ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Genomic evidence for the widespread presence of GH45 cellulases among soil invertebrates

Hannah Muelbaier[1,2] | Freya Arthen[1,2] | Gemma Collins[2,3] | Thomas Hickler[4,5] |
Karin Hohberg[6] | Ricarda Lehmitz[2,6] | Yannick Pauchet[7] | Markus Pfenninger[2,4,8] |
Anton Potapov[9,10,11] | Juliane Romahn[2,4,12] | Ina Schaefer[2,4,13] | Stefan Scheu[14] |
Clément Schneider[2,6] | Ingo Ebersberger[1,2,4] | Miklós Bálint[2,4,12]

[1]Applied Bioinformatics Group, Inst. of Cell Biology and Neuroscience, Goethe University, Frankfurt am Main, Germany

[2]LOEWE Centre for Translational Biodiversity Genomics, Frankfurt am Main, Germany

[3]Manaaki Whenua – Landcare Research, Auckland, New Zealand

[4]Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany

[5]Department of Physical Geography, Goethe University, Frankfurt/Main, Germany

[6]Senckenberg Museum of Natural History Görlitz, Görlitz, Germany

[7]Insect Symbiosis, Max Planck Institute for Chemical Ecology, Jena, Germany

[8]Institute for Molecular and Organismic Evolution, Johannes Gutenberg University, Mainz, Germany

[9]Senckenberg Museum for Natural History Görlitz, Görlitz, Germany

[10]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

[11]International Institute Zittau, TUD Dresden University of Technology, Zittau, Germany

[12]Institute of Insect Biotechnology, Justus-Liebig University, Giessen, Germany

[13]Animal Ecology, University of Goettingen, Goettingen, Germany

[14]J.F. Blumenbach Institute of Zoology and Anthropology, University of Goettingen, Goettingen, Germany

Correspondence
Miklós Bálint and Ingo Ebersberger,
LOEWE Centre for Translational
Biodiversity Genomics,
Senckenberganlage 25, 60325 Frankfurt
am Main, Germany.
Email: miklos.balint@agrar.uni-giessen.de
and ebersberger@bio.uni-frankfurt.de

Abstract

Lignocellulose is a major component of vascular plant biomass. Its decomposition is crucial for the terrestrial carbon cycle. Microorganisms are considered primary decomposers, but evidence increases that some invertebrates may also decompose lignocellulose. We investigated the taxonomic distribution and evolutionary origins of GH45 hydrolases, important enzymes for the decomposition of cellulose and hemicellulose, in a collection of soil invertebrate genomes. We found that these genes are common in springtails and oribatid mites. Phylogenetic analysis revealed that cellulase genes were acquired early in the evolutionary history of these groups. Domain architectures and predicted 3D enzyme structures indicate that these cellulases are functional. Patterns of presence and absence of these genes across different lineages prompt further investigation into their evolutionary and ecological benefits. The ubiquity of cellulase genes suggests that soil invertebrates may play a role in lignocellulose

Ingo Ebersberger and Miklós Bálint contributed equally to this work.

decomposition, independently or in synergy with microorganisms. Understanding the ecological and evolutionary implications might be crucial for understanding soil food webs and the carbon cycle.

KEYWORDS
arthropods, comparative genomics, decomposition, global change, horizontal gene transfer

