetation cover and soil intrinsic properties influence the soil bacterial community composition and diversity across different climatic regions of India. Vegetos 2019 32:3 32, 288–302. doi: 10.1007/S42535-019-00027-2.

Dini-Andreote, F., van Elsas, J. D., Olff, H., and Salles, J. F. (2018). Dispersal-competition tradeoff in microbiomes in the quest for land colonization. Scientific Reports 2018 8:1 8, 1–9. doi: 10.1038/s41598-018-27783-6.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. Journal of Vegetation Science 14, 927–930. doi: 10.1111/J.1654-1103.2003.TB02228.X.

Fahrig, L. (2017). Ecological Responses to Habitat Fragmentation Per Se. https://doi.org/10.1146/annurev-ecolsys-110316-022612 48, 1–23.

Fahrig, L., Baudry, J., Brotons, L., Burel, F. G., Crist, T. O., Fuller, R. J., et al. (2011). Functional landscape heterogeneity and animal biodiversity in agricultural landscapes. Ecol Lett 14, 101–112. doi: 10.1111/J.1461-0248.2010.01559.X.

French, K. E., Tkacz, A., and Turnbull, L. A. (2017). Conversion of grassland to arable decreases microbial diversity and alters community composition. Applied Soil Ecology 110, 43–52. doi: 10.1016/J.APSOIL.2016.10.015.

He, Z., Hu, Y., Yin, Z., Hu, Y., and Zhong, H. (2016). Microbial Diversity of Chromium-Contaminated Soils and Characterization of Six Chromium-Removing Bacteria. Environ Manage 57, 1319–1328. doi: 10.1007/S00267-016-0675-5/TABLES/3.

Hewitt, J., Thrush, S., Lohrer, A., and Townsend, M. (2010). A latent threat to biodiversity: Consequences of small-scale heterogeneity loss. Biodivers Conserv 19, 1315–1323. doi: 10.1007/S10531-009-9763-7/FIGURES/3.

Huais, P. Y. (2018). multifit: an R function for multi-scale analysis in landscape ecology. Landsc Ecol 33, 1023–1028. doi: 10.1007/S10980-018-0657-5/FIGURES/2.

Itakura, M., Saeki, K., Omori, H., Yokoyama, T., Kaneko, T., Tabata, S., et al. (2008). Genomic comparison of *Bradyrhizobium japonicum* strains with different symbiotic

nitrogen-fixing capabilities and other Bradyrhizobiaceae members. The ISME Journal 2009 3:3 3, 326–339. doi: 10.1038/ismej.2008.88.

Jackson, H. B., and Fahrig, L. (2012). What size is a biologically relevant landscape? Landsc Ecol 27, 929–941. doi: 10.1007/s10980-012-9757-9.

Janssen, P. J., van Houdt, R., Moors, H., Monsieurs, P., Morin, N., Michaux, A., et al. (2010). The Complete Genome Sequence of *Cupriavidus metallidurans* Strain CH34, a Master Survivalist in Harsh and Anthropogenic Environments. PLoS One 5, e10433. doi: 10.1371/JOURNAL.PONE.0010433.

Jensen, L. J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., et al. (2008). eggNOG: automated construction and annotation of orthologous groups of genes. Nucleic Acids Res 36, D250–D254. doi: 10.1093/NAR/GKM796.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44, D457–D462. doi: 10.1093/NAR/GKV1070.

Keet, J. H., Ellis, A. G., Hui, C., and le Roux, J. J. (2019). Strong spatial and temporal turnover of soil bacterial communities in South Africa's hyperdiverse fynbos biome. Soil Biol Biochem 136, 107541. doi: 10.1016/J.SOILBIO.2019.107541.

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., et al. (2010). Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26, 1463–1464. doi: 10.1093/BIOINFORMATICS/BTQ166.

Kim, H. S., Lee, S. H., Jo, H. Y., Finneran, K. T., and Kwon, M. J. (2021). Diversity and composition of soil Acidobacteria and Proteobacteria communities as a bacterial indicator of past land-use change from forest to farmland. Science of The Total Environment 797, 148944. doi: 10.1016/J.SCITOTENV.2021.148944.

Kudo, K., Yamaguchi, N., Makino, T., Ohtsuka, T., Kimura, K., Dong, D. T., et al. (2013). Release of arsenic from soil by a novel dissimilatory arsenate-reducing bacterium, *anaeromyxobacter* sp. strain PSR-1. Appl Environ Microbiol 79, 4635–4642. doi: 10.1128/AEM.00693-13/SUPPL_FILE/ZAM999104565SO1.PDF.

Kumar, U., Saqib, H. S. A., Islam, W., Prashant, P., Patel, N., Chen, W., et al. (2022). Landscape Composition and Soil Physical–Chemical Properties Drive the Assemblages of Bacteria and Fungi in Conventional Vegetable Fields. Microorganisms 10, 1202. doi: 10.3390/MICROORGANISMS10061202/S1.

Lan, G., Li, Y., Jatoi, M. T., Tan, Z., Wu, Z., and Xie, G. (2017). Change in Soil Microbial Community Compositions and Diversity Following the Conversion of Tropical Forest to Rubber Plantations in Xishuangbanan, Southwest China. Trop Conserv Sci 10. doi: 10.1177/1940082917733230/ASSET/IMAGES/LARGE/10.1177_1940082917733230-FIG2.JPEG.

Laurance, W. F., Sayer, J., and Cassman, K. G. (2014). Agricultural expansion and its impacts on tropical nature. Trends Ecol Evol 29, 107–116. doi: 10.1016/j.tree.2013.12.001.

Laurance, W. F., and Yensen, E. (1991). Predicting the impacts of edge effects in fragmented habitats. Biol Conserv 55, 77–92. doi: 10.1016/0006-3207(91)90006-U.

