

RESEARCH

Open Access



Long-read sequencing sheds light on key bacteria contributing to deadwood decomposition processes

Etienne Richy^{1*}, Priscila Thiago Dobbler¹, Vojtěch Tláškal^{1,2}, Rubén López-Mondéjar^{1,3}, Petr Baldrian¹ and Martina Kyselková^{1*}

Abstract

Background Deadwood decomposition is an essential ecological process in forest ecosystems, playing a key role in nutrient cycling and carbon sequestration by enriching soils with organic matter. This process is driven by diverse microbial communities encompassing specialized functions in breaking down organic matter, but the specific roles of individual microorganisms in this process are still not fully understood.

Results Here, we characterized the deadwood microbiome in a natural mixed temperate forest in Central Europe using PacBio HiFi long-read sequencing and a genome-resolved transcriptomics approach in order to uncover key microbial contributors to wood decomposition. We obtained high quality assemblies, which allowed attribution of complex microbial functions such as nitrogen fixation to individual microbial taxa and enabled the recovery of metagenome-assembled genomes (MAGs) from both abundant and rare deadwood bacteria. We successfully assembled 69 MAGs (including 14 high-quality and 7 single-contig genomes) from 4 samples, representing most of the abundant bacterial phyla in deadwood. The MAGs exhibited a rich diversity of carbohydrate-active enzymes (CAZymes), with Myxococcota encoding the highest number of CAZymes and the full complement of enzymes required for cellulose decomposition. For the first time we observed active nitrogen fixation by Steroidobacteraceae, as well as hemicellulose degradation and chitin recycling by Patescibacteria. Furthermore, PacBio HiFi sequencing identified over 1000 biosynthetic gene clusters, highlighting a vast potential for secondary metabolite production in deadwood, particularly in Pseudomonadota and Myxococcota.

Conclusions PacBio HiFi long-read sequencing offers comprehensive insights into deadwood decomposition processes by advancing the identification of functional features involving multiple genes. It represents a robust tool for unraveling novel microbial genomes in complex ecosystems and allows the identification of key microorganisms contributing to deadwood decomposition.

Keywords Deadwood microbiome, Biosynthetic gene cluster, Carbohydrate-active enzyme, Microbial decomposition, Metagenome-assembled genome, Metagenomics, Nitrogen fixation

*Correspondence:

Etienne Richy
etienne.richy@biomed.cas.cz

Martina Kyselková
martina.kyselkova@biomed.cas.cz

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

Deadwood constitutes a fundamental component of natural forest ecosystems, providing a variety of niches for microorganisms and representing a hotspot of biodiversity [1]. Decomposing deadwood releases carbon dioxide into the atmosphere, but also enriches the soil with nutrients and recalcitrant organic matter (OM), thereby promoting nutrient cycling and enhancing long-term carbon storage in forest soils [2]. This decomposition process is primarily driven by wood-decaying fungi, such as basidiomycetes, and bacteria, including Alphaproteobacteria, Gammaproteobacteria, Acidobacteriota, and Actinomycetota [3, 4]. Fungi initiate the decomposition of fresh deadwood by depolymerizing the wood biopolymers including lignin and cellulose, thereby enabling subsequent colonization of wood matrix and its utilization [5, 6]. Nitrogen-fixing bacteria also play a crucial role in early stages of decomposition by mitigating the high carbon-to-nitrogen ratio typical of fresh deadwood, which supports subsequent microbial activity [7]. As decomposition progresses, fungi and bacteria collaboratively degrade plant cell wall components like cellulose and hemicellulose, using a wide range of extracellular hydrolytic enzymes.

Second-generation (short-read) sequencing has significantly advanced our understanding of the taxonomic composition and functional potential of deadwood microbial communities [4, 8–11]. However, this approach often results in incomplete and fragmented metagenome-assembled genomes (MAGs) due to limited assembly capabilities, hampering the recovery of complex genomic features such as complete operons, large genomic islands, or repetitive sequences. These limitations hinder the ability to fully explore the functional potential of the microbial community and achieve accurate taxonomic classification [12]. In contrast, third-generation (long-read) sequencing technologies such as PacBio HiFi sequencing, produce highly accurate long reads, yielding more complete and less contaminated MAGs [13, 14]. The increased read length enhances the resolution of intricate genomic structures, providing a deeper insight into the ecological roles of microorganisms.

In this study, we aimed to characterize the key microorganisms involved in deadwood decomposition in a natural mixed temperate forest from Central Europe, using PacBio HiFi sequencing in combination with a genome-resolved transcriptomics approach. Using deadwood samples previously analyzed through bacterial genomics, metagenomics and metatranscriptomics with second generation high-throughput sequencing [9–11, 15], we also aimed to evaluate the added value of PacBio HiFi sequencing over short-read sequencing for characterizing the deadwood microbiome. We anticipated that

the long-read lengths and high accuracy of PacBio HiFi sequencing would yield more complete and less contaminated MAGs from the most abundant taxa and provide better resolution of operons and gene clusters encoding specific metabolic functions.

Material and methods

Sample description

Deadwood samples were collected in the core zone of the Žofínský prales National Nature Reserve in the Czech Republic (48°39′57″N, 14°42′24″E), a forest protected since 1838. Detailed site descriptions and deadwood sampling methods are provided in Tláškal et al. [9]. A map generated with the *r* package *sf* [16] representing the location of the sampling site is provided in Fig. 1. Four samples from *Fagus sylvatica* trunks representing class 2 (4–7 years of decomposition; sample 6—BioSample accession SAMN13925154 and sample 7—SAMN13925155) and class 4 (20–41 years of decomposition; sample 57—SAMN13925167 and sample 84—SAMN13925168) were analyzed. Samples were stored at – 80 °C prior DNA isolation. Short-read sequencing using Illumina HiSeq platform was previously performed on these samples [9, 10] and the short-read data (accessible via the BioProject accession number PRJNA603240) were used in the present study.

DNA isolation for PacBio HiFi sequencing

High molecular weight DNA was isolated from the four deadwood samples using a phenol/chloroform/isoamyl alcohol extraction method according to Sagova-Mareckova et al. [17] with several modifications; notably we replaced the bead-beating step by vortexing in order to prevent excessive DNA shearing. Samples were first thoroughly ground in a sterile mortar with liquid nitrogen using a rough pestle. Sample aliquots (~250 mg) were added to 2 mL tubes with a screw lid and mixed with 600 µL extraction buffer (50 mM Na-phosphate buffer pH 8, 50 mM NaCl, 500 mM Tris–HCl pH 8, and 5% sodium dodecyl sulfate) and 300 µL phenol (pH 8)/chloroform/isoamyl alcohol (25:24:1). The tubes were vortexed for 1 min at 4 °C, using the Vortex-Genie (Scientific Industries, Bohemia, NY) set at the maximum speed. The homogenized samples were centrifuged at 10,000 g for 3 min and supernatant was transferred to a clean tube. One volume of phenol/chloroform/isoamyl alcohol (25:24:1) was added and mixed with the supernatant by gentle tube inverting for 1 min. After centrifugation at 6000 g for 5 min, the supernatant was transferred to a clean tube, mixed (gentle tube inverting for 1 min) with 1 volume of chloroform/isoamyl alcohol (24:1), and centrifuged (6000 g, 5 min). The supernatant was transferred to a clean tube and mixed with NaCl

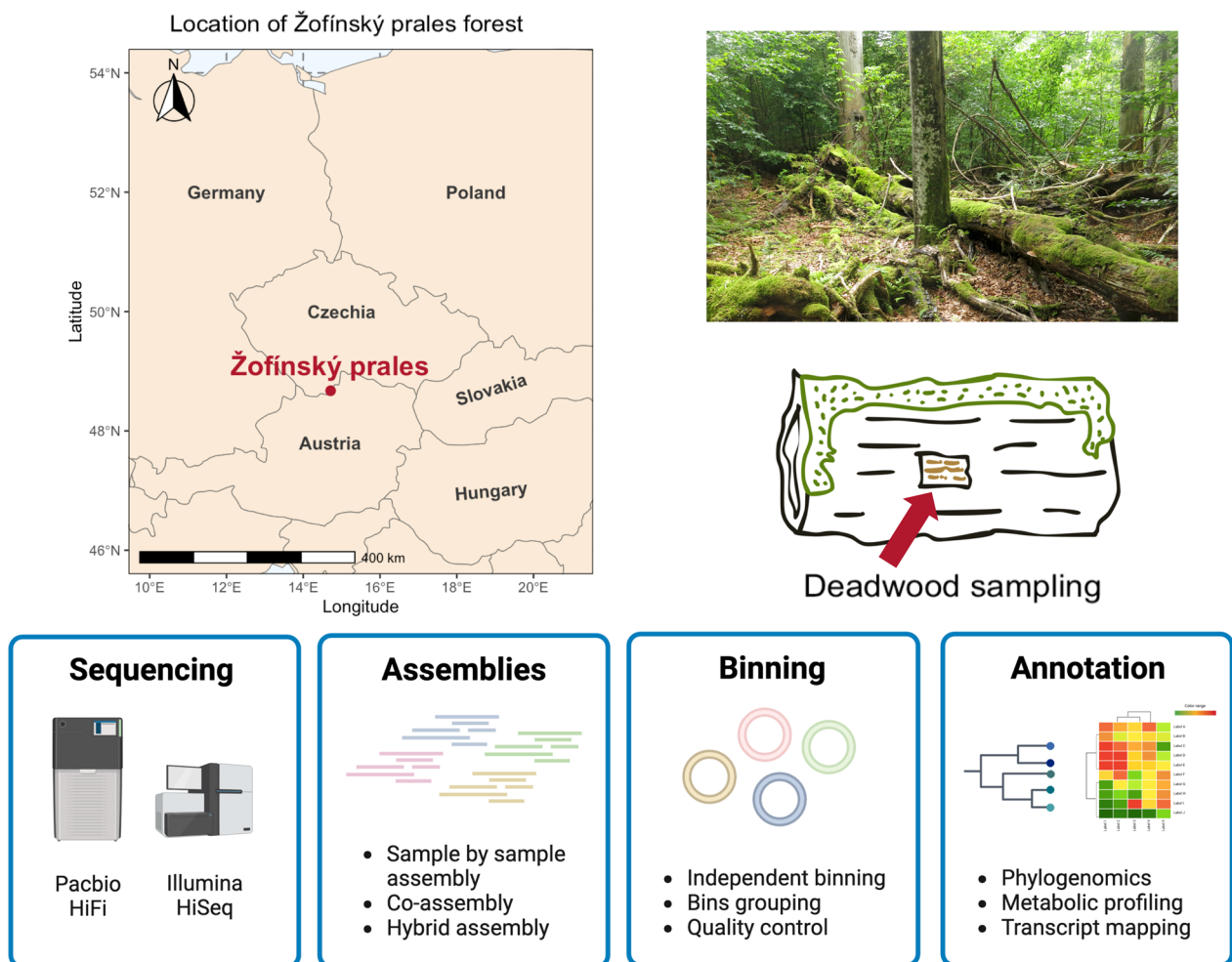


Fig. 1 Sampling strategy and metagenomic workflow. Deadwood samples ($n=4$) were collected from the Žofínský Prales National Nature Reserve, Czech Republic. For each sample, *Fagus sylvatica* trunks were drilled vertically from the middle of the top surface at five equidistant locations, and sawdust from the inside of the trunks was collected and pooled. The four samples were sequenced using PacBio HiFi and Illumina HiSeq. Sample-by-sample assembly and co-assembly was performed using short and long reads, as well as hybrid assembly. Assemblies were binned independently and bins were pooled, with the exception of the hybrid assembly, which gave unsatisfactory results. After quality control (completeness and contamination, dereplication and mis-binning detection), the phylogeny and metabolic capacities of the metagenomes-assembled genomes (MAGs) were analyzed. Transcripts were mapped to MAGs to validate the functionality of specific metabolism. The figure illustrating the metagenomic workflow was created with Biorender.com

