

Bacterial genomes recovered from litter's metagenomes in Amazonian Dark Earths

Anderson Santos de Freitas,¹ Luís Felipe Guandalin Zagatto,² Gabriel Silvestre Rocha,³ Thierry Alexandre Pellegrinetti,¹ Letícia de Cássia Malho Alves,¹ Vitor Moreira de Lara,¹ Jéssica Adrielle Mandro,¹ Guilherme Lucio Martins,¹ Aleksander Westphal Muniz,⁴ Rogério Eiji Hanada,⁵ Luiz Fernando Würdig Roesch,⁶ Siu Mui Tsai¹

AUTHOR AFFILIATIONS See affiliation list on p. 7.

ABSTRACT Here, we report 27 metagenome-assembled bacterial genomes (MAGs) from litter samples of a secondary forest located in Brazil over an Amazonian Dark Earth pool. The data set includes members from the phyla Pseudomonadota (14 MAGs), Actinomycetota (7 MAGs), Bacteroidota (4 MAGs), Bacillota (1 MAG), and Bdellovibrionota (1 MAG).

KEYWORDS next-generation sequencing, Amazon Rainforest, environmental DNA, genome assembly

Forest litter is a crucial component for nutrient cycling, harboring diverse microbes with biotechnological potential (1). Amazonian Dark Earths are small patches of high-fertility soil in the Amazon that were enriched by the pre-Columbian native people who inhabited these areas and manipulated the soil with ceramic pieces, charcoal, and animal bones (2, 3). Here, we present 27 bacterial metagenome-assembled genomes (MAGs) obtained from litter samples from a secondary ADE forest, a potential hotspot for prospecting microbes that could enhance agricultural practices and ecological restoration (4, 5).

Six litter samples were collected on November 2021 from the Caldeirão Experimental Research Station, located in Iranduba-AM, Brazil (60°23'00" W, 03°26'00"S). Each sample was obtained from distinct points (15 m apart from each) in a 40-year secondary forest with a horizon A/AB anthropic characterized as ADE using a 25 cm² frame randomly placed on the forest. This study was registered in the Brazilian National System for Management of Genetic Patrimonial and Associated Traditional Knowledge (SISGEN) under the access number AD13FB3.

For each sample, 1 g of chopped litter was added into a 50 mL sterile tube filled with 40 mL of autoclaved Phosphate-Buffered Saline (PBS; 1×, pH 7.4) and incubated at 4°C with a shaking speed of 300 rpm for 12 h. The extracts were filtered through a 2 mm sieve to remove leaf particles to another sterile 50 mL tube and centrifuged at 12,000 × *g* for 15 min. The pellet was used for DNA extraction using the DNeasy PowerLyzer PowerSoil Kit (Qiagen, Hilden, Germany) following the recommendations for tropical soils (6). Libraries were prepared using NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, USA), and samples were sequenced using the shotgun approach at the NovaSeq 6000 platform (Illumina Inc., San Diego, USA) by the method paired-end 2 × 150 bp.

Raw reads were processed in KBase (7). Default parameters were used for all software unless otherwise noted. Quality check was carried out using Trimmomatic v.0.36 (8). Three different methods were used to retrieve the MAGs. First, each sample was *de novo* assembled using metaSPAdes v3.15.5 assembler (9), evaluated with QUAST v5.2.0 (10), and bins were generated with MetaBAT 2 (11). Second, a co-assembly of the six

Editor Simon Roux, DOE Joint Genome Institute, Berkeley, California, USA

Address correspondence to Anderson Santos de Freitas, andersonfreitas@usp.br.

The authors declare no conflict of interest.

See the funding table on p. 7.