1 | INTRODUCTION

Most photosynthetically bound carbon on land ends up in woody plants as lignocellulose, a composite of cellulose, hemicelluloses, lignin and pectin. The decomposition of lignocellulose occurs predominantly in soils, which returns most of this carbon into the atmosphere (Post et al., 1990). Terrestrial ecosystems currently sequester about 29% of the anthropogenic carbon emissions, which implies an important but not fully understood role of terrestrial carbon cycling for climate regulation (Cragg et al., 2015). Microorganisms, especially bacteria and fungi, encode glycoside hydrolase cocktails for lignocellulose degradation in their genomes (Cragg et al., 2015), and are considered the main actors of decomposition (Bradford et al., 2017; Crowther et al., 2019; Pausas & Bond, 2020). The contribution of animals to decomposition of lignocellulose—beyond purely mechanical shredding—remains less understood. Experiments have shown that the presence of soil invertebrates can increase litter mass loss by up to 50% (García-Palacios et al., 2013). It is estimated that they decompose more deadwood in tropical forests than free-living microorganisms (Griffiths et al., 2019). Nevertheless, the mechanisms behind decomposition performed by soil invertebrates remain obscure and the ability of soil animals to degrade composite polysaccharides without relying on gut symbionts remains a long-standing debate in soil ecology (Berg et al., 2004; Cragg et al., 2015).

It was originally assumed that lignocellulose degradation by animals was entirely 'outsourced' to the gut microbiome (Briones, 2018; García-Palacios et al., 2013). However, evidence is emerging that at least some invertebrates, such as molluscs, crustaceans and phytophagous insects, can synthesize cellulase enzymes themselves (Busch et al., 2019; Chang & Lai, 2018; Cragg et al., 2015; Griffiths et al., 2021; Han et al., 2022; Kern et al., 2013; King et al., 2010; Shelomi et al., 2014; Watanabe et al., 1998). Scattered evidence also exists for the expression of active endogenous cellulases by distantly related soil invertebrates, for example, the earthworm *Pheretima hilgendorfi* (Nozaki et al., 2009), the Antarctic springtail *Cryptopygus antarcticus* (Hong et al., 2014), as well as few oribatid mites and other springtails (Busch et al., 2019). Based on these individual findings, we hypothesize that a larger fraction of soil invertebrates than previously thought may be directly contributing to decomposition of lignocellulose in dead plant matter in soils using endogenous cellulases, with or without relying on a microbiome. Given their global abundance and diversity in many soil ecosystems (FAO et al., 2020; Phillips et al., 2019; Potapov et al., 2023; van den Hoogen et al., 2019), soil invertebrates could, therefore, have an important but so far overlooked role in the

terrestrial carbon cycle which is distinct from the decomposition ability of microorganisms. To evaluate whether endogenous decomposition ability is a common feature shared by the main groups of soil invertebrates, we screened a diverse set of newly sequenced genomes of Collembola, Enchytraeidae, Gamasina, Myriapoda, Nematoda, Oribatida and Tardigrada for the presence and evolutionary origins of cellulase genes. We further investigated the hypothesis that cellulase genes in invertebrates were acquired by horizontal gene transfer (HGT) from microorganisms. Cellulases are classified into three different categories: Endoglucanases (EC 3.2.1.4) break down cellulose in smaller randomly sized fragments. Exoglucanases (EC 3.2.1.91) release cellobiose and cellotriose from the ends of cellulose molecules. Cellobiases (EC 3.2.1.21) convert cellobiose into glucose. Common to all cellulases is their activity to hydrolyse glycoside bonds, and the CAZy database currently lists 187 different families of glycoside hydrolases (Drula et al., 2022). Here, we focus on GH45 cellulases because these were already described in hexapods, and they were found to be functional (*Cryptopygus antarcticus* and *Phaedon cochleariae* (Girard and Jouanin 1999; Song et al. 2008; Busch et al., 2019)).

2 | MATERIALS AND METHODS

2.1 | Domain architectures of invertebrate GH45-type cellulases

Reviewed evidence exists for the presence of GH45-type cellulases in the Antarctic springtail (*Cryptopygus antarcticus*; Collembola, UniprotID D3GDK4) and the mustard beetle (*Phaedon cochleariae*; Insecta, UniprotID O97401). Since these experimentally confirmed cellulases harbour a Glyco-hydro 45 Pfam domain (PF02015) (Bankevich et al., 2012), we restricted our analysis to cellulase orthologues that carry this Pfam domain. Pfam domains were annotated with hmmscan from the HMMER package (Finn et al., 2015) using Pfam version 32 and applying the default e-value cut-off of 0.01.