(to a final concentration of 1.5 M) and CTAB (to a final concentration of 1%), and incubated at 65 °C for 35 min. The incubated solution was cooled at 4 °C for 5 min and then mixed with an equal volume of chloroform/isoamyl alcohol (24:1), and centrifuged at 3400 g for 20 min. The supernatant was mixed with 0.6 volume of isopropanol and 0.1 volume of 3 M sodium acetate in a clean tube and incubated at 4 °C for 45 min to precipitate the DNA. After centrifugation at 10,000 g for 20 min, the supernatant was removed and the DNA pellet was washed with 200 μ L cold 70% ethanol, air-dried and resuspended in 30 μ L 10 mM Tris buffer pH 8. Presence of high

molecular weight DNA was verified with DNA electrophoresis in 0.8% agarose gel and DNA concentration was measured with Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA). Three to six DNA aliquots per sample were pooled to obtain the minimal required DNA quantity of 5 μ g.

Sequencing and primary analyses of the metagenomes

The library preparation and PacBio HiFi sequencing on the Sequel II Instrument were performed at Brigham Young University Sequencing Centre, Utah. Samples 6 and 7 were pooled equimolarly during library preparation

and sequenced together on one 8 M SMRT Cell using 30-h movie. Samples 57 and 84 were sequenced separately on one 8 M SMRT Cell using 30-h movie. PacBio's movie subreads files were processed with CCS (<https://github.com/PacificBiosciences/pbbioconda>) for consensus Hi-Fi reads of >99% accuracy. After removing reads of less than 1000 bp, a total of 866,007 sequences with a total length of 16.1 Gb were obtained from the 4 samples. Taxonomic profiling of the raw PacBio HiFi reads was performed using DIAMOND+MEGAN-LR workflow. Reads were aligned with DIAMOND v2.1.8 [18] blastx against the NCBI nr database (version 09/2022) using the following parameter '-range-culling -top 5- F 5000' for long-read mode, followed by 'meganization' to ensure the LCA (last common ancestor) assignment for each read. For this, the MEGAN6 Community Edition [19] 'daa2rma' function was used with the following parameter '-longReads-lcaAlgorithm longReads-ram readCount'.

See Supplementary Material for details on primary analyses of the short-read metagenomes and metatranscriptomes.

Sequence assembly and taxonomic profiling

Sequences generated by PacBio HiFi sequencing were co-assembled using Hifiasm-meta r058 [20] with default parameters, and metaFlye v2.9.2 [21] with -pacbio-hifi and -meta setting [22]. Hifiasm-meta produced a total of 16,446 contigs with an N50 of 80,906bp, and metaFlye produced a total of 6589 contigs with a N50 of 90,042bp (Table S1). As a result, only Hifiasm-meta co-assembly was further analyzed. This assembler was also chosen for sample-by-sample assemblies. See Supplementary Material and Table S2 for details on the assembly of short-read and hybrid assemblies. Prior taxonomic classification, gene prediction was conducted using FragGeneScan v1.31 [23] on the assembled contigs from both co-assemblies and individual sample assemblies. The predicted genes were then aligned to the NCBI nr database (version 09/2022) using the DIAMOND blastp tool. Taxonomy was assigned based on the best-hit criteria (lowest e-value, highest identity, and highest bit-score), with a minimum e-value threshold of 10E-5.

MAG recovery and taxonomic identification

Bins were generated using PacBio HiFi sample-by-sample assembly and co-assembly in parallel. We employed three binning tools (MetaBAT v2.12.1, MaxBin v2.2.6, and CONCOCT v1.0.0) with default settings within metaWRAP v1.3.2 [24]. The resulting bins were next combined and refined using the Bin_refinement module of metaWRAP with the parameters '-c 50—×10' to obtain at least medium-quality (i.e., completeness >50%,

contamination <10%) draft MAGs. CheckM v1.0.12 [25] lineage workflow in metaWRAP was used to assess the quality (completeness and contamination) of the MAGs and seqkit v2.3.0 [26] was used to obtain basic MAG statistics. MAGs were dereplicated using dRep v3.4.0 [27]. First, genomes were grouped at ~90% ANI based on k-mer content using Mash [28] (primary clustering). These groups were then refined by aligning genomes within each primary cluster at 95% ANI with ANImf [29] (secondary clustering). Only non-redundant genomes with the highest completeness and the lowest contamination were kept for further analysis. We used GTDB-Tk v2.1.0 [30] to assign taxonomy to these MAGs based on the taxonomy R207_v2 from the Genome Taxonomy Database (GTDB). We further re-estimated the completeness and the contamination of the Patescibacteria MAGs using the recently developed Checkm2 v0.1.3 [31], providing better estimates for this phylum. We screened for the presence of ribosomal RNAs (i.e., 16S, 23S and/or 5S rRNA) in MAGs using barrnap v0.9 (<https://github.com/tseemann/barrnap>). The same procedure was used to generate the MAGs from Illumina HiSeq sequencing, but these MAGs were used for statistical comparison only (Figure S1, Supplementary Material).

Functional annotation of predicted ORFs was done with DRAM v1.4.0 [32], which relied on Pfam, KOfam, UniProt, dbCAN, and MEROPS. We estimated the relative abundances of the MAGs across the four metagenome datasets using the CoverM v0.6.1 tool [33], by assessing the coverage of mapped long-read in 'genome' mode. To obtain relative abundance, the average genome coverage was divided by the average total coverage of all genomes, then normalized by the ratio of mapped reads to total reads. A final contamination assessment and chimerism check on the final set of MAGs was done using GUNC v1.0.5 [34].

Integrating MAGs and metabarcoding data

The prokaryotic diversity of the samples was estimated using 16S rRNA amplicon sequencing data and compared with the 16S rRNA profiles of MAGs. First, Illumina MiSeq (2×250 bases) metabarcoding data for these samples were obtained from Tláskal et al., [10]. Amplicon sequencing data were then processed using the SEED v2.1.05 pipeline [35], following the methods of Richy et al. [36]. Paired-end reads were joined using fastq-join, and sequences with ambiguous bases or a mean quality score below 30 were excluded. Chimeric sequences were identified and removed using USEARCH v11.0.667 [37], followed by clustering with UPARSE within USEARCH [38] at 97% of similarity. The most abundant sequences were chosen as representatives for each Operational Taxonomic Unit (OTU), singletons were discarded, and

OTU tables were rarefied to a common sampling depth of 9164 reads per sample using the `rarefy_even_depth()` function in the *phyloseq* R package [39]. Species-level matches were determined by BLASTn against Silva v138.1 [40]. Hill numbers were used to calculate diversity indexes (observed richness = H_0 and exponential of Shannon = eH') using the `DivPart()` function in the *entropy* R package [41]. Finally, the 16S rRNA sequences of the MAGs were aligned and mapped to the 16S rRNA amplicon sequences using `blastn` (maximum e-value of 1×10^{-5}) and `bowtie2` v2.4.1 [42], respectively. Alignment and mapping results were consistent for 45 out of the 48 MAGs possessing a 16S rRNA copy in their genome (Table S3), and agreed with the MAG GTDB taxonomy.

Functional characterization of the deadwood microbial communities

Relevant functions in wood decomposition were screened [9]. All these analyses were performed on short reads, long reads and MAGs. EukRep v0.6.7 [43] were used to predict contigs of eukaryotic origin. Metatranscriptomes were used to validate the expression of specific metabolisms.

Characterization of carbohydrate-active enzymes (CAZymes)

Genes involved in the decomposition of the easily degradable carbohydrates (arabinogalactanases, xylanases/xyloglucanases, mannanases, cellobiases and xylobiases), complex plant cell wall biopolymers (endoglucanase, cellobiohydrolase, exoglucanase), microbial biomass (peptidoglycansases, beta-glucanases, chitinases) and reserve compounds (alpha-glucanases) were predicted using `FragGeneScan` v1.31 [23], and annotated using `run_dbcan.py` (v2) program [44]. Predicted prokaryotic and eukaryotic genes were compared to the dbCAN database V9 using HMMER 3.0 [45]. See Table S4 for the details of the CAZymes screened for each activity. To validate *Myxococcota* cellulose decomposition activity, we mapped transcripts from sample 7 to *Myxococcota* MAGs (both *Myxococcota* genomes were assembled from this sample) using `minimap2` v2.24 [46] with the '-x sr' setting. Transcript counts were summed using `dirseq` v0.4.3 [47] based on gene coordination from the gene prediction.

Nitrogen metabolism investigation

Predicted genes from long-read and short-read co-assemblies were first screened with HMMER v3.3.2 against the KOfam database (accessed in June 2022) using the predefined adaptive thresholds [48]. After detecting the metabolic pathways involved in the nitrogen cycle found in a single contig, we further analyzed the contigs in which regions associated with N_2 fixation were found (i.e. *nif* genes + *fix* genes + *lsc* system genes or

Suf system genes). We performed gene prediction using `GeneMarkS` v1.14 [49] with default setting, and the annotation using `eggno-mapper` v2.1.11 [50] and `MMseqs` after translating input CDS to proteins. The phylogenetic relationships between the *nifH* sequences were analyzed using neighbor-joining trees generated using the maximum likelihood algorithm with 1000 bootstrap iterations. First, multiple sequence alignment was generated using `mafft` v7.490 [51] with the option '-maxiterate 1000' and '-localpair', then sequences were trimmed using `BMGE` v2.0 [52] software and the BLOSUM30 matrix. Finally, phylogenetic tree was computed using `iqtree` v2.2.0.3 [53] and the LG + C20 model and visualized using `FigTree` v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>). Additional analyses on the nitrogen fixation potential in *Steroidobacteraceae* MAG are described in the Supplementary Material.