Received 21 April 2024

Accepted 10 May 2024

Published 4 June 2024

Copyright © 2024 Freitas et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

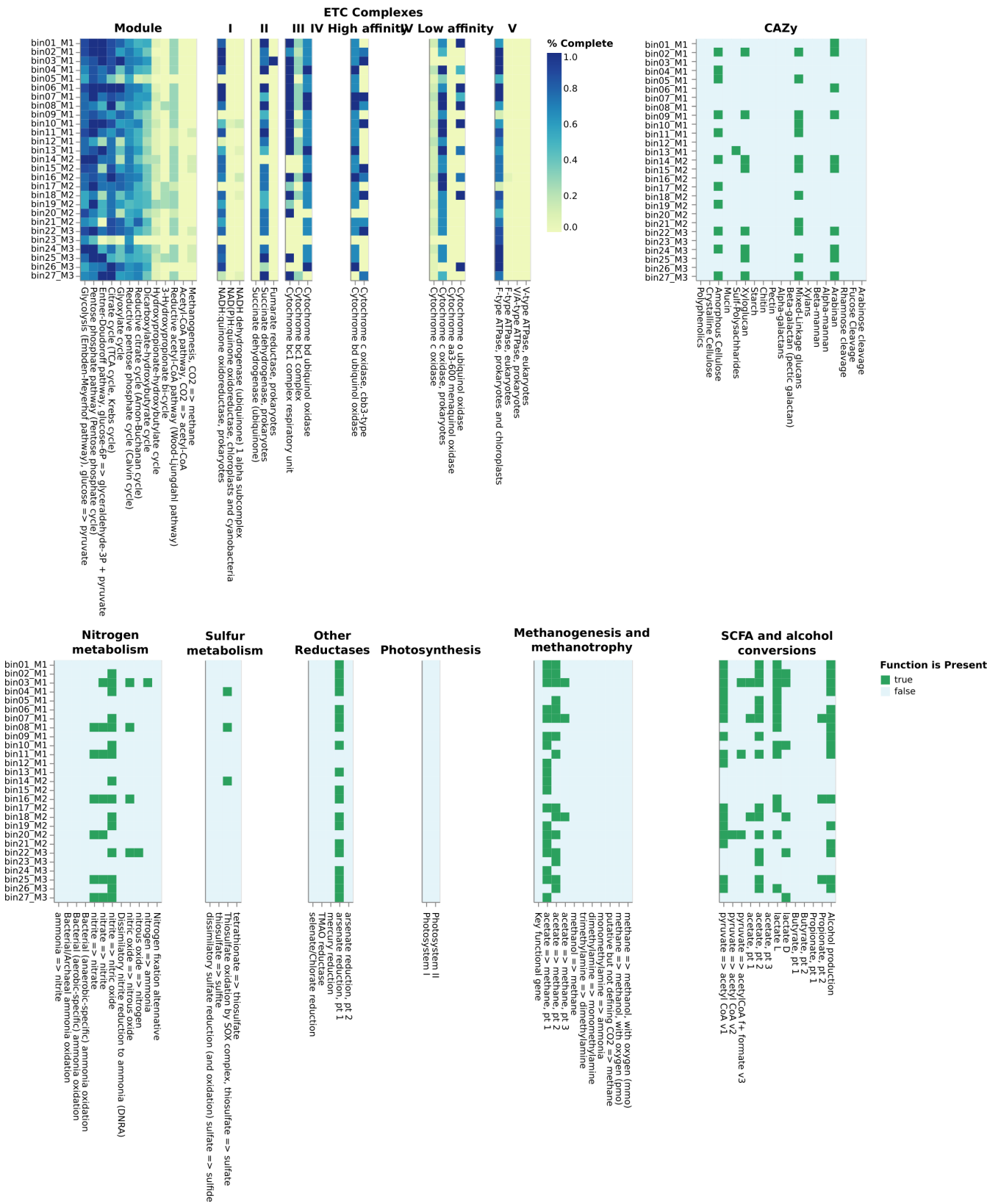


FIG 1 DRAM annotations of the 27 MAGs recovered from the metagenomes of litter in Amazonian Dark Earths.

TABLE 1 Description of 27 MAGs obtained from litter samples in a secondary Amazonian Dark Earth forest^a