Li, C., Quan, Q., Gan, Y., Dong, J., Fang, J., Wang, L., et al. (2020). Effects of heavy metals on microbial communities in sediments and establishment of bioindicators based on microbial taxa and function for environmental monitoring and management. Science of The Total Environment 749, 141555. doi: 10.1016/J.SCITOTENV.2020.141555.

Li, P., Liu, J., Jiang, C., Wu, M., Liu, M., and Li, Z. (2019). Distinct Successions of Common and Rare Bacteria in Soil Under Humic Acid Amendment – A Microcosm Study. Front Microbiol 10, 2271. doi: 10.3389/FMICB.2019.02271/BIBTEX.

Liu, H., Yuan, R., Sarkodie, E. K., Tang, J., Jiang, L., Miao, B., et al. (2022a). Insight into functional microorganisms in wet–dry conversion to alleviate the toxicity of chromium fractions in red soil. Front Microbiol 13, 977171–977171. doi: 10.3389/FMICB.2022.977171/FULL.

Liu, L., Chen, X., Hu, S., Zhan, Q., and Peng, W. (2021). Genetic diversity and distribution of rhizobia associated with soybean in red soil in Hunan Province. Arch Microbiol 203, 1971–1980. doi: 10.1007/S00203-020-02120-6/TABLES/2.

Liu, W., Yang, X., Jiang, L., Guo, L., Chen, Y., Yang, S., et al. (2022b). Partitioning of beta-diversity reveals distinct assembly mechanisms of plant and soil microbial communities in response to nitrogen enrichment. Ecol Evol 12, e9016. doi: 10.1002/ECE3.9016.

Liu, Y. H., Wang, E. T., Jiao, Y. S., Tian, C. F., Wang, L., Wang, Z. J., et al. (2018). Symbiotic characteristics of *Bradyrhizobium diazoefficiens* USDA 110 mutants associated with shrubby sophora (*Sophora flavescens*) and soybean (*Glycine max*). Microbiol Res 214, 19–27. doi: 10.1016/J.MICRES.2018.05.012.

Lu, J., Breitwieser, F. P., Thielen, P., and Salzberg, S. L. (2017). Bracken: Estimating species abundance in metagenomics data. PeerJ Comput Sci 2017. doi: 10.7717/peerj-cs.104.

Mathis, J. N., McMillin, D. E., Champion, R. A., and Hunt, P. G. (1997). Genetic Variation in Two Cultures of *Bradyrhizobium japonicum* 110 Differing in Their Ability to Impart Drought Tolerance to Soybean. Current Microbiology 1997 35:6 35, 363–366. doi: 10.1007/S002849900269.

McGarigal, K., Cushman, S., and Ene, E. (2012). FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst.

Medan, D., Torretta, J. P., Hodara, K., de la Fuente, E. B., and Montaldo, N. H. (2011). Effects of agriculture expansion and intensification on the vertebrate and invertebrate diversity in the Pampas of Argentina. Biodivers Conserv 20, 3077–3100. doi: 10.1007/s10531-011-0118-9.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., et al. (2008). The metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9, 1–8. doi: 10.1186/1471-2105-9-386/FIGURES/4.

Michel, D. C., Passos, S. R., Simões-Araujo, J. L., Baraúna, A. C., da Silva, K., Parma, M. M., et al. (2017). *Bradyrhizobium centrolobii* and *Bradyrhizobium macuxiense* sp. nov. isolated from *Centrolobium paraense* grown in soil of Amazonia, Brazil. Arch Microbiol 199, 657–664. doi: 10.1007/S00203-017-1340-Y/FIGURES/3.

Monsieurs, P., Moors, H., van Houdt, R., Janssen, P. J., Janssen, A., Coninx, I., et al. (2011). Heavy metal resistance in *Cupriavidus metallidurans* CH34 is governed by an intricate transcriptional network. BioMetals 24, 1133–1151. doi: 10.1007/S10534-011-9473-Y/FIGURES/4.

Montagna, M., Berruti, A., Bianciotto, V., Cremonesi, P., Giannico, R., Gusmeroli, F., et al. (2018). Differential biodiversity responses between kingdoms (plants, fungi, bacteria and metazoa) along an Alpine succession gradient. Mol Ecol 27, 3671–3685. doi: 10.1111/MEC.14817.

Muñoz-Arenas, L. C., Fusaro, C., Hernández-Guzmán, M., Dendooven, L., Estrada-Torres, A., and Navarro-Noya, Y. E. (2020). Soil microbial diversity drops with land-

use change in a high mountain temperate forest: a metagenomics survey. Environ Microbiol Rep 12, 185–194. doi: 10.1111/1758-2229.12822.

Nasko, D. J., Koren, S., Phillippy, A. M., and Treangen, T. J. (2018). RefSeq database growth influences the accuracy of k-mer-based lowest common ancestor species identification. Genome Biol 19, 1–10. doi: 10.1186/S13059-018-1554-6/FIGURES/5.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). MetaSPAdes: A new versatile metagenomic assembler. Genome Res 27, 824–834. doi: 10.1101/gr.213959.116.

Ohigashi, T., Schloter, M., Schulz, S., Munthali, K., and Uchida, Y. (2021). Changes in soil prokaryotic diversity in response to land-use changes in Sub-Saharan Africa. Soil Syst 5, 62. doi: 10.3390/SOILSYSTEMS5040062/S1.

Ondov, B. D., Bergman, N. H., and Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. BMC Bioinformatics 12, 1–10. doi: 10.1186/1471-2105-12-385/FIGURES/4.

Oren, A., Arahal, D. R., Rosselló-Móra, R., Sutcliffe, I. C., and Moore, E. R. B. (2021). Emendation of rules 5b, 8, 15 and 22 of the international code of nomenclature of prokaryotes to include the rank of phylum. Int J Syst Evol Microbiol 71, 004851. doi: 10.1099/IJSEM.0.004851/CITE/REFWORKS.