Biosynthetic gene clusters

`AntiSMASH` v6.1.1 [54] was used to predict secondary metabolite clusters. BGCs were predicted with '-hmm-detection-strictness strict' and '-asf' settings. The options '-cb-general', '-cb-subclusters' and '-cb-known-clusters' were added in order to compare identified clusters against `antiSMASH`-predicted clusters database, known subclusters responsible for synthesizing precursors and known gene clusters from the MIBiG database [55] respectively. To cluster the BGCs by similarity, we used `BiG-SCAPE` v1.1.5 [56] with the '-mibig' and '-cutoffs 0.6' flags to group the BGC detected and the MIBiG database v3.1 entries at 60% of similarity.

Results

PacBio HiFi sequencing yields high-quality assemblies

The PacBio HiFi sequencing approach produced an average of 216,502 sequences per sample, with an average read length of 18,340 bp (Table S5). Eighty-five percent of the raw reads were assigned to Prokaryotes (mostly Bacteria) and 13% to Eukaryotes (Table S6). The co-assembly of PacBio HiFi reads yielded 16,446 contigs, with an N50 of 66,188 bp, including 16,439 contigs > 10 kb and 9696 contigs > 50 kb (Table S1). The largest contig reached 5.8 Mb. On average, 41% of the long reads mapped to the assembled contigs (Table S7). Most contigs were assigned to Bacteria, while 10% were assigned to Eukaryotes (Table S6). See Tables S2 and Tables S6 for details of sample-by-sample assembly.

MAGs from both abundant and rare bacterial taxa recovered with PacBio HiFi data

Binning based on PacBio HiFi assemblies generated a total of 69 unique bacterial MAGs composed of 26 ± 21 contigs with an average length of $148,129 \pm 386,395$ bp

(Table S8). Among these, 14 were High-quality draft MAGs (i.e., completeness > 90% and contamination < 5%, presence of the 23S, 16S, and 5S rRNA genes and at least 18 tRNAs) and 55 were Medium-quality draft MAGs (completeness > 50%, contamination < 10%, Figure S1). Twenty-two MAGs were composed of less than 10 contigs, among which 7 were composed of a single contig and considered nearly finished MAGs (Table S8).

These MAGs were assigned to Pseudomonadota (38 genomes), Bacteroidota (10), Verrucomicrobiota (6), Patescibacteria (6), Acidobacteriota (4), Actinomycetota (3), and Myxococcota (2) (Fig. 2A). Genomes from the most abundant bacterial phyla identified through metabarcoding were successfully binned (Fig. 2B), and the number of recovered MAGs was strongly correlated with the OTU richness of these phyla (Pearson correlation coefficient, $r = 0.84$, $p = 0.008$). However, no genomes from the Planctomycetota phylum were recovered, although their contigs were successfully

assembled (Figure S2). This likely reflects the unique community structure of this clade, which had relatively low 16S rRNA abundance (ranking as the sixth most abundant phylum) but was the second most diverse phylum after Pseudomonadota (Fig. 2B).

Out of the 48 MAGs containing a copy of the 16S rRNA gene, 44 MAGs were successfully linked to a unique OTU recovered from metabarcoding data. Among these, 24 were associated to OTUs exhibiting a relative abundance greater than 0.1% (i.e., among the 172 OTUs, Fig. 2C), including five MAGs originating from the 12 most abundant OTUs assigned to Alphaproteobacteria (3) and Myxococcota (2). Additionally, 20 MAGs corresponded to OTUs with a relative abundance below 0.1%, including two MAGs assigned to Patescibacteria and two to Alphaproteobacteria. The proportional relationship between MAG recovery and OTU relative abundance (Fig. 2D) confirmed that genomes from both abundant and rare taxa were successfully recovered.

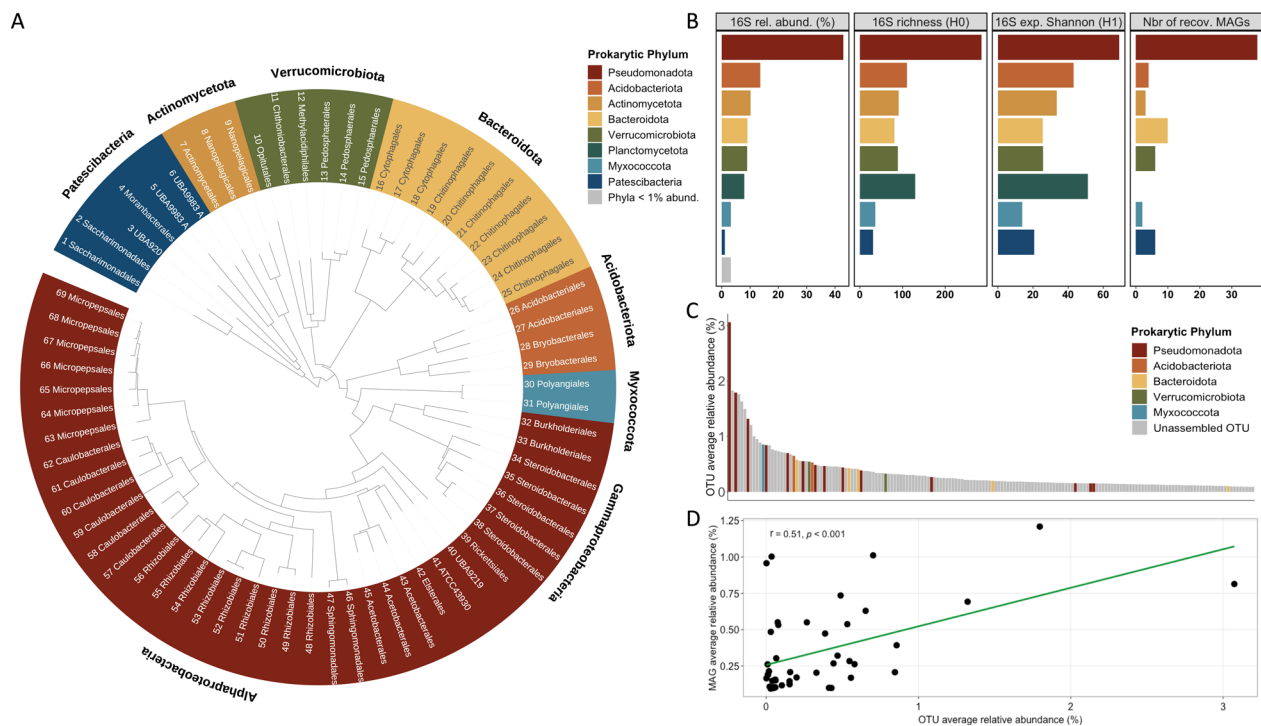


Fig. 2 Integrating de novo genome assembly and metabarcoding. **A** Maximum likelihood phylogenomic tree, classifying MAGs based on their position in the reference tree, relative evolutionary divergence, and average nucleotide identity (ANI) compared to reference genomes. The tree was generated using GTDB-tk based on the concatenated phylogeny of 120 bacterial single-copy marker genes and includes 69 unique bacterial genomes (MAGs) with completeness > 50% and contamination < 10%. **B** Relative abundance of sequences (%), phylum richness (H0) and exponential of Shannon (H1) index estimated from 16S rRNA amplicon metabarcoding and averaged across the four samples, as well as number of recovered MAGs per bacterial phylum. **C** Barplot displaying the averaged OTU relative abundance across all four samples. Only OTUs with an average relative abundance > 0.1% (top 172) are shown and ranked in descending order. The bar color corresponds to the phylum when a corresponding MAG was assembled. **D** Correlation between MAG relative abundance and sequence relative abundance of corresponding OTUs

Functional profiling reveals key bacteria involved in OM decomposition

The abundance of CAZymes was comparable across short reads, long reads, and MAGs, with alphaglucanases, peptidoglycanases, cellobiases, and xylobiases being the most common, while exoglucanases and cellobiohydrolases were relatively rare (Figure S3). MAGs assigned to Myxococcota (Polyangiaceae family) encoded the highest number of CAZymes involved in the degradation of both easily decomposable and recalcitrant biopolymers (Fig. 3). These MAGs encoded the full complement of enzymes required for cellulose decomposition, including endoglucanases, exoglucanases, and cellobiohydrolases. Importantly, while endoglucanases and exoglucanases were transcribed by other MAGs, transcript mapping revealed that cellobiohydrolases were exclusively transcribed by Polyangiaceae (Table S9). Bacteroidota also exhibited a high number of CAZymes targeting easily degradable carbohydrates, as well as microbial biomass. In contrast, Acidobacteriota and Verrucomicrobiota encoded fewer enzymes for microbial biomass recycling but many for easily degradable carbohydrates and reserve

compounds. Pseudomonadota exhibited functional versatility, encoding numerous CAZymes for microbial biomass degradation, though the distribution of enzymes for other substrates was inconsistent. The number of CAZymes found in MAGs was proportional to their genome size ($r=0.76$, $p\text{-value}<0.001$). Consequently, the phyla encoding the smallest number of CAZymes were Actinomycetota and Patescibacteria. Actinomycetota assembled in deadwood (Nanopelagiacaceae and Microbacteriaceae families) encoded mainly CAZymes targeting reserve compounds and easily degradable carbohydrates, and lacked the metabolic capacity to degrade complex plant cell wall biopolymers and microbial biomass (Fig. 3). Patescibacteria exhibited only few enzymes involved in decomposition of easily degradable carbohydrates, targeting microbial biomass and reserve compounds. However, transcript mapping confirmed expression of CAZymes by Patescibacteria, exceeding the transcription of Acidobacteriota and Pseudomonadota for mannanase and betaglucanases activity (Table S9). In addition, while overall chitinase expression was low among the MAGs recovered (mean coverage=3.0),