MAG	GTDB classification	SRA identifier	GenBank identifier	Contigs	Completeness	Contamination	Total length	GC%	N50	CDS	rRNA	tRNA	tmRNA
bin01_M1	d__Bacteri__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales; f__Cellvibrionaceae;g__Cellvibrio;s__	SAMN40501439	JBCECX0000000000	74	97.04	0.38	5,128,080	47.47	128,266	4,306	0	28	1
bin02_M1	d__Bacteri__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales; f__Pseudomonadaceae;g__Pseudomonas_E;s__	SAMN40501440	JBCECW0000000000	653	76.94	0.82	4,102,145	62.12	7,014	3,894	0	25	0
bin03_M1	d__Bacteri__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales; f__Rhodocyclaceae;g__Dactylopiibacterium;s__	SAMN40501441	JBCECV0000000000	99	96.78	1.24	3,325,840	68.58	53,501	3,030	0	41	1
bin04_M1	d__Bacteri__Pseudomonadota;c__Gammaproteobacteria; o__Burkholderialesf__Burkholderiaceae;g__Oxalicibacterium; s__Oxalicibacterium faecigallinarum	SAMN40501442	JBCECU0000000000	164	95.59	0.43	2,979,859	54.97	25,452	2,906	0	38	1
bin05_M1	d__Bacteri__Actinomycetota;c__Actinomycetia; o__Actinomycetalesf__Microbacteriaceae;g__Curtobacterium; s__Curtobacterium sp900086645	SAMN40501443	JBCECT0000000000	428	79.38	2.87	3,447,546	70.94	10,039	3,415	0	48	1
bin06_M1	d__Bacteri__Pseudomonadota;c__Gammaproteobacteria; o__Xanthomonadalesf__Rhodanobacteraceae;g__Luteibacter; s__Luteibacter sp000745055	SAMN40501444	JBCECS0000000000	37	98.78	0.63	4,572,070	64.98	293,888	4,040	1	48	1
bin07_M1	d__Bacteri__Pseudomonadota;c__Gammaproteobacteria; o__Burkholderialesf__Burkholderiaceae_Bg__ALPHA2B;s__	SAMN40501445	JBCECR0000000000	672	98.98	0.82	5,438,842	69.19	10,528	5,030	0	42	0
bin08_M1	d__Bacteri__Pseudomonadota;c__Gammaproteobacteria; o__Burkholderialesf__Burkholderiaceae_C;g__Bordetella; s__Bordetella sp002261335	SAMN40501446	JBCECQ0000000000	199	90.52	0.62	5,097,444	70.34	41,725	4,827	1	42	1

(Continued on next page)

TABLE 1 Description of 27 MAGs obtained from litter samples in a secondary Amazonian Dark Earth forest^a (Continued)

MAG	GTDB classification	SRA identifier	GenBank identifier	Contigs	Completeness	Contamination	Total length	GC%	N50	CDS	rRNA	tRNA	tmRNA
bin09_M1	d__Bacteriopl__Actinomycetota;c__Acti- nomycetia; o__Actinomycetales;f__Microbacteria- ceae;g__Leifsonia;s__	SAMN40501447	JBCECP0000000000	71	89.32	0.88	5,023,944	58.24	221,190	4,662	0	41	1
bin10_M1	d__Bacteriopl__Pseudomona- dotax__Alphaproteobacteria; o__Rhizobiales;f__Rhizobiaceae;g__Agro- bacterium;	SAMN40501448	JBCECO0000000000	180	97.3	0.87	3,634,223	70.85	35,746	3,617	0	39	1
bin11_M1	s__Agrobacterium cavaerae d__Bacteriopl__Actinomycetota;c__Acti- nomycetia; o__Actinomycetales;f__Cellulomonada- ceae;	SAMN40501449	JBCECN0000000000	181	95.17	0.0	1,609,008	35.46	12,035	1,429	0	24	1
bin12_M1	g__Cellulomonas;s__ d__Bacteriopl__Pseudomona- dotax__Gammaproteobacteria; o__Diplorickettsiales;f__Diplorickettsia- ceae;g__Rickettsiella_B's__	SAMN40501450	JBCECM0000000000	319	86.43	0.87	3,062,255	68.66	12,901	3,100	0	40	0
bin13_M1	d__Bacteriopl__Pseudomona- dotax__Alphaproteobacteria; o__Caulobacteriales;f__Caulobactera- ceae;g__Brevundimonas;s__	SAMN40501451	JBCECL0000000000	139	87.69	1.93	3,434,383	72.37	38,335	3,111	0	47	0
bin14_M2	d__Bacteriopl__Bacteroidota;c__Bacteroi- diao__Sphingobacteriales; f__Sphingobacteriaceae;g__Mucilagini- bacter;s__	SAMN40501452	JBCECK0000000000	325	93.71	3.97	6,024,919	41.41	33,895	5,116	1	38	1
bin15_M2	d__Bacteriopl__Bacteroidota;c__Bacteroi- diao__Flavobacteriales; f__Flavobacteriaceae;g__Flavobacte- rium;s__Flavobacterium lindanitolerans	SAMN40501453	JBCECJ0000000000	221	96.37	0.35	3,626,744	37.68	29,068	3,211	0	29	1
bin16_M2	d__Bacteriopl__Pseudomona- dotax__Alphaproteobacteria; o__Rhizobiales;f__Rhizobia- ceae;g__Ochrobactrum_A;	SAMN40501454	JBCECI0000000000	344	92.05	2.83	4,239,545	53.72	18,488	4,060	0	38	1
bin17_M2	s__Ochrobactrum_A pseudogrignonensis d__Bacteriopl__Actinomycetota;c__Acti- nomycetia;o__Actinomycetales;	SAMN40501455	JBCECH0000000000	342	83.51	1.73	3,382,307	69.93	15,560	3,286	1	34	0