Oren, A., and Garrity, G. M. (2021). Valid publication of the names of forty-two phyla of prokaryotes. Int J Syst Evol Microbiol 71, 005056. doi: 10.1099/IJSEM.0.005056/CITE/REFWORKS.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42, D206–D214. doi: 10.1093/NAR/GKT1226.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20, 289–290. doi: 10.1093/BIOINFORMATICS/BTG412.

Pessoa-Filho, M., Barreto, C. C., dos Reis Junior, F. B., Fragoso, R. R., Costa, F. S., de Carvalho Mendes, I., et al. (2015). Microbiological functioning, diversity, and structure of bacterial communities in ultramafic soils from a tropical savanna. Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology 107, 935–949. doi: 10.1007/S10482-015-0386-6/FIGURES/6.

Procópio, L., and Barreto, C. (2021). The soil microbiomes of the Brazilian Cerrado. J Soils Sediments 21, 2327–2342. doi: 10.1007/S11368-021-02936-9/TABLES/1.

Quirino, B. F., Pappas, G. J., Tagliaferro, A. C., Collevatti, R. G., Neto, E. L., da Silva, M. R. S. S., et al. (2009). Molecular phylogenetic diversity of bacteria associated with soil of the savanna-like Cerrado vegetation. Microbiol Res 164, 59–70. doi: 10.1016/J.MICRES.2006.12.001.

Raven, P. H., and Wagner, D. L. (2021). Agricultural intensification and climate change are rapidly decreasing insect biodiversity. Proc Natl Acad Sci U S A 118, e2002548117. doi: 10.1073/PNAS.2002548117/ASSET/E4B53E21-F3B6-4826-B889-A0DB0FC7A08C/ASSETS/IMAGES/LARGE/PNAS.2002548117FIG01.JPG.

Rodrigues, J. L. M., Pellizari, V. H., Mueller, R., Baek, K., Jesus, E. D. C., Paula, F. S., et al. (2013). Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. Proc Natl Acad Sci U S A 110, 988–993. doi: 10.1073/PNAS.1220608110/SUPPL_FILE/PNAS.201220608SI.PDF.

Sanford, R. A., Cole, J. R., and Tiedje, J. M. (2002). Characterization and description of *Anaeromyxobacter dehalogenans* gen. nov., sp. nov., an Aryl-halorespiring facultative

anaerobic myxobacterium. Appl Environ Microbiol 68, 893–900. doi: 10.1128/AEM.68.2.893-900.2002.

Santos, J. S. dos, Dodonov, P., Oshima, J. E. F., Martello, F., Santos de Jesus, A., Eduardo Ferreira, M., et al. (2021). Landscape ecology in the Anthropocene: an overview for integrating agroecosystems and biodiversity conservation. Perspect Ecol Conserv 19, 21–32. doi: 10.1016/j.pecon.2020.11.002.

Santos, J. S. dos, Silva-Neto, C. M., Silva, T. C., Siqueira, K. N., Ribeiro, M. C., and Collevatti, R. G. (2022). Landscape structure and local variables affect plant community diversity and structure in a Brazilian agricultural landscape. Biotropica 54, 239–250. doi: 10.1111/BTP.13054.

Schröder, T., and Fleig, F. D. (2017). Spatial patterns and edge effects on soil organic matter and nutrients in a forest fragment of southern Brazil. Soil Research 55, 649–656. doi: 10.1071/SR16186.

Schulz, S., Brankatschk, R., Dümig, A., Kögel-Knabner, I., Schloter, M., and Zeyer, J. (2013). The role of microorganisms at different stages of ecosystem development for soil formation. Biogeosciences 10, 3983–3996. doi: 10.5194/bg-10-3983-2013.

Seitz, T. J., Schütte, U. M. E., and Drown, D. M. (2021). Soil Disturbance Affects Plant Productivity via Soil Microbial Community Shifts. Front Microbiol 12, 76. doi: 10.3389/FMICB.2021.619711/BIBTEX.

Sheik, C. S., Mitchell, T. W., Rizvi, F. Z., Rehman, Y., Faisal, M., Hasnain, S., et al. (2012). Exposure of Soil Microbial Communities to Chromium and Arsenic Alters Their Diversity and Structure. PLoS One 7, e40059. doi: 10.1371/JOURNAL.PONE.0040059.

Singh, B. K., Quince, C., Macdonald, C. A., Khachane, A., Thomas, N., Al-Soud, W. A., et al. (2014). Loss of microbial diversity in soils is coincident with reductions in

some specialized functions. Environ Microbiol 16, 2408–2420. doi: 10.1111/1462-2920.12353.

Slimane, M., and El-Hafid, N. (2021). Microbiome response under heavy metal stress. Heavy Metals in the Environment: Impact, Assessment, and Remediation, 39–56. doi: 10.1016/B978-0-12-821656-9.00003-1.

Souza, R. C., Mendes, I. C., Reis-Junior, F. B., Carvalho, F. M., Nogueira, M. A., Vasconcelos, A. T. R., et al. (2016). Shifts in taxonomic and functional microbial diversity with agriculture: How fragile is the Brazilian Cerrado? BMC Microbiol 16, 1–15. doi: 10.1186/S12866-016-0657-Z/FIGURES/6.

Tatusov, R. L., Koonin, E. v., and Lipman, D. J. (1997). A Genomic Perspective on Protein Families. Science (1979) 278, 631–637. doi: 10.1126/SCIENCE.278.5338.631.

Torsvik, V., and Øvreås, L. (2002). Microbial diversity and function in soil: From genes to ecosystems. Curr Opin Microbiol 5, 240–245. doi: 10.1016/S1369-5274(02)00324-7.

Veldkamp, E., Schmidt, M., Powers, J. S., and Corre, M. D. (2020). Deforestation and reforestation impacts on soils in the tropics. Nature Reviews Earth & Environment 2020 1:11 1, 590–605. doi: 10.1038/s43017-020-0091-5.