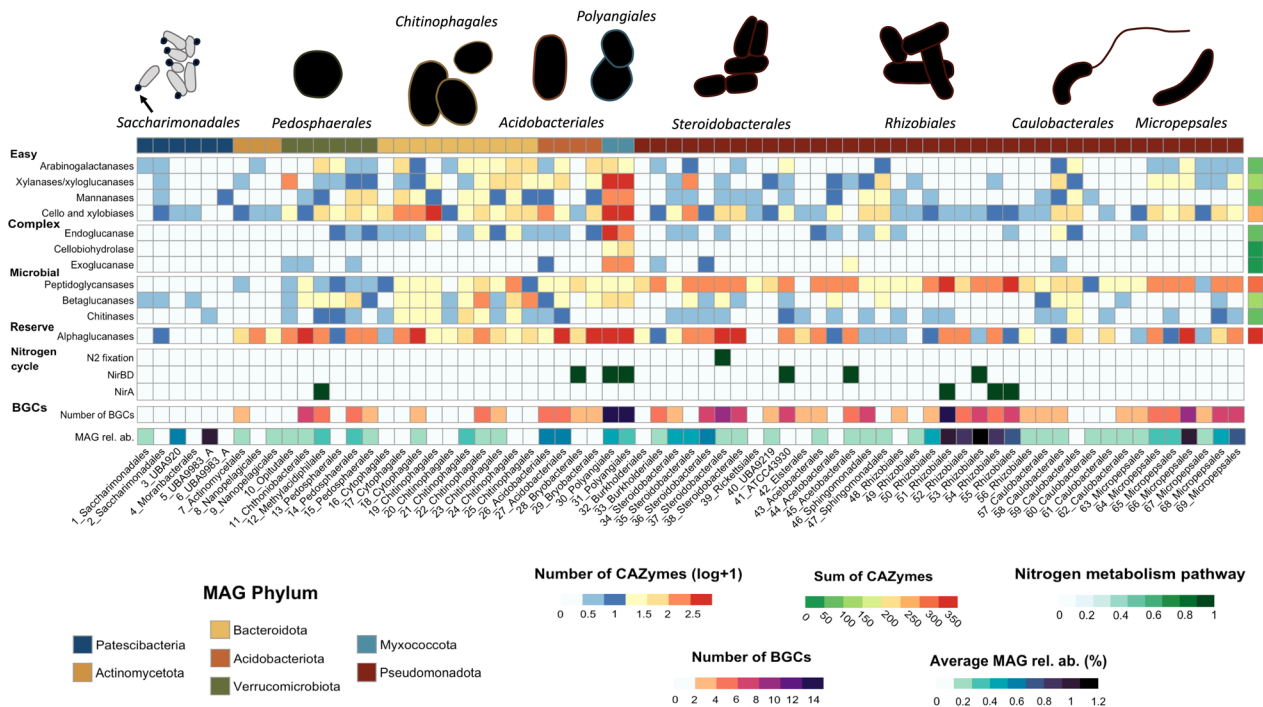


Fig. 3 Functional profiles of deadwood-associated MAGs. Heatmap representing the metabolic capacities involved in the decomposition of the easily decomposable carbohydrates (arabinogalactanases, xylanases/xyloglucanases, mannanases, cellobiases and xylobiases) and complex plant cell wall polymers (endoglucanases, cellobiohydrolases, exoglucanases), microbial biomass (peptidoglycanases, betaglucanases, chitinases), and reserve compounds (alphaglucanases), nitrogen fixation, dissimilatory nitrate reduction (nitrite to ammonia only) and assimilatory nitrate reduction (nitrite to ammonia only), and the number of biosynthetic gene clusters found in the 69 MAGs. The top row represents the phylum to which the MAG was assigned. The number of CAZymes were log transformed + 1 to improve the visualization, and the total counts of CAZymes were summed on the right last column per activity or by class of enzyme. The bottom row represents the average relative abundance of the MAG. The shape of the bacteria was drawn based on the closest characterized clades of each MAG lineages

Patescibacteria exhibited the second-highest chitinase expression after Bacteroidota (6.6 and 7.3 respectively).

New nitrogen-fixing bacteria identified in deadwood

While no complete nitrogen cycle pathways were detected on contigs from short-read sequencing, PacBio HiFi contigs and MAGs revealed the presence of dissimilatory and assimilatory nitrate reduction pathways, along with nitrogen fixation (Table S10). The *nirB* and *nirD* genes, which convert nitrite to ammonia in the dissimilatory nitrate reduction pathway, were found in Pseudomonadota (including three Alphaproteobacteria MAGs), Myxococcota (two Polyangiales), and Acidobacteriota (Bryobacteriales) (Fig. 3). The *nirA* gene, responsible for the conversion of nitrite to ammonia in the assimilatory nitrate reduction pathway, was identified in Verrucomicrobiota (Methylacidiphilales) and Alphaproteobacteria (three Rhizobiales). Nitrogenase genes (*nifH*, *nifD*, *nifK*), which catalyze the biological reduction of dinitrogen to ammonia, were exclusively present in the

MAG 36_Steroidobacteriales (a gammaproteobacterium assigned to the family Steroidobacteraceae).

The discovery of nitrogenase genes within the Steroidobacteraceae family was unexpected, but supported by subsequent analyses which indicated accurate binning, no recent N_2 -fixation gene transfer, and active nitrogen fixation by this MAG (Fig. 4 and Supplementary Material). To further characterize the nitrogen-fixing capabilities of the MAG 36_Steroidobacteriales, we compared its genomic regions associated with N_2 fixation to other contigs generated by PacBio HiFi co-assembly (Fig. 4). We identified four additional contigs containing essential nitrogen fixation genes, all assigned to Alphaproteobacteria (Table S11). The minimum set of genes required for nitrogen fixation, specifically *nifHDK* and *nifENB*, were present. Additionally, the *fixABCRUX* genes were identified, preferred over the Rnf complex for producing reduced ferredoxin/flavodoxin. Although 36_Steroidobacteriales lacked *nifS*, it encoded the complete *Isc* system (*iscR*, *iscS*, *iscU*, *iscA*, *hscB*, *hscA*, *fdx*, and *iscX*), essential

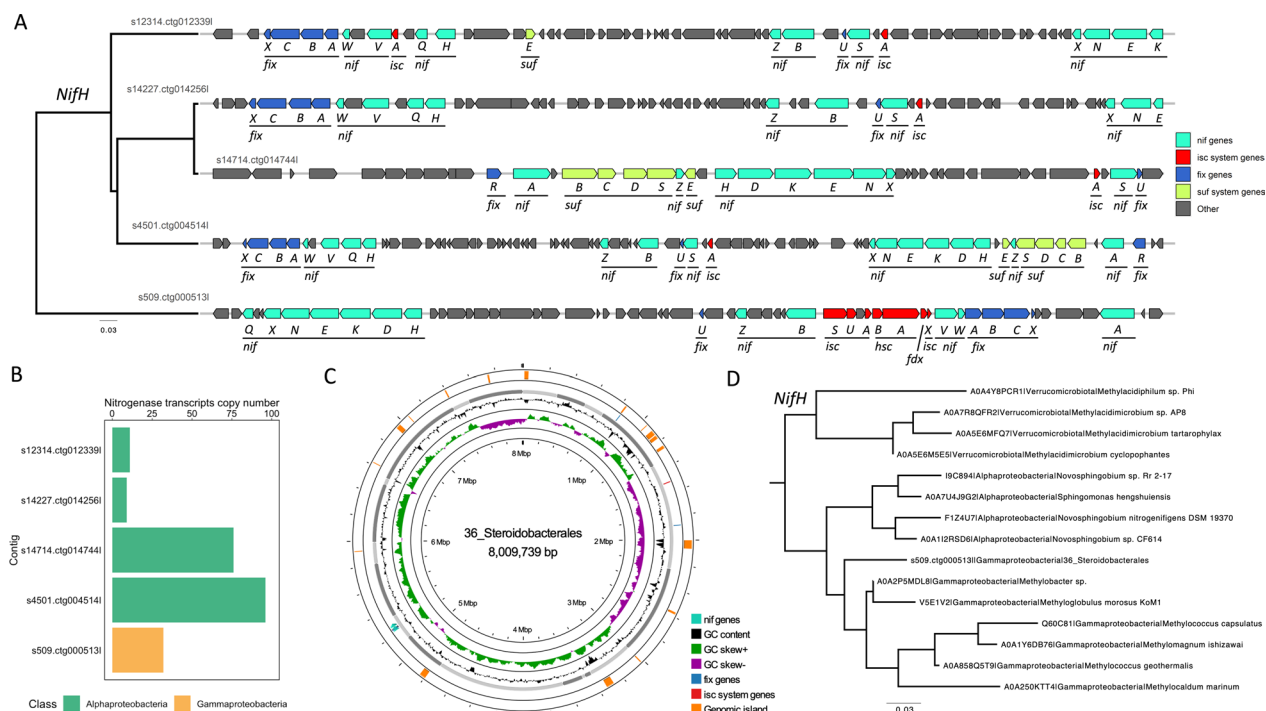


Fig. 4 Characterization of the genomic regions associated with N_2 fixation. **A** Gene organization of the five genomic regions associated with N_2 fixation assembled from PacBio HiFi assembly. Clustering is based on the phylogenetic relationships of the corresponding *NifH* sequences. Only regions associated with N_2 fixation were plotted. The genes are colored by operons and systems: in cyan the *nif* genes, in blue the *fix* genes, in red the *Isc* system genes and in green the *Suf* system genes. **B** Numbers of transcripts mapping to the nitrogenase genes (*nifHDK*) found in the five contigs (from A), belonging either to Alphaproteobacteria (green) or Gammaproteobacteria (yellow). The contig s509.ctg000513l corresponds to the contig found in the MAG 36_Steroidobacteriales. **C** Circular plot illustrating genomic islands of the MAG 36_Steroidobacteriales computed with Islandviewer4. The exact location of the nitrogen fixation contig s509.ctg000513l is between 5,118,742 and 5,335,241 bp, where no genomic island was detected. **D** Phylogenetic analysis placing the *nifH* gene found in the MAG 36_Steroidobacteriales within Gammaproteobacteria. Neighbor-joining trees were generated using the maximum likelihood algorithm with 1000 bootstrap iterations and rooted using minimal ancestor deviation. First, the *NifH* amino acid sequences from the InterPro database (accessed on August 1st, 2023) clustered at 97% of similarity were included, followed by sub-selection of the closest related taxa

for the maturation of [Fe-S] proteins. In contrast, the Alphaproteobacteria contigs lacked the *Isc* system but contained the *Suf* system (*sufB*, *sufC*, *sufD*, *sufS*, and *sufE*) instead.

Deadwood bacterial communities exhibit a wide range of biosynthetic gene clusters

A total of 1089 putative BGCs were identified in the PacBio HiFi long-read co-assembly, presenting a stark contrast to the three BGCs found in the short-read co-assembly (Figure S4). A comparison with the MIBig database highlighted the novelty of these BGCs, with a staggering 98% showing less than 50% shared genes with MIBig entries. The primary BGC types identified in the long-read assembly were also found in MAGs, enabling the association of biosynthetic potential with taxonomic groups (Table S12).

Pseudomonadota, particularly MAGs from Rhizomicrobium, Rhizobiales (Alphaproteobacteria), and Steroidibacteriaceae (Gammaproteobacteria), carried the majority of BGCs, exceeding 10 in some instances (Fig. 5A). Pseudomonadotal BGCs presented a wide diversity, encompassing various types of compounds such as thioamitides, lasso peptides, ranthipeptides, azol(in)e-containing linear peptides, linaridines, or ladderanes, each holding bioprospecting potential. The two Myxococcota MAGs displayed the highest numbers of BGCs (14 and 15), encoding a wide array of compounds, including thioamitides and ranthipeptides. Verrucomicrobiota,

Acidobacteriota, Bacteroidota, and Actinomycetota MAGs harbored 0–8 BGCs, with terpene, RiPP, polypeptide, NRPS, and arylpolyene BGCs being the most prevalent. Group-specific BGCs, such as flexirubins in Chitinophagaceae (Bacteroidota) MAGs, were also identified. Patascibacteria MAGs did not contain any BGCs. BiG-SCAPE analysis revealed that the 271 BGCs identified in MAGs were categorized into 254 unique gene cluster families (Fig. 5B). The abundance of singletons underscored the vast diversity of BGCs in deadwood MAGs. For instance, seven BGCs encoding thioamitides each belonged to a distinct gene cluster family (Fig. 5C). BGCs belonging to one cluster family always belonged to taxonomically related MAGs, indicating that BGCs in deadwood bacteria are phylogenetically conserved.