2.2 | Genome assembly pipeline

The genome assemblies provided by the MetaInvert Project (Bioproject ID: PRJNA758215) cover a phylogenetically diverse set of soil-living invertebrates collected from the field or obtained from cultures. Short read Illumina sequencing (300 bp paired-end) with the NovaSeq 6000 platform was done at Novogene Europe

(Cambridge, UK), reads were trimmed with Trimmomatic, human contaminating reads were filtered with Kraken2 (Wood et al., 2019) and contig assembly was done with SPAdes (Bankevich et al., 2012). The resulting contigs were taxonomically assigned with Blobtools2 (Challis et al., 2020) using the NCBI non-redundant protein database as a reference, and only contigs with an assignment to the phylum of the target species together with unassigned contigs were kept. Redundancy reduction and scaffolding were done with Redundans (Pryszcz & Gabaldón, 2016), and genome assembly completeness was assessed with BUSCO (v 4.1.4) using the precomputed metazoan (obd10) reference set. The genome data and metadata were published and further analysed by Collins et al. (2023) (in press: 10.1038/s42003-023-05621-4). Collection (geographic coordinates) and vouchering data of the used specimens are provided by Collins et al. (2023). For our orthologue search, we selected genomes with at least 50% BUSCO completeness (176 assemblies, Table S2).

2.3 | RefSeq gene set collection

We downloaded gene sets for all 18,412 taxa represented in the NCBI RefSeq Genome release 207 (O'Leary et al., 2016). The resulting taxon collections comprised 16,401 bacteria, 910 archaea, 409 fungi, 262 invertebrates and 430 vertebrates. The taxon list together with the accession numbers is provided in Table S1.

2.4 | Taxonomic assignment and contaminant detection

To rule out that fungal or bacterial contaminations of the underlying genome assemblies are responsible for the animal cellulase orthologues (Steinegger & Salzberg, 2020), we followed a multi-step procedure. First, during the invertebrate genome assembly (described in Collins et al., 2023, not part of this paper), all contigs were used as input for a Blobtools2 (Challis et al., 2020) analysis using the NCBI non-redundant protein and nucleotide databases as a reference. Only contigs assigned to the phylum of the target species and unassigned contigs were passed on to the downstream analyses. Second, we taxonomically classified each of the detected invertebrate orthologues. In brief, we used the orthologue sequence as a query for a Diamond version 2.0.13 (Buchfink et al., 2015) search against the NCBI non-redundant protein database (downloaded January 2022). From the resulting hit list, we excluded the trivial hit against itself and then assigned the query sequence to the last common ancestor of the taxa within a 10% bit score margin of the best hit (Huson et al., 2016). A sequence was flagged as a putative contaminant if its taxonomic assignment was not placed on the lineage from the species whose genome was analysed to the root of the tree of cellular life. The workflow is implemented into the software package taXaminer (https://github.com/BIONF/taxaminer), and it provided no evidence for a foreign origin of these sequences (Table S3). As a showcase example, we further investigated the origin of two GH45

cellulase orthologues. In the mite *Carabodes femoralis*, the orthologue is located on a short contig with no neighbouring genes. We taxonomically assigned this gene to the Acariformes. In the springtail *Pogonognathellus longicornis*, the GH45 cellulase is located together with other genes on a contig of 41 kbp in length. The cellulase and neighbouring genes on the same contig were consistently assigned to Entomobryomorpha as a result (Figure S5).