Wang, I. J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. Evolution (N Y) 67, 3403–3411. doi: 10.1111/EVO.12134.

Wolińska, A., Górniak, D., Zielenkiewicz, U., Goryluk-Salmonowicz, A., Kuźniar, A., Stępniewska, Z., et al. (2017). Microbial biodiversity in arable soils is affected by agricultural practices. Int Agrophys 31, 259–271. doi: 10.1515/INTAG-2016-0040.

Wood, D. E., and Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15, R46. doi: 10.1186/gb-2014-15-3-r46.

Wu, Q., Sanford, R. A., and Löffler, F. E. (2006). Uranium(VI) reduction by *Anaeromyxobacter dehalogenans* strain 2CP-C. Appl Environ Microbiol 72, 3608–3614. doi: 10.1128/AEM.72.5.3608-3614.2006/ASSET/5D5A7137-5B98-4369-9D51-F4309F577569/ASSETS/GRAPHIC/ZAM0050667730006.JPEG.

Yao, Y., Sui, X. H., Zhang, X. X., Wang, E. T., and Chen, W. X. (2015). *Bradyrhizobium erythrophlei* sp. nov. and *Bradyrhizobium ferriligni* sp. nov., isolated from effective nodules of *Erythrophleum fordii*. Int J Syst Evol Microbiol 65, 1831–1837. doi: 10.1099/IJS.0.000183/CITE/REFWORKS.

Ye, S. H., Siddle, K. J., Park, D. J., and Sabeti, P. C. (2019). Benchmarking Metagenomics Tools for Taxonomic Classification. Cell 178, 779–794. doi: 10.1016/J.CELL.2019.07.010.

Zhong, X., Chen, Z., Ding, K., Liu, W. S., Baker, A. J. M., Fei, Y. H., et al. (2022). Heavy metal contamination affects the core microbiome and assembly processes in metal mine soils across Eastern China. J Hazard Mater 443, 130241. doi: 10.1016/J.JHAZMAT.2022.130241.

Zhou, X., Khashi u Rahman, M., Liu, J., and Wu, F. (2021). Soil acidification mediates changes in soil bacterial community assembly processes in response to agricultural intensification. Environ Microbiol 23, 4741–4755. doi: 10.1111/1462-2920.15675.

Zuur, A. F., Ieno, E. N., and Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. Methods Ecol Evol 1, 3–14. doi: 10.1111/j.2041-210x.2009.00001.x.

Appendix

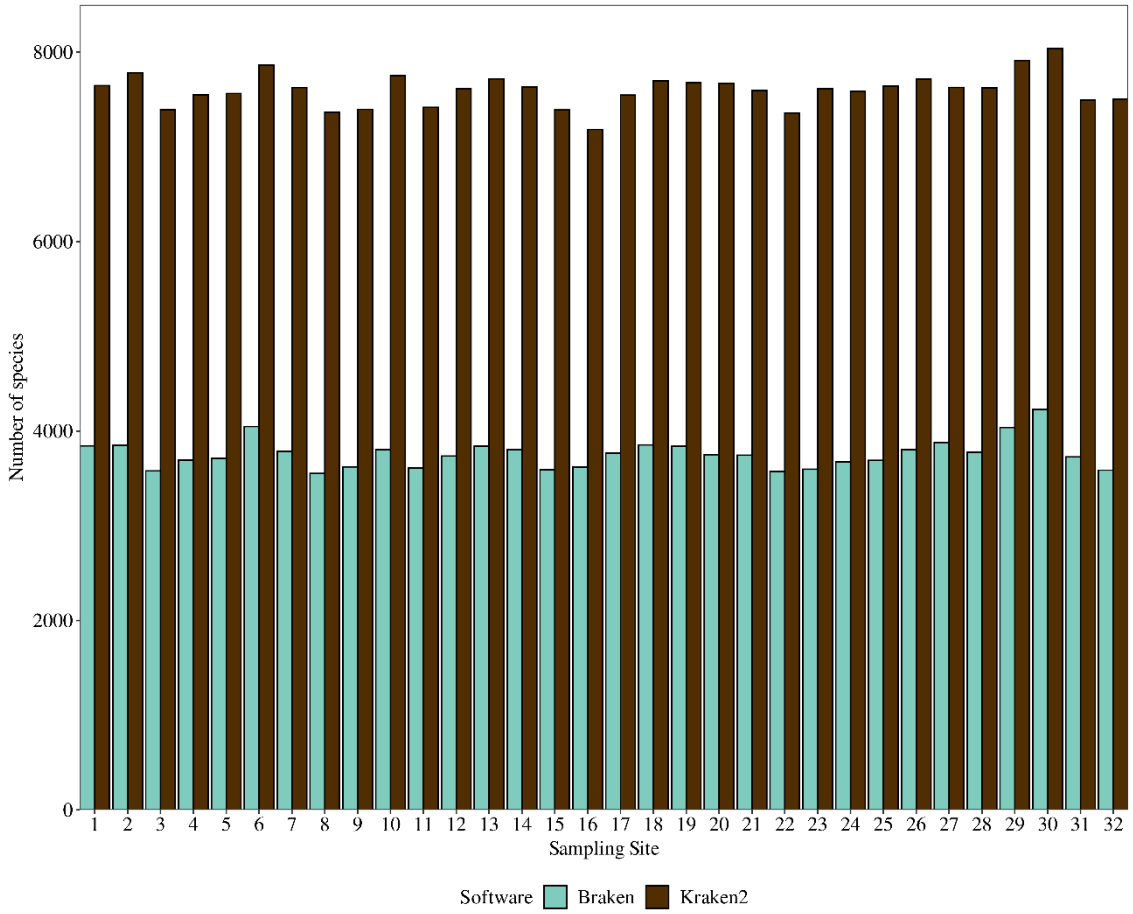

Figure S 1. Number of species assigned with Kraken 2 according to RefSeq database and the number of species reclassified by Bracken for each sampling site.