Discussion

PacBio HiFi metagenomic sequencing has demonstrated outstanding efficiency for reconstructing genomes from deadwood microbial communities, including Myxococcota, which could not be previously recovered through cultivation or short-read metagenomics from these samples [10, 15]. This technology enabled the assembly of both abundant and rare bacterial genomes, allowing better characterization of the deadwood bacterial communities. A total of 69 bacterial genomes were generated from 4 deadwood samples, including 67 newly reconstructed MAGs, 14 high-quality and 7 composed of a single contig. While PacBio HiFi generated a comparable amount

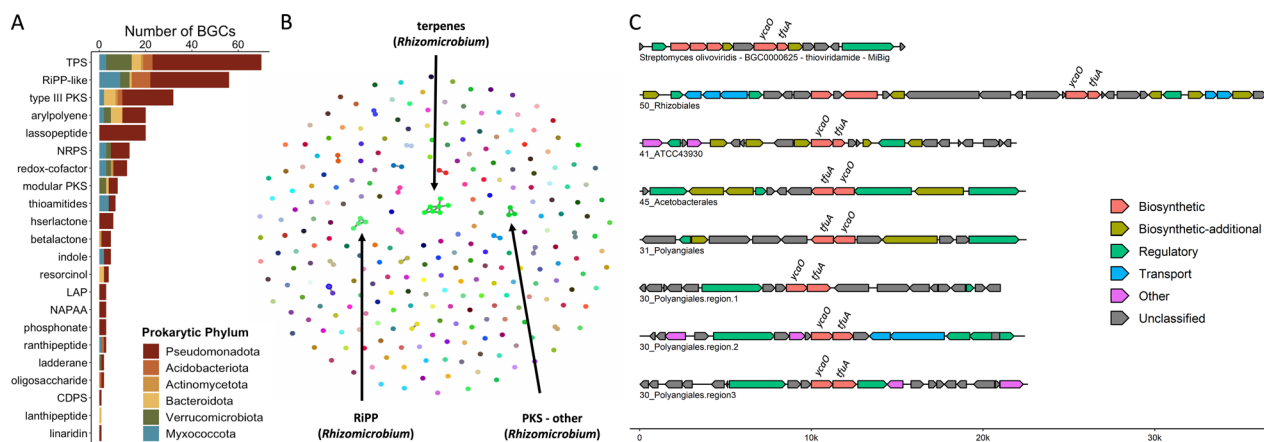


Fig. 5 Biosynthetic gene cluster diversity in MAGs. **A** Barplot summarizing the number of BGCs identified in MAGs derived from long-read co-assembly. BGCs are categorized by type and ranked in descending order. Colors indicate the number of BGCs per phylum. TPS, terpenes; RiPP, ribosomally synthesised and post-translationally modified peptides; PKS, polyketide synthase product; NRPS, non-ribosomal peptide synthetase product; LAP, linear azol(in)e-containing peptides; NAPAA, non-alpha poly-amino acids like e-polylysine; CDPS, tRNA-dependent cyclodipeptide synthase product. **B** Similarity network of 279 BGCs found in 69 MAGs. BGCs belonging to the same gene cluster family are linked. Gene cluster families with more than 2 members are labeled with the corresponding product type and taxonomic classification of the MAGs in which they occurred (in parenthesis, all BGCs within one gene cluster family always originated from the same genus). **C** Organization of BGCs encoding thioamitides. Seven thioamitide BGCs found in 5 MAGs (each belonging to a unique gene cluster family) are displayed together with thioviridamide BGC from *Streptomyces olivoviridis* as a reference. The core biosynthetic genes *ycaO* and *tfuA* are labeled

of data to Illumina sequencing (an average of 4.0 Gb versus 4.9 Gb per sample respectively, Table S5), it outperformed Illumina in binning (69 versus 11 MAGs), yielding less fragmented, less contaminated, and more complete MAGs. In comparison to related studies, our findings overtake those of Tláskal et al. [10], who assembled 58 MAGs from 25 short-read metagenomes of deadwood from the same forest. Our investigation extended to the eight most abundant deadwood bacterial phyla, showing successful genome assembly for all phyla except Planctomycetota. Our results suggest that while genome size and relative abundance do not appear to be the ultimate limiting factors for successful binning, the inherent complexity of Planctomycetota, characterized by high alpha diversity, challenges their genome recovery.

Employing PacBio HiFi sequencing, we underscored the important role of Myxococcota in deadwood decomposition. While Myxococcota have been previously assembled from various ecosystems, including aquatic, terrestrial, host-associated, and built environments [57], our study marks the first assembly of Myxococcota genomes from deadwood samples. The successful recovery of these genomes using long-read sequencing—where previous short-read sequencing efforts failed [10]—is likely due to the large size of Myxococcota genomes (>8 Mbp, according to the GTDB database r202) and the frequent presence of complex genomic features, such as insertion sequences, prophages, genomic islands, and hypervariable regions with high single nucleotide polymorphism density [58]. These features are challenging to assemble into MAGs with short-read data. Long-read sequencing, with its capacity to span repetitive and variable regions, is thus more effective at capturing the complete and contiguous genomes of Myxococcota.

Myxococcota exhibit unique traits, encompassing motility [59], predation [60], fruiting bodies formation [61], and the ability to decompose cellulose [62] and catabolize aromatic lignin depolymerization products [63]. Here, we successfully assembled two abundant Myxococcota (Polyangiaceae) genomes expressing high cellulolytic activity, including endoglucanase, exoglucanase, and cellobiohydrolase – enzymes exclusively found in Myxococcota MAGs. While Polyangiaceae also expressed CAZymes targeting bacterial biomass, the lack of chitinase prevented them from recycling fungal biomass (Fig. 3). Conversely, chitinases were consistently present in Bacteroidota and detected in Verrucomicrobiota, Pseudomonadota and Patescibacteria. Notably, although chitinase was found in only one copy in 5_UBA9983_A, this patescibacterial genome exhibited the second-highest chitinase expression.

Patescibacteria had been identified in a wide range of ecosystems [57, 64] including deadwood habitats [10, 65],

yet their ecological role remains enigmatic. While their limited genome size, lacking crucial metabolic genes, suggests an obligate epibiotic lifestyle [66], their parasitic status remains subject to debate [67]. In our study, UBA9983_A demonstrated a possibility for specialized opportunistic lifestyle, primarily recycling chitin. Additionally, Patescibacteria of the order Saccharimonadales exhibited high expression of mannase, an enzyme involved in hemicellulose degradation, highlighting their active contribution in wood decomposition processes. These findings collectively imply that Patescibacteria may not solely depend on host resources. Further analyses are needed to elucidate whether hemicellulose-derived residues and chitin serve as energy sources for ATP generation or benefits to the host (e.g., commensalism or mutualism). Nevertheless, our results suggest a broader ecological relationship for Patescibacteria beyond parasitism.

Nitrogen concentration in deadwood is low [9] and represents a limiting factor for efficient decomposition. As metabolic pathways involved in nitrogen assimilation were scarce in the studied samples (Table S10), microbial biomass could represent an alternative source of nitrogen [68]. We found that CAZymes targeting bacterial biomass were more frequent than CAZymes targeting fungal biomass, but that peptidoglycanases were 2 times less expressed than chitinases (Table S4, Table S9). This result is surprising since fungal biomass contains less nitrogen than bacterial biomass [69], but these processes might simply be regulated by the availability of bacterial and eukaryotic biomass. While microbial biomass decomposition represents an interesting strategy for recycling nitrogen during decay, an external source of nitrogen is required to initiate decomposition of fresh dead wood. The initial nitrogen input is likely provided by bacterial nitrogen fixation, as illustrated by nitrogen fixation rates being eight times higher in young deadwood compared to old deadwood [9]. The conversion of atmospheric N_2 to NH_3 , being a highly energy-consuming process [70], is only realized if no better suitable nitrogen sources are available [71]. The expression of nitrogenase genes at intermediate stages of decomposition (>4 years) thus indicates that the recycling of microbial biomass does not completely meet the microbial nitrogen demand throughout the decomposition process.

Despite the fact that biological nitrogen fixation has been extensively investigated (e.g. [72, 73, 74]), our study marks the first observation of this function within the Steroidobacteraceae family. Beyond reporting a novel nitrogen-supplying bacteria in a nitrogen-limited environment, PacBio HiFi sequencing facilitated the investigation of genes involved in the conversion of N_2 into ammonium. In the deadwood ecosystem,

Pseudomonadota catalyze the production of reduced ferredoxin/ferredoxin using the *fix* operon, particularly advantageous under oxygen-limited conditions [75]. In addition, Steroidobacteraceae employ the *isc* system for biosynthesizing [Fe-S] proteins, crucial under elevated oxygen conditions [76]. The coexistence of the *fix* operon and the *isc* system likely allows Steroidobacteraceae to maintain nitrogen fixation activity under oxygen concentration fluctuations. Alphaproteobacteria appear less sensitive to elevated oxygen, favoring the *suf* system over the *isc* system, which is known to be more beneficial for bacterial growth in the presence of hydrogen peroxide [77]. Although the proportional contribution of Steroidobacteraceae (Gammaproteobacteria) and Alphaproteobacteria to global deadwood nitrogen fixation requires further exploration, our study provides empirical evidence of their in situ activity.

PacBio HiFi sequencing unveiled the remarkable potential of deadwood microorganisms for the production of diverse secondary metabolites, an important feature that would remain unnoticed in short-read data from this study. We identified over a thousand mostly novel BGCs, showcasing the extensive diversity of these BGCs in MAGs. This diversity suggests that deadwood bacteria display multiple interactions with other microorganisms in deadwood. Furthermore, our investigation highlighted bacterial groups in deadwood that hold promise for the production of novel bioactive compounds. Similar to observations in soil [78], the abundance of BGC types varied by taxonomy. In deadwood, we observed high numbers of BGCs in Myxococcota and Pseudomonadota, as well as in other groups such as Verrucomicrobiota and Acidobacteriota, whose biosynthetic potential has only recently been reported [79, 80]. Notably, we did not identify a shared BGC family across different bacterial taxa, indicative of a compound with broader importance in deadwood. In contrast, phylogenetic conservation of BGC families in deadwood MAGs was observed. The presence of BGCs for flexirubins, pigments typical of Bacteroidota [81], in several Chitinophagaceae MAGs constitutes convincing proof of validity of our metagenome analysis.