2.5 | Orthology-based phylogenetic profiles of fungal GH45 cellulase

Profile-based targeted orthologue searches in annotated gene sets were performed with fDOG (Birikmen et al., 2021) using the GH45 cellulase of the fungus *Rhizoctonia solani* (NCBI Accession XP_043186467.1) as the seed. For the training of the initial profile Hidden Markov model, we used the parameter --*minDist genus* and --*maxDist phylum* limiting the number of training sequences to 6 (see Table S4 for more information). Candidate orthologues were filtered for the presence of the Pfam glyco-hydro 45 domain (PF02015; see Table S5 for a list of discarded orthologues). Orthologue search in the unannotated MetaInvert genome assemblies was performed with the fDOG extension fDOG-Assembly. In brief, genomic regions likely containing a GH45-type cellulase were identified with a tBLASTn search using the consensus sequence included in the initial core gh45 core group from fDOG as query. The hit region was extended by 500 nucleotides on either side and genes in the resulting candidate genomic region were annotated with MetaEuk v5.34c21f2 (Levy Karin et al., 2020) using the OMA database (released in December 2021; Nguyen et al., 2015) as the reference database for the gene prediction. The corresponding protein sequences were then tested for orthology using the routines of fDOG and afterwards, features were annotated with FAS (Dosch et al., 2023). The fDOG-assembly workflow is available from https://github.com/BIONF/fDOG/tree/fdog_goes_assembly. We completed the taxon collection by adding four individual genomes publicly available on GenBank (Table S6) and performed the same orthologue searches described above. The results from fDOG and fDOG-Assembly were merged and visualized with PhyloProfile (Tran et al., 2018).

2.6 | GH45 cellulase gene tree reconstruction

To investigate the evolutionary history of the GH45 cellulases, we used the identified orthologues for a gene tree reconstruction. If the orthologue search obtained more than one co-orthologue, we used the one that is most similar to the seed protein from *R. solani* for the tree reconstruction. Sequences were aligned with Muscle v3.8.1551 (Edgar, 2004) and alignment columns comprising more than 50% gaps were removed with a custom perl script. The resulting multiple sequence alignment was used as input for a maximum likelihood tree reconstruction with IQ-TREE (Nguyen et al., 2015) version 1.6.8. Branch support was assessed with 1000 bootstrap replicates using

the ultrafast bootstrap approach. The SH-aLRT branch test was performed, and the optimal number of cores was automatically detected via IQ-TREE (− nt AUTO). All gene trees were visualized with iTOL (Letunic & Bork, 2021). Animals are paraphyletic in this tree, and a topology test using the AU test (Shimodaira & Hasegawa, 1999) confirmed that a tree with monophyletic animal cellulases explained the data significantly worse (p-AU = 9.2E−4). A second gene tree containing all identified orthologues and co-orthologues was reconstructed with the same approach. The two tree files in Newick format are available as Supplement Files F1 and F2.

2.7 | Phylogenies of springtails and oribatids

Phylogenies of springtail and oribatid organisms included in the MetaInvert project were computed separately with a supermatrix approach. BUSCO version 5.4.2 (Simão et al., 2015) with the precomputed BUSCO Arthropoda gene set (db10) was used to search for orthologues in all genome assemblies. BUSCO genes were discarded from the analysis which were present in less than 75% of all species. Multiple sequence alignments were computed with MAFFT using local pairwise alignment and at maximum 1000 iterations (7.481) (Katoh & Standley, 2013), trimmed with clipkit (1.3.0) (Steenwyk et al., 2020) and concatenated with FASconCAT-G (1.04) (Kück & Longo, 2014). Four phylogenetic trees per taxon group were reconstructed with IQ-TREE (Nguyen et al., 2015). Branch support was assessed with 1000 bootstrap replicates using the ultrafast bootstrap approach. The best-fit model was Q.insect + F + R9 for oribatids, and Q.insect + F + R10 for springtails, automatically chosen by ModelFinder according to BIC. The final consensus tree was computed with splitstree (4.19.0) (Huson & Bryant, 2006) by summarizing the four IQ-TREEs into a consensus tree. The consensus trees were outgroup-rooted using *Sarcoptes scabiei* (GCA_020844145.1) for oribatids, and *Machilis hrabei* (GCA_003456935.1), *Drosophila albomicans* (GCA_009650485.1) and *Tyrophagus putrescentiae* for springtails as outgroups. The species trees and reproductive mode of the corresponding species (taken from Collins et al. 2023) were visualized with iTOL. Chi-squared tests were performed with R.