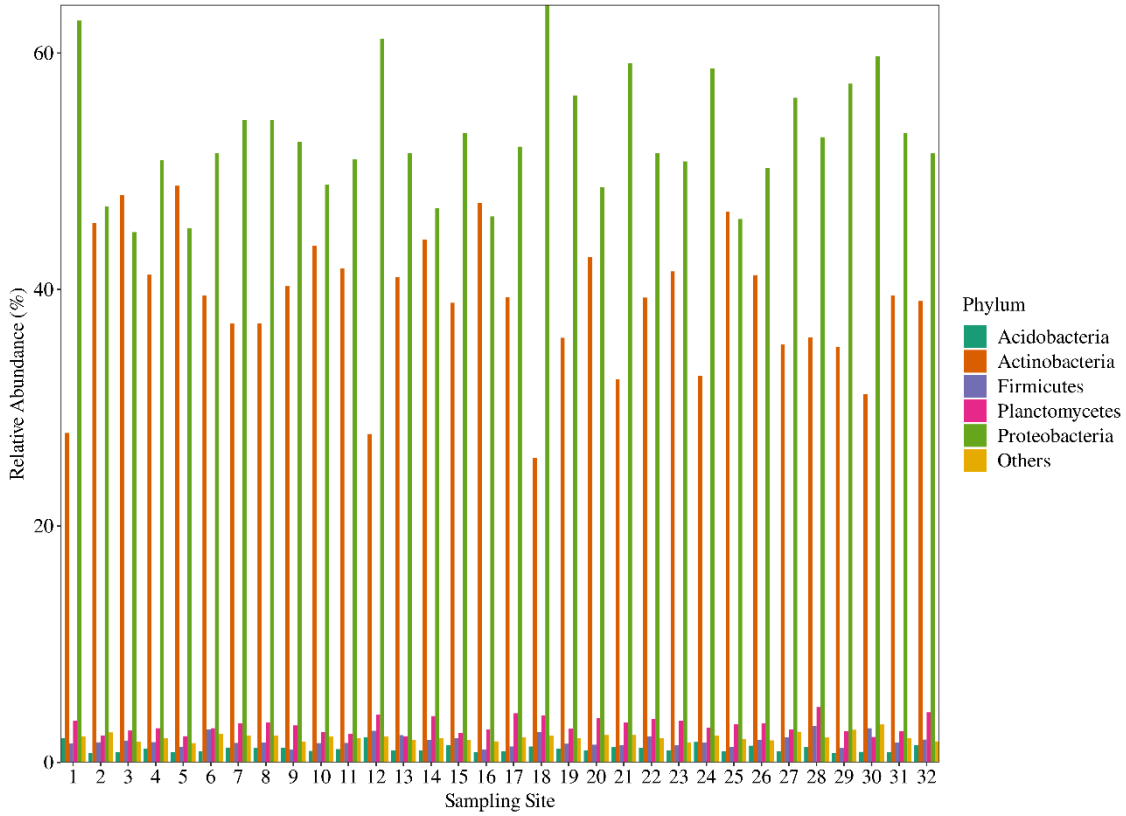

Figure S 2. Relative abundance of each phylum assigned by Kraken 2 for each sampling site.

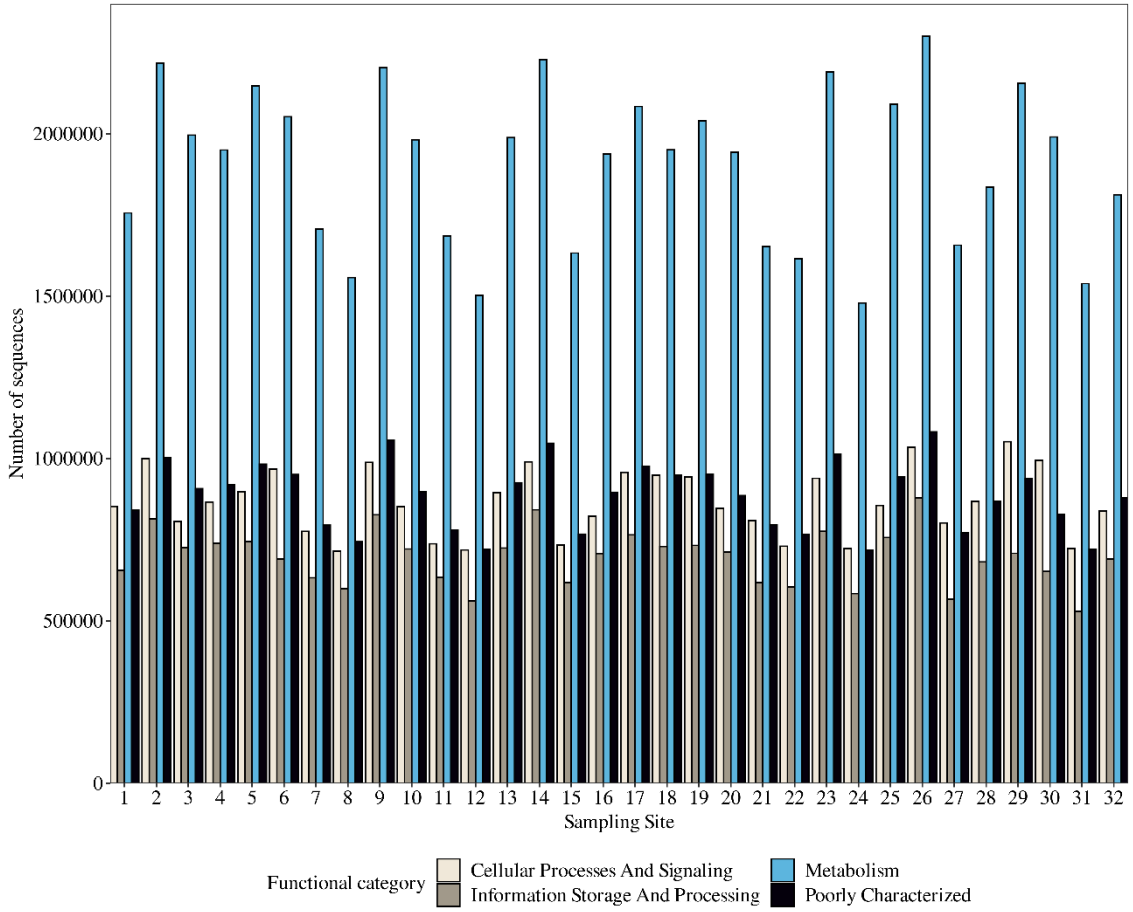

Figure S 3. Number of sequences for each functional category and sampling site obtained with MG-RAST against the COG database.