Long-read metagenomic analysis, while powerful, still has limitations. More effort is needed to appropriately describe the Planctomycetota: the high alpha diversity of this phylum in deadwood may require an increased sequencing depth for their genome recovery, as illustrated by the fact that Tláskal et al. [9] was able to recover one Planctomycetota MAG when combining short-read data from 25 deadwood samples (139 Gb) from the same forest. Nevertheless, long-read metagenomic sequencing provides substantial advantages in uncovering microbiome functions that are challenging to capture using

short-read metagenomics. In our study, the short-read approach appeared reliable for identifying single-gene metabolic functions such as CAZymes, but failed to capture genomic features involving multiple genes, such as nitrogen fixation and biosynthetic gene clusters. Additionally, PacBio HiFi assemblies yielded more eukaryotic contigs than Illumina assemblies, although both platforms generated a proportional number of eukaryotic reads. This discrepancy is thus unlikely to stem from differences in the initial sample preparation for long-read sequencing (e.g., less intensive DNA extraction) but rather reflects the longer sequencing lengths and higher accuracy of PacBio HiFi sequencing [82]. Eukaryotic genomes, characterized by intricate features like repetitive regions, introns, and exons [83], require greater metagenomic sequencing depth and sample size to be successfully assembled compared to prokaryotic genomes. For instance, Saraiva et al., [84] needed 6000 terrestrial metagenomes to assemble 197 eukaryotic bins, whereas Ma et al., [85] achieved the assembly of 40,039 prokaryotic MAGs from 2941 soil metagenomes.

In conclusion, the PacBio HiFi sequencing enabled the assembly of both rare and abundant bacterial genomes, facilitating the assembly of novel genomes of bacteria with key roles in deadwood decomposition, such as cellulose decomposition in Polyangiaceae and nitrogen fixation in Steroidobacteraceae. Additionally, it revealed the significant contribution of Patescibacteria to wood decomposition processes and identified a wealth of new biosynthetic gene clusters with potential ecological and biotechnological significance.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40793-024-00639-5>.

Supplementary material 1.

Supplementary material 2.

Supplementary material 3.

Acknowledgements

The authors are grateful to the Nature Conservation Agency of the Czech Republic for granting the permit to perform environmental sampling within the Zofínský prales National Natural Reserve. The authors acknowledge BioRender, which was used for figure creation.

Author contributions

M.K. and P.B. designed the study. E.R., M.K. and P.B. directed the analyses. M.K. and V.T. performed HMW DNA isolation. P.T.D. and E.R. carried out the computational analyses and visual representations. E.R. and M.K. analyzed the data with the help of P.T.D., V.T., R.L.M. and P.B. E.R. and M.K. wrote the original manuscript, with the input of P.T.D., V.T., R.L.M. and P.B. All authors have read and approved the final manuscript.

Funding

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (CZ.02.01.01/00/22_008/0004635—AdAgriF—Advanced

methods of greenhouse gases emission reduction and sequestration in agri-culture and forest landscape for climate change mitigation).

Availability of data and materials

Descriptions of the deadwood samples 6, 7, 57 and 84 are available at the NCBI BioSample repository (<https://www.ncbi.nlm.nih.gov/biosample/>), under accession numbers SAMN13925154, SAMN13925155, SAMN13925167, and SAMN13925168, respectively. The raw PacBio HiFi sequences from the four deadwood samples are available at the NCBI Sequence Read Archive repository (<https://www.ncbi.nlm.nih.gov/sra/>), under accession numbers SRR28211698—SRR28211701. The co-assembly of PacBio HiFi sequences is available at NCBI GenBank (<https://www.ncbi.nlm.nih.gov/nuccore/>) under accession number JBBBCBH000000000. The 69 metagenome-assembled genomes from PacBio HiFi sequences are available at the NCBI GenBank repository, under accession numbers JBBBCFX000000000—JBBBCIN000000000. The raw Illumina HiSeq metagenome sequences corresponding to samples 6, 7, 57 and 84 are available at the NCBI Sequence Read Archive repository, under accession numbers SRR10968229, SRR10968228, SRR10968259, and SRR10968258, respectively. The Illumina HiSeq metatranscriptome sequences of samples 6 and 7 are available at the NCBI Sequence Read Archive repository, under accession numbers SRR10968251 and SRR10968250. The Illumina MiSeq 16S rRNA amplicon sequencing data for samples 6, 7, 57 and 84 are available at the NCBI Sequence Read Archive repository, under accession numbers SRR12914735, SRR12914734, SRR12914771, and SRR12914770. The biosynthetic gene clusters recovered from Illumina short-read metagenome, PacBio HiFi long-read metagenome and from PacBio HiFi MAGs, are available at the Zenodo repository (<https://zenodo.org>), under record numbers 10550953, 10529179, and 10529038, respectively.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Laboratory of Environmental Microbiology, Institute of Microbiology of the Czech Academy of Sciences, Videňská 1083, 14200 Prague 4, Czech Republic. ²Institute of Soil Biology and Biogeochemistry, Biology Centre of the Czech Academy of Sciences, Na Sádkách 7, 37005 České Budějovice, Czech Republic. ³Department of Soil and Water Conservation and Waste Management, CEBAS-CSIC, Campus Universitario de Espinardo, 30100 Murcia, Spain.

Received: 14 May 2024 Accepted: 11 November 2024

Published online: 03 December 2024

References

- Seibold S, Rammer W, Hothorn T, Seidl R, Ulyshen MD, Lorz J, Cadotte MW, Lindennmayer DB, Adhikari YP, Aragón R, Bae S, Baldrian P, Barimani Varandi H, Barlow J, Bässler C, Beauchêne J, Berenguer E, Bergamin RS, Birkemoe T, Boros G, Brandl R, Brustel H, Burton PJ, Cakpo-Tossou YT, Castro J, Cateau E, Cobb TP, Farwig N, Fernández RD, Firn J, Gan KS, González G, Gossner MM, Habel JC, Hébert C, Heibl C, Heikkala O, Hemp A, Hemp C, Hjältén J, Hotes S, Kouki J, Lachat T, Liu J, Liu Y, Luo Y-H, Macandog DM, Martina PE, Mukul SA, Nachin B, Nisbet K, O'Halloran J, Oxbrough A, Pandey JN, Pavlíček T, Pawson SM, Rakotondranary JS, Ramanamanjato J-B, Rossi L, Schmidl J, Schulze M, Seaton S, Stone MJ, Stork NE, Suran B, Sverdrup-Thygeson A, Thorn S, Thyagarajan G, Wardlaw TJ, Weisser WW, Yoon S, Zhang N, Müller J. The contribution of insects to global forest deadwood decomposition. *Nature*. 2021;597:77–81. <https://doi.org/10.1038/s41586-021-03740-8>.
- Baldrian P, López-Mondéjar R, Kohout P. Forest microbiome and global change. *Nat Rev Microbiol*. 2023;21:487–501. <https://doi.org/10.1038/s41579-023-00876-4>.
- Baldrian P. Forest microbiome: diversity, complexity and dynamics. *FEMS Microbiol Rev*. 2016. <https://doi.org/10.1093/femsre/fuw040>.
- Johnston SR, Boddy L, Weightman AJ. Bacteria in decomposing wood and their interactions with wood-decay fungi. *FEMS Microbiol Ecol*. 2016;92:179. <https://doi.org/10.1093/femsec/fiw179>.
- de Boer W, Folman LB, Summerbell RC, Boddy L. Living in a fungal world: impact of fungi on soil bacterial niche development. *FEMS Microbiol Rev*. 2005;29:795–811. <https://doi.org/10.1016/j.femsre.2004.11.005>.
- Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, Aerts A, Asiegbu FO, Baker SE, Barry K, Bendiksby M, Blumentritt M, Coutinho PM, Cullen D, De Vries RP, Gathman A, Goodell B, Henrissat B, Ihrmark K, Kauserud H, Kohler A, LaButti K, Lapidus A, Lavin JL, Lee Y-H, Lindquist E, Lilly W, Lucas S, Morin E, Murat C, Oguiza JA, Park J, Pisabarro AG, Riley R, Rosling A, Salamov A, Schmidt O, Schmutz J, Skrede I, Stenlid J, Wiebenga A, Xie X, Kües U, Hibbett DS, Hoffmeister D, Höglberg N, Martin F, Grigoriev IV, Watkinson SC. The plant cell wall—decomposing machinery underlies the functional diversity of forest fungi. *Science*. 2011;333:762–5. <https://doi.org/10.1126/science.1205411>.
- Rinne-Garmston KT, Peltoniemi K, Chen J, Peltoniemi M, Fritze H, Mäkipää R. Carbon flux from decomposing wood and its dependency on temperature, wood N₂ fixation rate, moisture and fungal composition in a Norway spruce forest. *Glob Change Biol*. 2019;25:1852–67. <https://doi.org/10.1111/gcb.14594>.
- Hoppe B, Krüger D, Kahl T, Arnstadt T, Buscot F, Bauhus J, Wubet T. A pyrosequencing insight into sprawling bacterial diversity and community dynamics in decaying deadwood logs of *Fagus sylvatica* and *Picea abies*. *Sci Rep*. 2015;5:9456. <https://doi.org/10.1038/srep09456>.
- Tláškal V, Brabcová V, Větrovský T, Jomura M, López-Mondéjar R, Oliveira Monteiro LM, Saraiva JP, Human ZR, Cajthami T, Nunes Da Rocha U, Baldrian P. Complementary roles of wood-inhabiting fungi and bacteria facilitate deadwood decomposition. *mSystems*. 2021;6:e01078-20. <https://doi.org/10.1128/mSystems.01078-20>.
- Tláškal V, Brabcová V, Větrovský T, López-Mondéjar R, Monteiro LMO, Saraiva JP, Da Rocha UN, Baldrian P. Metagenomes, metatranscriptomes and microbiomes of naturally decomposing deadwood. *Sci Data*. 2021;8:198. <https://doi.org/10.1038/s41597-021-00987-8>.
- Tláškal V, Zrůstová P, Vrška T, Baldrian P. Bacteria associated with decomposing dead wood in a natural temperate forest. *FEMS Microbiol Ecol*. 2017. <https://doi.org/10.1093/femsec/fix157>.
- Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol*. 2017;2:1533–42. <https://doi.org/10.1038/s41564-017-0012-7>.
- Bickhart DM, Kolmogorov M, Tseng E, Portik DM, Korobeynikov A, Tolstoganov I, Uritskiy G, Liachko I, Sullivan ST, Shin SB, Zorea A, Andreu VP, Panke-Buisse K, Medema MH, Mizrahi I, Pevzner PA, Smith TPL. Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. *Nat Biotechnol*. 2022;40:711–9. <https://doi.org/10.1038/s41587-021-01130-z>.
- Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, Wollenberg RD, Albertsen M. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat Methods*. 2022;19:823–6. <https://doi.org/10.1038/s41592-022-01539-7>.
- Tláškal V, Baldrian P. Deadwood-inhabiting bacteria show adaptations to changing carbon and nitrogen availability during decomposition. *Front Microbiol*. 2021;12:685303. <https://doi.org/10.3389/fmicb.2021.685303>.
- Pebesma E. Simple features for R: standardized support for spatial vector data. *R J*. 2018;10:439. <https://doi.org/10.32614/RJ-2018-009>.
- Sagova-Mareckova M, Cermak L, Novotna J, Plhaczkova K, Forstova J, Kopecky J. Innovative methods for soil DNA purification tested in soils with widely differing characteristics. *Appl Environ Microbiol*. 2008;74:2902–7. <https://doi.org/10.1128/AEM.02161-07>.
- Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 2021;18:366–8. <https://doi.org/10.1038/s41592-021-01101-x>.
- Huson DH, Beier S, Flade I, Górski A, El-Hadidi M, Mitra S, Ruscheweyh H-J, Tappu R. MEGAN community edition—interactive exploration and