(Continued on next page)

TABLE 1 Description of 27 MAGs obtained from litter samples in a secondary Amazonian Dark Earth forest^a (Continued)

MAG	GTDB classification	SRA identifier	GenBank identifier	Contigs	Completeness	Contamination	Total length	GC%	N50	CDS	rRNA	tRNA	tmRNA
bin18_M2	f__Microbacteriaceae;g__Microbacterium;s__												
	d__Bacteriia;p__Pseudomonadota;c__Alphaproteobacteria;o__Acetobacteriales;f__Acetobacteraceae;	SAMN40501456	JBCECG0000000000	817	81.88	4.27	4,547,200	72.04	6,566	4,322	1	43	1
	g__Paracraurococcus;s__Paracraurococcus sp013112485												
	d__Bacteriia;p__Actinomycetota;c__Thermoleophila;	SAMN40501457	JBCECF0000000000	675	86.37	0.98	4,429,271	71.80	8,477	4,226	1	47	1
	o__Solirubrobacterales;f__Solirubrobacteraceae;												
bin20_M2	g__JAGIB01;s__												
	d__Bacteriia;p__Actinomycetota;c__Actinomyces;	SAMN40501458	JBCECE0000000000	637	74.89	0.69	3,163,046	71.50	5,539	3,188	0	61	1
	o__Propionibacteriales;f__Propionibacteriaceae;												
bin21_M2	g__Propionimonas;s__												
	d__Bacteriia;p__Bdellovibrionota;c__Bdellovibrionales;f__Bdellovibrionaceae;g__Bdellovibrionaceae;	SAMN40501459	JBCECD0000000000	473	89.66	0.51	3,024,043	46.47	7,644	3,038	1	30	0
	lovibionia;												
bin22_M3	d__Bacteriia;p__Bacteroidota;c__Bacteroidia;	SAMN40501460	JBCECC0000000000	148	97.45	1.43	5,295,192	33.77	70,500	4,594	1	43	1
	o__Flavobacteriales;f__Flavobacteriaceae;g__Flavobacterium;												
	s__Flavobacterium nitrogenifigens												
bin23_M3	d__Bacteriia;p__Bacillota;c__Bacilliales;f__Bacillaceae;g__Bacillus;	SAMN40501461	JBCECB0000000000	8	95.49	0.75	807,964	26.76	184,592	1,405	3	27	1
	o__Mycoplasmatales;												
	f__VBWQ01;g__Spiroplasma_D5												
bin24_M3	d__Bacteriia;p__Bacteroidota;c__Bacteroidia;	SAMN40501462	JBCECA0000000000	27	97.13	0.0	3,903,772	43.98	297,924	4,362	2	32	1
	o__Sphingobacteriales;f__Sphingobacteriaceae;g__Pedobacter;s__												
	d__Bacteriia;p__Actinomycetota;c__Actinomyces;	SAMN40501463	JBCEBZ0000000000	597	89.51	3.24	7,655,958	68.18	17,547	7,051	0	47	1
bin25_M3	o__Mycobacteriales;f__Mycobacteriaceae;g__Rhodococcus												

(Continued on next page)

TABLE 1 Description of 27 MAGs obtained from litter samples in a secondary Amazonian Dark Earth forest^a (Continued)

MAG	GTDB classification	SRA identifier	GenBank identifier	Contigs	Completeness	Contamination	Total length	GC%	N50	CDS	rRNA	tRNA	tmRNA
bin26_M3	d__Bacteriopl__Pseudomona-	SAMN40501464	JBCEBY0000000000	569	87.0	1.15	3,290,807	38.92	6,445	3,075	0	31	1
	dota;c__Gammaproteobacteria;												
	o__Pseudomonadales;f__Moraxella-												
	ceae;g__Acinetobacter;												
bin27_M3	s__Acinetobacter pittii	SAMN40501465	JBCEBX0000000000	394	82.8	1.73	2,719,478	56.58	8,186	2,687	0	31	0
	d__Bacteriopl__Pseudomona-												
	dota;c__Alphaproteobacteria;												
	o__Caulobacteriales;f__Caulobactera-												
	ceae;g__Asticcacaulis;												