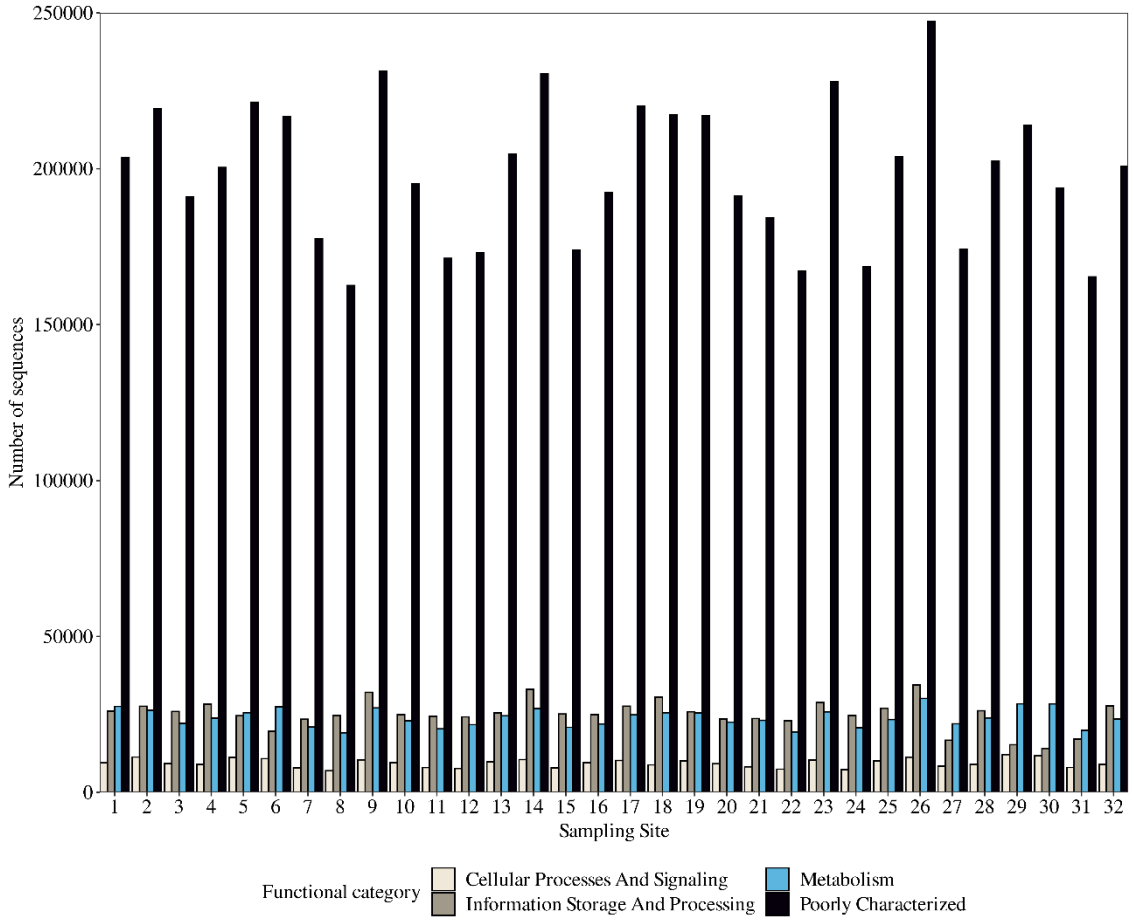

Figure S 4. Number of sequences for each functional category and sampling site obtained with MG-RAST against the NOG database.