2.8 | Correspondence of GH45 tree and phylogenies

The BUSCO-based phylogenies of springtails and oribatids were outgroup-rooted and merged for the comparison with the GH45 cellulase gene tree. The tanglegram matches taxa by links and was computed with R using the packages phytools v1.4-0 (Revell, 2012) and castor v1.7.6 (Louca & Doebeli, 2018).
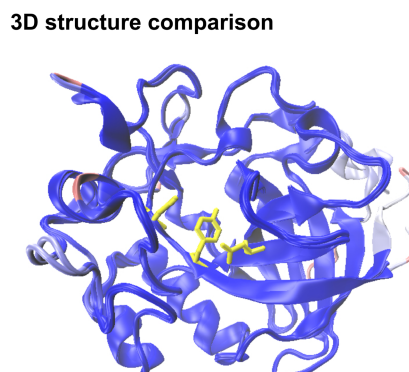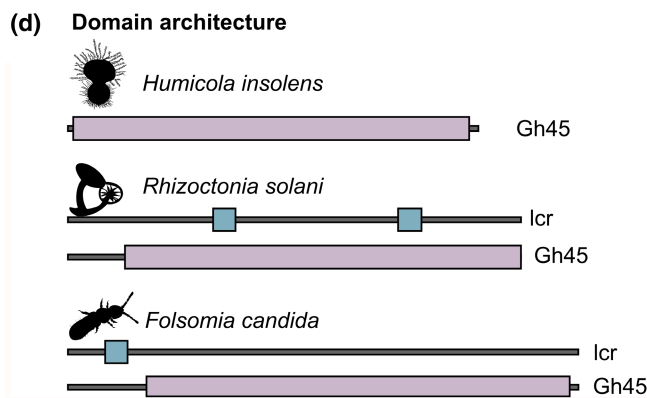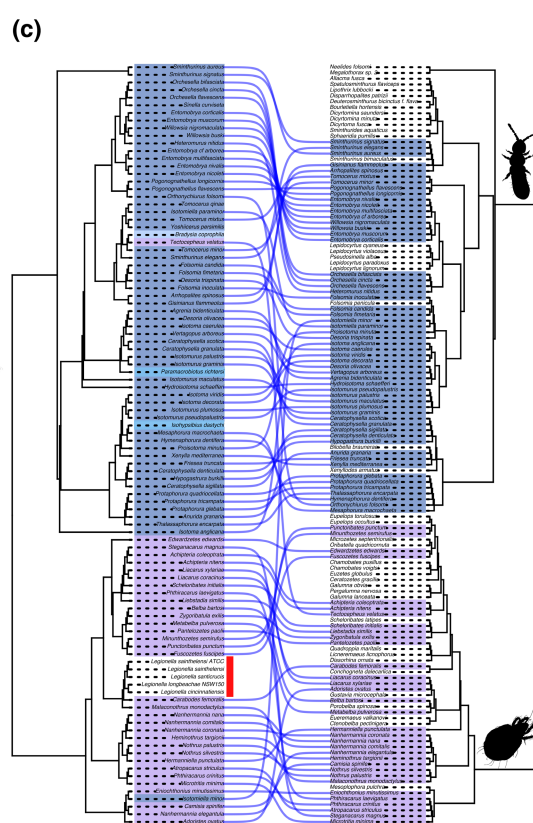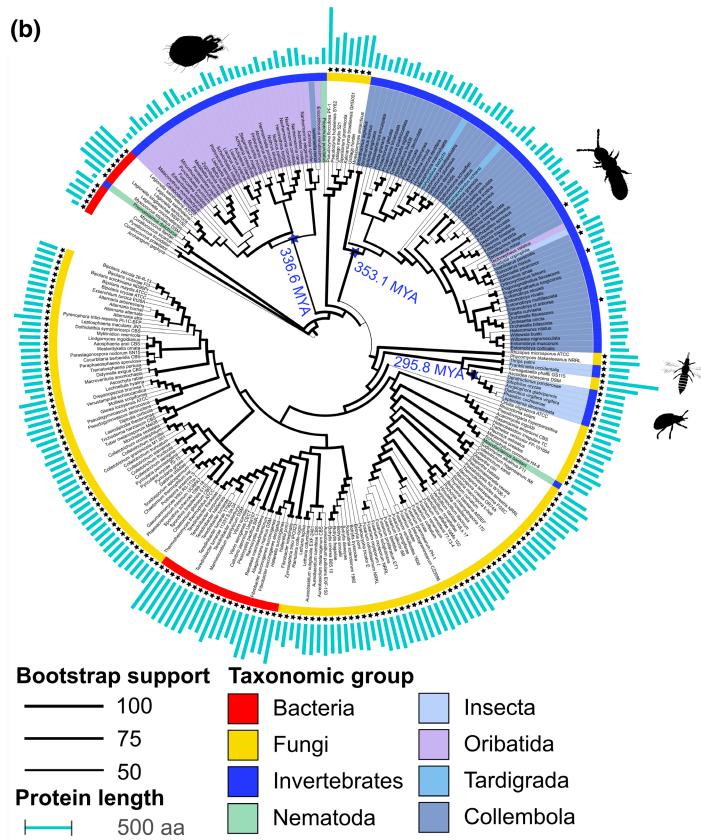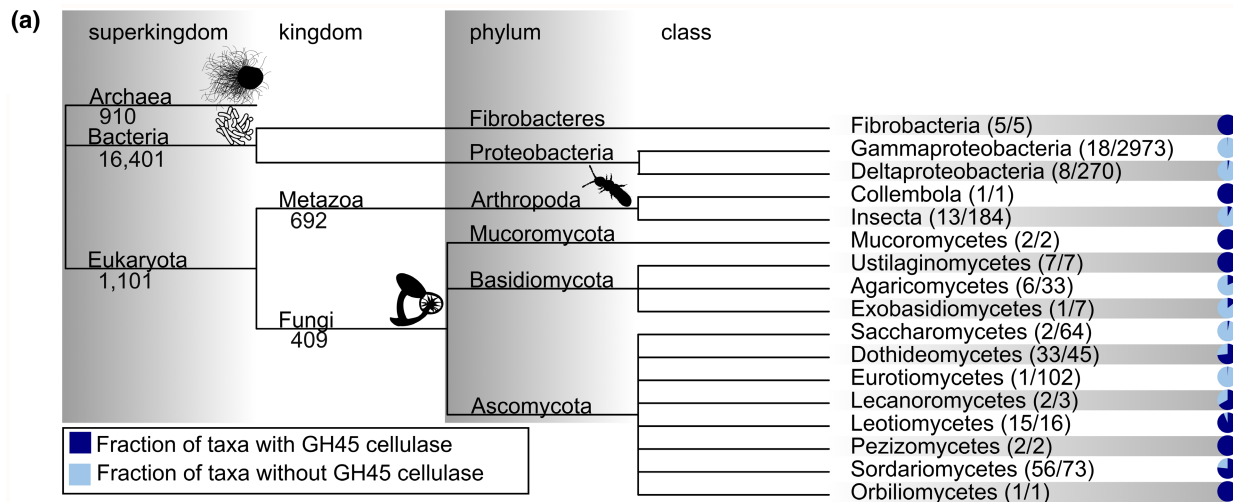
2.9 | 3D structure comparison