^aProtein coding sequence.

samples (merged fastq files) was performed using metaSPAdes. Bins were retrieved using MetaBAT 2, MaxBin2 v2.2.4 (12), and Concoct v1.1 (13). The three binned results were added to DASTool v1.1.2 for refinement and de-replication (14). The third method was performed by importing the co-assembled reads into the COMEBin v1.0.3 (15). All bins generated in all methods were de-replicated using dRep v3.1.0 (16). De-replicated bins were quality-checked considering completeness and contamination (17) with CheckM v1.0.18 (18). The best quality MAGs (>70% completeness, <5% contamination) were then taxonomically annotated with GTDB-Tk v2.3.2 (19), and Prokka v1.14.5 (20), and functionally annotated with DRAM v0.1.2 (21), respectively (Table 1)(Fig. 1).

Our data provide a better understanding of the bacterial roles in nutrient cycling within ADE litter. We expect these results to be useful for future research in soil biotechnology.

ACKNOWLEDGMENTS

This work was supported by the São Paulo Research Foundation (FAPESP; grant numbers 2020/08927-0, 2022/13186-4, 2022/13186-4, and 2014/50320-4; and fellowships 2021/10.626-0, 2021/10.716-0, and 2021/11.386-3), the National Council for Scientific and Technological Development (CNPq; grant numbers 423437/2021-6 and 314806/2021-0), the Amazonas Research Foundation (FAPEAM; grant number 01.02.016301.00293/2021), the Coordination for the Improvement of Higher Education Personnel–Brasil (CAPES; fellowship 887.597909/2021-00) – Finance Code 001, and the PRPI-USP (22.1.08498.01.0).

AUTHOR AFFILIATIONS

¹Center for Nuclear Energy in Agriculture, University of Sao Paulo, Piracicaba, Brazil

²Department of Terrestrial Ecology, Netherlands Institute of Ecology, Wageningen, the Netherlands

³Department of Microbial Ecology, Netherlands Institute of Ecology, Wageningen, the Netherlands

⁴Western Amazon Agroforestry Research Center, Brazilian Agricultural Research Corporation, Manaus, Brazil

⁵General Coordination of Research, Training, and Extension, National Institute of Amazonian Research, Manaus, Brazil

⁶Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida, USA

AUTHOR ORCIDs

Anderson Santos de Freitas  <http://orcid.org/0000-0002-7726-0180>

Luís Felipe Guandalin Zagatto  <http://orcid.org/0000-0001-9984-9044>

Gabriel Silvestre Rocha  <http://orcid.org/0000-0002-3188-6268>

Thierry Alexandre Pellegrinetti  <http://orcid.org/0000-0001-9386-6273>

Jéssica Adrielle Mandro  <https://orcid.org/0000-0003-3454-0702>

Guilherme Lucio Martins  <http://orcid.org/0000-0001-8953-2529>

Aleksander Westphal Muniz  <http://orcid.org/0000-0001-9182-0669>

Rogério Eiji Hanada  <http://orcid.org/0000-0002-4544-4882>

Luiz Fernando Würdig Roesch  <http://orcid.org/0000-0003-1450-8828>

Siu Mui Tsai  <https://orcid.org/0000-0002-3733-6312>

FUNDING

Funder	Grant(s)	Author(s)
Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)	2021/10.626-0	Anderson Santos de Freitas

Funder	Grant(s)	Author(s)
Conselho Nacional de Desenvolvimento Científico e Tecnológico	314806/2021-0	Siu Mui Tsai
Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM)	01.02.016301.00293/2021	Rogério Eiji Hanada
Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)	2020/08927-0	Siu Mui Tsai

AUTHOR CONTRIBUTIONS

Anderson Santos de Freitas, Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review and editing | Luís Felipe Guandalin Zagatto, Conceptualization, Investigation, Writing – original draft | Gabriel Silvestre Rocha, Validation, Writing – original draft | Thierry Alexandre Pellegrinetti, Formal analysis, Methodology, Software, Writing – review and editing | Letícia de Cássia Malho Alves, Investigation, Visualization | Vitor Moreira de Lara, Formal analysis, Investigation | Jéssica Adrielle Mandro, Methodology, Validation | Guilherme Lucio Martins, Investigation, Validation, Visualization | Aleksander Westphal Muniz, Resources, Supervision | Rogério Eiji Hanada, Resources, Writing – review and editing | Luiz Fernando Würdig Roesch, Investigation, Methodology, Writing – review and editing | Siu Mui Tsai, funding acquisition, project administration, Resources, Supervision, Writing – review and editing