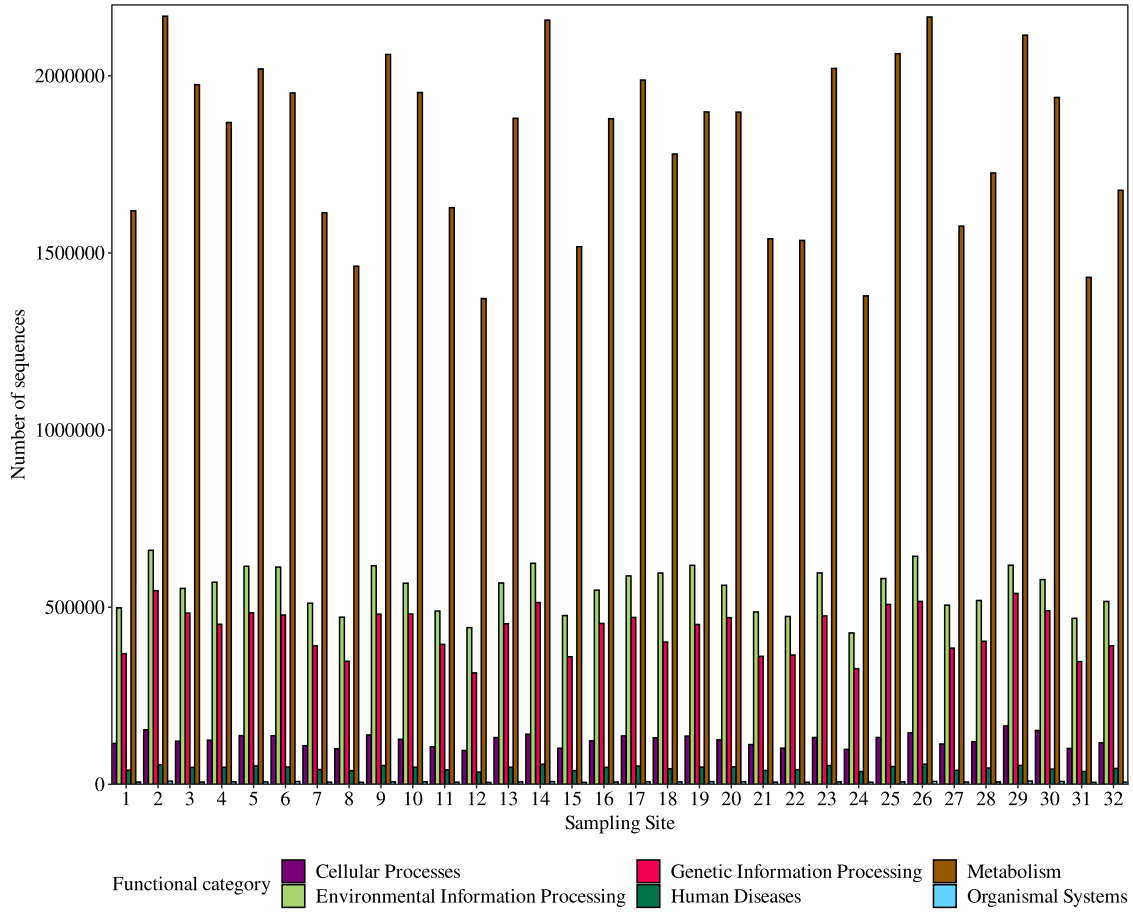

Figure S 5. Number of sequences for each functional category and sampling site obtained with MG-RAST against the KEGG database.

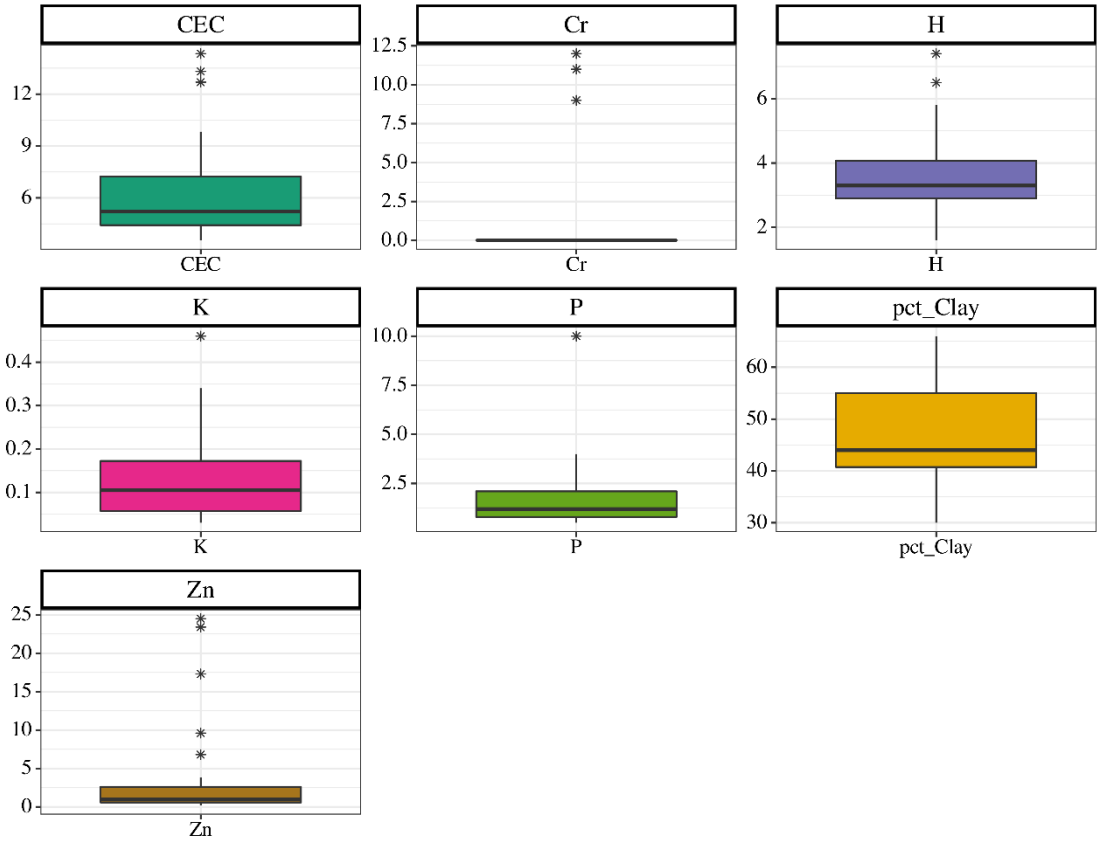

Figure S 6. Local variables selected after multicollinearity analysis. CEC cation exchange capacity; Cr chromium; H hydrogen; K potassium; P phosphorus; pct_Clay percentage of Clay; Zn zinc.