3D structures of *R. solani* GH45 cellulase and of the endoglucanase V from *Humicola insolens* were retrieved from precomputed predictions from UniProt (accession number A0A0B7FQX1 and P43316). The 3D structure of the *F. candida* GH45 cellulase was locally computed with AlphaFold (Jumper et al., 2021). The structures were visualized and compared with VMD (Humphrey et al., 1996) and the extensions MultiSeq (Roberts et al., 2006) in combination with the alignment tool STAMP (Russell & Barton, 1992). The active site was identified by highlighting in yellow the catalytic residues found by Davies et al., 1993. A structure-guided sequence alignment of all three GH45 sequences was computed with expresso from the t-coffee package (Armougom et al., 2006) (Figure S4).

3 | RESULTS

3.1 | Endogenous GH45 cellulases are mainly found in fungi and invertebrates

We used a fungal sequence as a starting point to obtain a comprehensive overview of the taxonomic distribution of the GH45 cellulase family. In an initial screen, we considered all species that are represented by a genome assembly in NCBI RefSeq. We decided to base this analysis on fungal GH45 cellulase because (i) existing results suggest that invertebrate GH45 cellulases were likely horizontally transferred from fungi (Busch et al., 2019) and (ii) fungal GH45

FIGURE 1 GH45 cellulases on the tree of life. (a) Abundance of GH45 cellulases across the three domains of life; (b) Maximum likelihood phylogeny of the GH45 cellulase family. Branch lengths are not drawn to scale and line weights indicate percent bootstrap support. Species represented by a genome assembly in the NCBI RefSeq or GenBank databases are indicated by an asterisk. Pictograms identify the four main soil invertebrate clades: oribatid mites, springtails, thrips and beetles (clockwise). Bars indicate the protein length of the respective GH45 cellulase in amino acids. Internal node labels provide age estimates of the respective clades (Kumar et al., 2022); (c) Tanglegram between the GH45 gene tree (left) and a phylogenomic reconstruction of the species tree (right). Only cellulases identified in oribatids and springtails are considered. Lines connect the GH45 cellulases in the gene tree with the corresponding taxa they were identified in the species tree; (d) Left: Comparison of protein domain architecture of two fungal (*H. insolens* (Uniprot ID: P43316), *R. solani* (Uniprot ID: A0A0B7FQX1)) and one invertebrate GH45 cellulase (*F. candida* (NCBI Acc: XP_021945337.1)). lcr—low complexity region; GH45—Pfam Glyco-hydro 45 domain (PF02015). Right: 3D structure alignment of the same three GH45 cellulases. Thirty amino acids from the N termini of the *R. solani* and *F. candida* proteins are not considered in the structural comparison because the corresponding alpha helix could not be confidently placed in the structure by AlphaFold (per-residue confidence score (pLDDT) <70; (Tunyasuvunakool et al., 2021)). Structural similarity is colour coded and ranges from red (very low) to very high (blue). The catalytic residues located in the active groove of the *H. insolens* cellulase (Davies et al., 1993) are highlighted in yellow. A structure-guided amino acid sequence alignment is shown in Figure S4 and reveals the conservation of these catalytic residuals in the other two proteins.

(a)

superkingdom   kingdom   phylum   class

Archaea 910
Bacteria 16,401

Fibrobacteres — Fibrobacteria (5/5)
Proteobacteria — Gammaproteobacteria (18/2973)
— Deltaproteobacteria (8/270)

Eukaryota 1,101

Metazoa 692
Arthropoda — Collembola (1/1)
— Insecta (13/184)

Fungi 409
Mucoromycota — Mucoromycetes (2/2)
Basidiomycota — Ustilaginomycetes (7/7)
— Agaricomycetes (6/33)
— Exobasidiomycetes (1/7)
Ascomycota — Saccharomycetes (2/64)
— Dothideomycetes (33/45)
— Eurotiomycetes (1/102)
— Lecanoromycetes (2/3)
— Leotiomycetes (15/16)
— Pezizomycetes (2/2)
— Sordariomycetes (56/73)
— Orbiliomycetes (1/1)

Fraction of taxa with GH45 cellulase
Fraction of taxa without GH45 cellulase

(b)

353.1 MYA
336.6 MYA
295.8 MYA

Bootstrap support
— 100
— 75
— 50

Protein length
├─┤ 500 aa

Taxonomic group
- Bacteria
- Fungi
- Invertebrates
- Nematoda
- Insecta
- Oribatida
- Tardigrada
- Collembola