- analysis of large-scale microbiome sequencing data. *PLOS Comput Biol*. 2016;12:e1004957. <https://doi.org/10.1371/journal.pcbi.1004957>.
20. Feng X, Cheng H, Portik D, Li H. Metagenome assembly of high-fidelity long reads with hifiasm-meta. *Nat Methods*. 2022;19:671–4. <https://doi.org/10.1038/s41592-022-01478-3>.
 21. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, Pevzner PA. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods*. 2020;17:1103–10. <https://doi.org/10.1038/s41592-020-00971-x>.
 22. Kim CY, Ma J, Lee I. HiFi metagenomic sequencing enables assembly of accurate and complete genomes from human gut microbiota. *Nat Commun*. 2022;13:6367. <https://doi.org/10.1038/s41467-022-34149-0>.
 23. Rho M, Tang H, Ye Y. FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res*. 2010;38:e191–e191. <https://doi.org/10.1093/nar/gkq747>.
 24. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*. 2018;6:158. <https://doi.org/10.1186/s40168-018-0541-1>.
 25. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25:1043–55. <https://doi.org/10.1101/gr.186072.114>.
 26. Shen W, Le S, Li Y, Hu F (2016) SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *Plos One* 11(10):e0163962. <https://doi.org/10.1371/journal.pone.0163962>
 27. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J*. 2017;11:2864–8. <https://doi.org/10.1038/ismej.2017.126>.
 28. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol*. 2016;17:132. <https://doi.org/10.1186/s13059-016-0997-x>.
 29. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci*. 2009;106:19126–31. <https://doi.org/10.1073/pnas.0906412106>.
 30. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics*. 2019. <https://doi.org/10.1093/bioinformatics/btz848>.
 31. Chklovskii A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods*. 2023;20:1203–12. <https://doi.org/10.1038/s41592-023-01940-w>.
 32. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narroze AB, Rodríguez-Ramos J, Bolduc B, Gazitúa MC, Daly RA, Smith GJ, Vik DR, Pope PB, Sullivan MB, Roux S, Wrighton KC. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res*. 2020;48:8883–900. <https://doi.org/10.1093/nar/gkaa621>.
 33. Aroney STN, Newell RJP, Nissen J, Camargo AP, Tyson GW, Woodcroft BJ. CoverM: read coverage calculator for metagenomics; 2024. <https://doi.org/10.5281/ZENODO.10531253>
 34. Orakov A, Fullam A, Coelho LP, Khedkar S, Szklarczyk D, Mende DR, Schmidt TSB, Bork P. GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome Biol*. 2021;22:178. <https://doi.org/10.1186/s13059-021-02393-0>.
 35. Větrovský T, Baldrian P, Morais D. SEED 2: a user-friendly platform for amplicon high-throughput sequencing data analyses. *Bioinformatics*. 2018;34:2292–4. <https://doi.org/10.1093/bioinformatics/bty071>.
 36. Richy E, Fort T, Odriozola I, Kohout P, Barbi F, Martinovic T, Tupek B, Adamczyk B, Lehtonen A, Mäkipää R, Baldrian P. Phosphorus limitation promotes soil carbon storage in a boreal forest exposed to long-term nitrogen fertilization. *Glob Change Biol*. 2024;30:e17516. <https://doi.org/10.1111/gcb.17516>.
 37. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26:2460–1. <https://doi.org/10.1093/bioinformatics/btq461>.
 38. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*. 2013;10:996–8. <https://doi.org/10.1038/nmeth.2604>.
 39. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*. 2013;8:e61217.
 40. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2012;41:D590–6. <https://doi.org/10.1093/nar/gks1219>.
 41. Marcon E, Hérault B. entropart: an R package to measure and partition diversity. *J Stat Softw*. 2015;68:1–26.
 42. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9. <https://doi.org/10.1038/nmeth.1923>.
 43. West PT, Probst AJ, Grigoriev IV, Thomas BC, Banfield JF. Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome Res*. 2018;28:569–80. <https://doi.org/10.1101/gr.228429.117>.
 44. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2018;46:W95–101.
 45. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39:W29–37. <https://doi.org/10.1093/nar/gkr367>.
 46. Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics*. 2021;37:4572–4. <https://doi.org/10.1093/bioinformatics/btab705>.
 47. Woodcroft BJ, Singleton CM, Boyd JA, Evans PN, Emerson JB, Zayed AAF, Hoelzle RD, Lamberton TO, McCalley CK, Hodgkins SB, Wilson RM, Purvine SO, Nicora CD, Li C, Frolking S, Chanton JP, Crill PM, Saleska SR, Rich VI, Tyson GW. Genome-centric view of carbon processing in thawing permafrost. *Nature*. 2018;560:49–54. <https://doi.org/10.1038/s41586-018-0338-1>.
 48. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*. 2020;36:2251–2. <https://doi.org/10.1093/bioinformatics/btz859>.
 49. Besemer J. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res*. 2001;29:2607–18. <https://doi.org/10.1093/nar/29.12.2607>.
 50. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol*. 2021;38:5825–9. <https://doi.org/10.1093/molbev/msab293>.
 51. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. 2019;20:1160–6. <https://doi.org/10.1093/bib/bbx108>.
 52. Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol*. 2010;10:210. <https://doi.org/10.1186/1471-2148-10-210>.
 53. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32:268–74. <https://doi.org/10.1093/molbev/msu300>.
 54. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res*. 2021;49:W29–35. <https://doi.org/10.1093/nar/gkab335>.
 55. Terlouw BR, Blin K, Navarro-Muñoz JC, Avalon NE, Chevrette MG, Egbert S, Lee S, Meijer D, Recchia MJ, Reitz ZL, van Santen JA, Selem-Mojica N, Tørring T, Zaroubi L, Alanjary M, Aleti G, Aguilera C, Al-Salihi SAA, Augustijn HE, Avelar-Rivas JA, Avitia-Domínguez LA, Barona-Gómez F, Bernaldo-Aguero J, Bielinski VA, Biermann F, Booth TJ, Carrion Bravo VJ, Castelo-Branco R, Chagas FO, Cruz-Morales P, Du C, Duncan KR, Gavrilidou A, Gayraud D, Gutiérrez-García K, Haslinger K, Helfrich EJJ, van der Hoof JJJ, Jati AP, Kalkreuter E, Kalyvas N, Kang KB, Kautsar S, Kim W, Kunjapur AM, Li Y-X, Lin G-M, Loureiro C, Louwen JJR, Louwen NLL, Lund G, Parra J, Philmus B, Pourmohsenin B, Pronk LJ, Rego A, Rex DAB, Robinson S, Rosas-Becerra LR, Roxborough ET, Schorn MA, Scobie DJ, Singh KS, Sokolova N, Tang X, Udway D, Vigneshwari A, Vind K, Vromans SPJM, Waschulin V, Williams SE, Winter JM, Witte TE, Xie H, Yang D, Yu J, Zdouc M, Zhong Z, Collemare J, Lington RG, Weber T, Medema MH. MIBIG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res*. 2023;51:D603–10. <https://doi.org/10.1093/nar/gkac1049>.