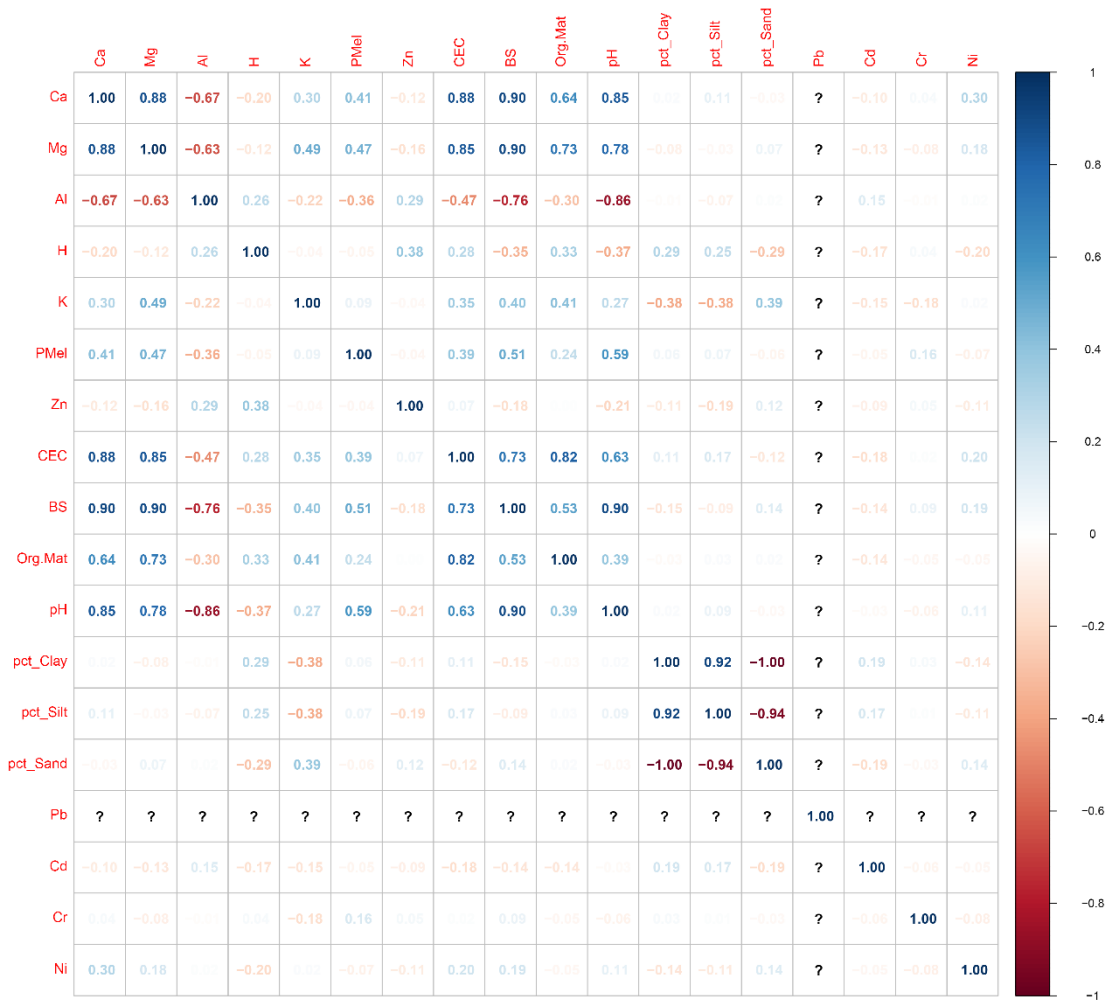

Figure S 7. Pearson's correlation between soil physicochemical variables. Color gradient represents correlation strength between variables.

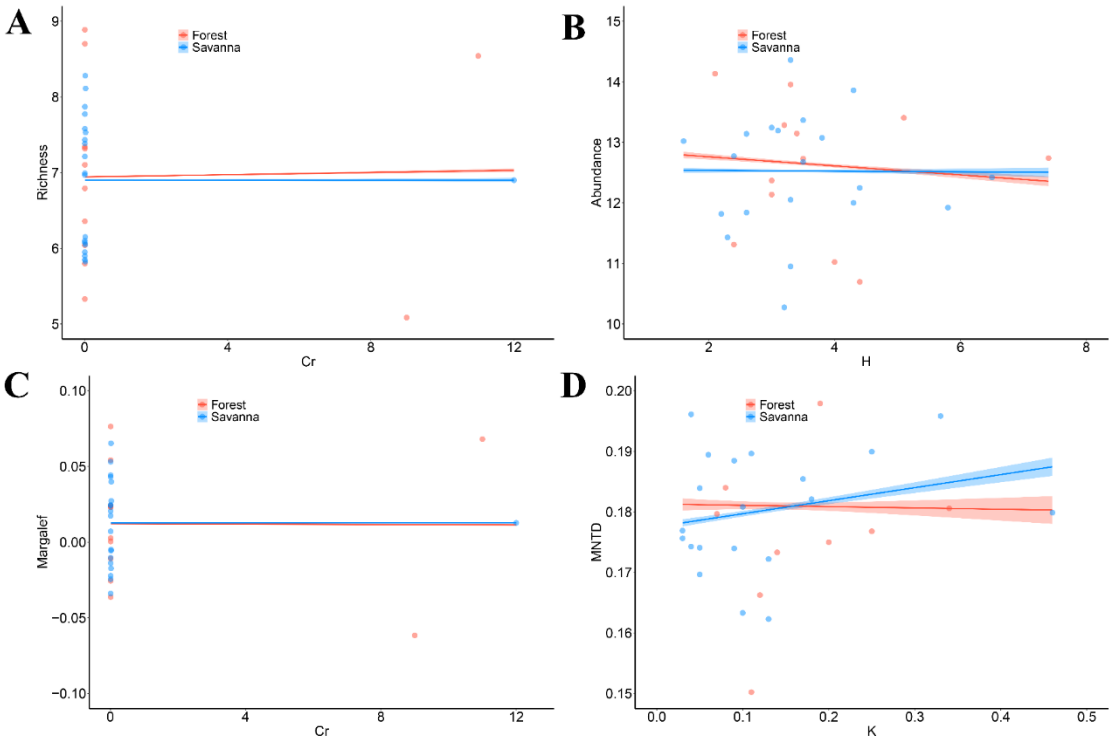

Figure S 8. Local variables influence on alpha diversity of microbiome community in agriculture landscapes. The lines correspond to the linear regression fit to each vegetation type, and the shaded area indicates the 95% confidence interval.