DATA AVAILABILITY

This Whole Genome Shotgun project has been deposited in GenBank under accession no. [PRJNA1088826](https://doi.org/10.1093/bioinformatics/btt086). The raw reads can be found in SRA under the accession numbers [SRR25306242](https://doi.org/10.1093/bioinformatics/btt086), [SRR25306243](https://doi.org/10.1093/bioinformatics/btt086), [SRR25306244](https://doi.org/10.1093/bioinformatics/btt086), [SRR25306239](https://doi.org/10.1093/bioinformatics/btt086), [SRR25306240](https://doi.org/10.1093/bioinformatics/btt086), and [SRR25306241](https://doi.org/10.1093/bioinformatics/btt086). The average number of reads per sample was 40.1 ± 7.1 million reads. The version described in this paper is the first version. The annotation files can be found on GitHub at the address <https://github.com/FreitasAndy/LitterMAGS>.

REFERENCES

- Krishna MP, Mohan M. 2017. Litter decomposition in forest ecosystems: a review. *Energ Ecol Environ* 2:236–249. <https://doi.org/10.1007/s40974-017-0064-9>
- Navarrete AA, Cannavan FS, Taketani RG, Tsai SM. 2010. A molecular survey of the diversity of microbial communities in different Amazonian agricultural model systems. *Diversity* 2:787–809. <https://doi.org/10.3390/d2050787>
- Silva LCR, Corrêa RS, Wright JL, Bomfim B, Hendricks L, Gavin DG, Muniz AW, Martins GC, Motta ACV, Barbosa JZ, Melo V de F, Young SD, Broadley MR, Santos RV. 2021. A new hypothesis for the origin of Amazonian Dark Earths. *Nat Commun* 12:127. <https://doi.org/10.1038/s41467-020-20184-2>
- Freitas A de, Zagatto LFG, Rocha GS, Muchalak F, Silva S dos S, Muniz AW, Hanada RE, Tsai SM. 2023. Amazonian dark earths enhance the establishment of tree species in forest ecological restoration. *Front Soil Sci* 3. <https://doi.org/10.3389/fsoil.2023.1161627>
- Barbosa Lima A, Cannavan FS, Navarrete AA, Teixeira WG, Kuramae EE, Tsai SM. 2015. Amazonian dark earth and plant species from the Amazon region contribute to shape rhizosphere bacterial communities. *Microb Ecol* 69:855–866. <https://doi.org/10.1007/s00248-014-0472-8>
- Venturini AM, Nakamura FM, Gontijo JB, da França AG, Yoshiura CA, Mandro JA, Tsai SM. 2020. Robust DNA protocols for tropical soils. *Heliyon* 6:e03830. <https://doi.org/10.1016/j.heliyon.2020.e03830>
- Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, et al. 2018. KBase: the United States department of energy systems biology knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27:824–834. <https://doi.org/10.1101/gr.213959.116>
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. <https://doi.org/10.7717/peerj.7359>
- Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32:605–607. <https://doi.org/10.1093/bioinformatics/btv638>
- Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. *Nat Methods* 11:1144–1146. <https://doi.org/10.1038/nmeth.3103>
- Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 3:836–843. <https://doi.org/10.1038/s41564-018-0171-1>
- Wang Z, You R, Han H, Liu W, Sun F, Zhu S. 2024. Effective binning of metagenomic contigs using contrastive multi-view representation learning. *Nat Commun* 15:585. <https://doi.org/10.1038/s41467-023-44290-z>
- Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* 11:2864–2868. <https://doi.org/10.1038/ismej.2017.126>

17. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloie-Fadrosch EA, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 35:725–731. <https://doi.org/10.1038/nbt.3893>
18. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
19. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>
20. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
21. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrowe AB, Rodríguez-Ramos J, Bolduc B, Gazitúa MC, Daly RA, Smith GJ, Vik DR, Pope PB, Sullivan MB, Roux S, Wrighton KC. 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* 48:8883–8900. <https://doi.org/10.1093/nar/gkaa621>