56. Navarro-Muñoz JC, Selem-Mojica N, Mullooney MW, Kautsar SA, Tryon JH, Parkinson EI, De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S, Roeters A, Lokhorst W, Fernandez-Guerra A, Cappellini LTD, Goering AW, Thomson RJ, Metcalf WW, Kelleher NL, Barona-Gomez F, Medema MH. A computational framework to explore large-scale biosynthetic diversity. *Nat Chem Biol*. 2020;16:60–8. <https://doi.org/10.1038/s41589-019-0400-9>.
57. Nayfach S, Roux S, Seshadri R, Udway D, Varghese N, Schulz F, Wu D, Paez-Espino D, Chen I-M, Huntemann M, Palaniappan K, Ladau J, Mukherjee S, Reddy TBK, Nielsen T, Kirton E, Faria JP, Edirisinghe JN, Henry CS, Jungbluth SP, Chivian D, Dehal P, Wood-Charlson EM, Arkin AP, Tringe SG, Visel A, IMG/M Data Consortium, Abreu H, Acinas SG, Allen E, Allen MA, Alteio LV, Andersen G, Anesio AM, Attwood G, Avila-Magaña V, Badis Y, Bailey J, Baker B, Baldrian P, Barton HA, Beck DAC, Becraft ED, Beller HR, Beman JM, Bernier-Latmani R, Berry TD, Bertagnolli A, Bertilsson S, Bhatnagar JM, Bird JT, Blanchard JL, Blumer-Schuetz SE, Bohannan B, Borton MA, Brady A, Brawley SH, Brodie J, Brown S, Brum JR, Brune A, Bryant DA, Buchan A, Buckley DH, Buongiorno J, Cadillo-Quiroz H, Caffrey SM, Campbell AN, Campbell B, Carr S, Carroll J, Cary SC, Cates AM, Cattolico RA, Cavicchioli R, Chistoserdova L, Coleman ML, Constant P, Conway JM, Mac Cormack WP, Crowe S, Crump B, Currie C, Daly R, DeAngelis KM, Denev V, Denman SE, Desta A, Dionisi H, Dodsworth J, Dombrowski N, Donohue T, Dopson M, Driscoll T, Dunfield P, Dupont CL, Dynarski KA, Edgcomb V, Edwards EA, Elshahed MS, Figueroa I, Flood B, Fortney N, Fortunato CS, Francis C, Gachon CMM, Garcia SL, Gazitua MC, Gentry T, Gerwick L, Gharechahi J, Girguis P, Gladden J, Gradoville M, Grasby SE, Gräver K, Grottenberger CL, Gruninger RJ, Guo J, Habteselassie MY, Hallam SJ, Hatzenpichler R, Hausmann B, Hazen TC, Hedlund B, Henny C, Herfort L, Hernandez M, Hershey OS, Hess M, Hollister EB, Hug LA, Hunt D, Jansson J, Jarett J, Kadnikov VV, Kelly C, Kelly R, Kelly W, Kerfeld CA, Kimbrel J, Klassen JL, Konstantinidis KT, Lee LL, Li W-J, Loder AJ, Loy A, Lozada M, MacGregor B, Magnabosco C, Maria da Silva A, McKay RM, McMahon K, McSweeney CS, Medina M, Meredith L, Mizzi J, Mock T, Mopper L, Moran MA, Morgan-Lang C, Moser D, Muyzer G, Myrold D, Nash M, Nesbø CL, Neumann AP, Neumann RB, Noguera D, Northen T, Norton J, Nowinski B, Nüsslein K, O'Malley MA, Oliveira RS, Maia De Oliveira V, Onstott T, Osvatic J, Ouyang Y, Pachiadaki M, Parnell J, Partida-Martinez LP, Peay KG, Pelletier D, Peng X, Pester M, Pett-Ridge J, Peura S, Pjevac P, Plominsky AM, Poehlein A, Pope PB, Ravin N, Redmond MC, Reiss R, Rich V, Rinke C, Rodrigues JLM, Rodriguez-Reillo W, Rossmassler K, Sackett J, Salekdeh GH, Saleska S, Scarborough M, Schachtman D, Schadt CW, Schrenk M, Sczyrba A, Sengupta A, Setubal JC, Shade A, Sharp C, Sherman DH, Shubenkova OV, Sierra-Garcia IN, Simister R, Simon H, Sjöling S, Slonczewski J, Correa De Souza RS, Spear JR, Stegen JC, Stepanauskas R, Stewart F, Suen G, Sullivan M, Sumner D, Swan BK, Swingle W, Tarn J, Taylor GT, Teeling H, Tekere M, Teske A, Thomas T, Thrash C, Tiedje J, Ting CS, Tully B, Tyson G, Ulloa O, Valentine DL, Van Goethem MW, VanderGheynst J, Verbeke TJ, Vollmers J, Vuillemin A, Waldo NB, Walsh DA, Weimer BC, Whitman T, Van Der Wielen P, Wilkins M, Williams TJ, Woodcroft B, Wooley J, Wrighton K, Ye J, Young EB, Youssef NH, Yu FB, Zemskaya TI, Ziels R, Woyke T, Mouncey NJ, Ivanova NN, Kyrpides NC, Eloe-Fadrosh EA. A genomic catalog of Earth's microbiomes. *Nat Biotechnol*. 2021;39:499–509. <https://doi.org/10.1038/s41587-020-0718-6>.
58. Whitworth DE, Sydney N, Radford EJ. Myxobacterial genomics and post-genomics: a review of genome biology, genome sequences and related 'omics studies. *Microorganisms*. 2021;9:2143. <https://doi.org/10.3390/microorganisms9102143>.
59. Nan B, Bandaria JN, Moghtaderi A, Sun I-H, Yildiz A, Zusman DR. Flagella stator homologs function as motors for myxobacterial gliding motility by moving in helical trajectories. *Proc Natl Acad Sci*. 2013. <https://doi.org/10.1073/pnas.1219982110>.
60. Thiery S, Kaimer C. The predation strategy of *Myxococcus xanthus*. *Front Microbiol*. 2020;11:2. <https://doi.org/10.3389/fmicb.2020.00002>.
61. Muñoz-Dorado J, Marcos-Torres FJ, García-Bravo E, Moraleda-Muñoz A, Pérez J. Myxobacteria: moving, killing, feeding, and surviving together. *Front Microbiol*. 2016. <https://doi.org/10.3389/fmicb.2016.00781>.
62. López-Mondéjar R, Tláskal V, Da Rocha UN, Baldrian P. Global distribution of carbohydrate utilization potential in the prokaryotic tree of life. *mSystems*. 2022;7:e00829-22. <https://doi.org/10.1128/mSystems.00829-22>.
63. Wilhelm RC, Singh R, Eltis LD, Mohn WW. Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing. *ISME J*. 2019;13:413–29. <https://doi.org/10.1038/s41396-018-0279-6>.
64. Richy E, Cabello-Yeves PJ, Hernandez-Coutinho F, Rodriguez-Valera F, González-Álvarez I, Gandois L, Rigal F, Lauga B. How microbial communities shape peatland carbon dynamics: new insights and implications. *Soil Biol Biochem*. 2024;191: 109345. <https://doi.org/10.1016/j.soilbio.2024.109345>.
65. Choi BY, Lee S, Kim J, Park H, Kim J-H, Kim M, Park S-J, Kim K-T, Ryu H, Shim D. Comparison of endophytic and epiphytic microbial communities in surviving and dead Korean fir (*Abies koreana*) using metagenomic sequencing. *Forests*. 2022;13:1932. <https://doi.org/10.3390/f13111932>.
66. Kuroda K, Yamamoto K, Nakai R, Hirakata Y, Kubota K, Nobu MK, Narihiro T. Symbiosis between *Candidatus* patescibacteria and archaea discovered in wastewater-treating bioreactors. *MBio*. 2022;13:e01711-e1722. <https://doi.org/10.1128/mbio.01711-22>.
67. Wang Y, Gallagher LA, Andrade PA, Liu A, Humphreys IR, Turkarslan S, Cutler KJ, Arrieta-Ortiz ML, Li Y, Radey MC, McLean JS, Cong Q, Baker D, Baliga NS, Peterson SB, Mougous JD. Genetic manipulation of Patescibacteria provides mechanistic insights into microbial dark matter and the epibiotic lifestyle. *Cell*. 2023;186:4803–4817.e13. <https://doi.org/10.1016/j.cell.2023.08.017>.
68. López-Mondéjar R, Brabcová V, Štursová M, Davidová A, Jansa J, Cajthaml T, Baldrian P. Decomposer food web in a deciduous forest shows high share of generalist microorganisms and importance of microbial biomass recycling. *ISME J*. 2018;12:1768–78. <https://doi.org/10.1038/s41396-018-0084-2>.
69. Paul EA, Frey SD, editors. *Soil microbiology, ecology, and biochemistry*. 5th ed. Amsterdam: Elsevier; 2024.
70. Cherkasov N, Ibhaden AO, Fitzpatrick P. A review of the existing and alternative methods for greener nitrogen fixation. *Chem Eng Process Process Intensif*. 2015;90:24–33. <https://doi.org/10.1016/j.cep.2015.02.004>.
71. Burris RH, Roberts GP. Biological nitrogen fixation. *Annu Rev Nutr*. 1993;13:317–35. <https://doi.org/10.1146/annurev.nu.13.070193.001533>.
72. Davies-Barnard T, Friedlingstein P. The global distribution of biological nitrogen fixation in terrestrial natural ecosystems. *Glob Biogeochem Cycles*. 2020;34:e2019GB006387. <https://doi.org/10.1029/2019GB006387>.
73. Dos Santos PC, Fang Z, Mason SW, Setubal JC, Dixon R. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics*. 2012;13:162. <https://doi.org/10.1186/1471-2164-13-162>.
74. Zehr JP, Capone DG. Changing perspectives in marine nitrogen fixation. *Science*. 2020;368:eaay9514. <https://doi.org/10.1126/science.aay9514>.
75. Alleman AB, Garcia Costas A, Mus F, Peters JW. Rnf and fix have specific roles during aerobic nitrogen fixation in *Azotobacter vinelandii*. *Appl Environ Microbiol*. 2022;88:e01049-e1122. <https://doi.org/10.1128/aem.01049-22>.
76. Johnson DC, Unciuleac M-C, Dean DR. Controlled expression and functional analysis of iron-sulfur cluster biosynthetic components within *Azotobacter vinelandii*. *J Bacteriol*. 2006;188:7551–61. <https://doi.org/10.1128/JB.00596-06>.
77. Tokumoto U. Interchangeability and Distinct properties of bacterial Fe-S cluster assembly systems: functional replacement of the ISC and SUF operons in *Escherichia coli* with the nifSU-like operon from *Helicobacter pylori*. *J Biochem (Tokyo)*. 2004;136:199–209. <https://doi.org/10.1093/jb/mvh104>.
78. Sharrar AM, Crits-Christoph A, Méheust R, Diamond S, Starr EP, Banfield JF. Bacterial secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation type. *MBio*. 2020;11:e00416-e420. <https://doi.org/10.1128/mbio.00416-20>.
79. Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature*. 2018;558:440–4. <https://doi.org/10.1038/s41586-018-0207-y>.
80. Waschulin V, Borsetto C, James R, Newsham KK, Donadio S, Corre C, Wellington E. Biosynthetic potential of uncultured Antarctic soil bacteria revealed through long-read metagenomic sequencing. *ISME J*. 2022;16:101–11. <https://doi.org/10.1038/s41396-021-01052-3>.
81. Brinkmann S, Kurz M, Patras MA, Hartwig C, Marner M, Leis B, Billion A, Kleiner Y, Bauer A, Toti L, Pöverslein C, Hammann PE, Vilcinskas A, Glaeser J, Spohn M, Schäberle TF. Genomic and chemical decryption of the bacteroidetes phylum for its potential to biosynthesize natural products.

- Microbiol Spectr. 2022;10:e02479–e2521. <https://doi.org/10.1128/spectrum.02479-21>.
82. Uliano-Silva M, Ferreira JGRN, Krashennikova K, Darwin Tree of Life Consortium, Blaxter M, Mieszkowska N, Hall N, Holland P, Durbin R, Richards T, Kersey P, Hollingsworth P, Wilson W, Twyford A, Gaya E, Lawniczak M, Lewis O, Broad G, Martin F, Hart M, Barnes I, Formenti G, Abueg L, Torrance J, Myers EW, Durbin R, Blaxter M, McCarthy SA. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics*. 2023;24:288. <https://doi.org/10.1186/s12859-023-05385-y>.
 83. Galagan JE, Henn MR, Ma L-J, Cuomo CA, Birren B. Genomics of the fungal kingdom: insights into eukaryotic biology. *Genome Res*. 2005;15:1620–31. <https://doi.org/10.1101/gr.3767105>.
 84. Saraiva JP, Bartholomäus A, Toscan RB, Baldrian P, Nunes da Rocha U. Recovery of 197 eukaryotic bins reveals major challenges for eukaryote genome reconstruction from terrestrial metagenomes. *Mol Ecol Resour*. 2023;23:1066–76. <https://doi.org/10.1111/1755-0998.13776>.
 85. Ma B, Lu C, Wang Y, Yu J, Zhao K, Xue R, Ren H, Lv X, Pan R, Zhang J, Zhu Y, Xu J. A genomic catalogue of soil microbiomes boosts mining of biodiversity and genetic resources. *Nat Commun*. 2023;14:7318. <https://doi.org/10.1038/s41467-023-43000-z>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.