(c)

(d) Domain architecture

Humicola insolens
Gh45

Rhizoctonia solani
lcr
Gh45

Folsomia candida
lcr
Gh45

3D structure comparison

cellulases have been thoroughly functionally characterized (Berto et al., 2019; Bharadwaj et al., 2020). We searched for orthologues to fungal GH45 cellulases in 16,401 bacteria, 910 archaea and 1101 eukaryotes (Table S1). The resulting phylogenetic profile revealed that GH45 cellulases are abundant only in fungi, where 128 out of 409 investigated taxa possess at least one GH45 cellulase gene. However, they are limited to individual taxonomic groups (Figure 1a). In contrast, only 31 out of 16,401 bacterial species were found to harbour a GH45 cellulase. Next to fungi, orthologues to GH45 cellulases were largely confined to animals (14 out of 692) (Figure 1a). We found no evidence that these comprise fungal contaminations (see taxonomic assignment and contaminant detection in Section 4). In metazoans, the majority of orthologues were found in arthropods, and they were completely absent in vertebrates.

3.2 | Most springtails and oribatid mites possess endogenous GH45 cellulases

The analysis of publicly available genomes sheds light on the general distribution of GH45 cellulases across the Tree of Life. However, it lacks the resolution to investigate the occurrence of this cellulase in soil invertebrates. We extended the analysis by including novel genome assemblies for an additional 176 species representing a diverse selection of soil invertebrates (NCBI BioProject PRJNA758215) into the analysis. This revealed a high occurrence of GH45 cellulases in springtails (56/78 analysed species) and in oribatid mites (33/54 analysed species, Table S2). Additionally, we detected cellulases in Coleoptera and Thysanoptera (2/2 species, Table S2). In three out of nine nematode species we also found GH45 cellulases, whereas none were found in 30 representatives of Chilopoda and Diplopoda (Table S2).

GH45 cellulases are present in three of four known main springtail lineages (Poduromorpha, Entomobryomorpha and Symphypleona), missing only from the earliest branching Neelipleona. Within Symphypleona and Entomobryomorpha, cellulases are consistently absent in one clade each (Figure S1). GH45 cellulases are present in almost all representatives of the basal clades in Oribatida (Enarthronota, Mixonomata, Holosomata; Table S2). In contrast, they were missing in half of the investigated species from the later-branching Brachypylina (Figure S2). Integrating the presence–absence pattern of GH45 cellulases with the reproduction mode of the respective species (sexually vs. parthenogenetic) reveals no significant correlation for Oribatida and Collembola respectively (χ^2 test, $p > .05$). Besides Oribatida, we investigated a second mite taxon, the Gamasina, which are represented by two species in our data sets (Table S2). We found in neither species a GH45 cellulase.

3.3 | The evolutionary history of invertebrate GH45 cellulases is complex

We reconstructed the phylogenetic relationships of the sequences identified both in RefSeq assemblies and in soil invertebrate genomes

(Figure 1b) to better understand the evolutionary trajectory resulting in the present-day distribution of soil invertebrate GH45 cellulases. In the resulting tree, 104 out of 107 animal proteins are placed in only four distinct and taxonomically largely homogenous clades, one each for the Collembola, Oribatida, Thysanoptera and Coleoptera. On a larger scale, we found that the invertebrate cellulase clades are embedded into an evolutionary background formed by mostly fungal and very few bacterial sequences. This leaves metazoans as a taxonomic group paraphyletic with respect to their cellulases and indicates that GH45 cellulases have been independently introduced into the animal kingdom at least four times, very likely from fungal donors. We noted that several bacterial cellulases, all from the genus *Legionella*, form a monophyletic clade placed within the diversity of the oribatid cellulases. This finding suggests a horizontal transfer of an animal gene into a bacterial clade. However, the corresponding branch is very long (1.78 substitutions per site; SupplData) and further investigations will be necessary to consolidate this hypothesis. To assess whether the evolutionary relationships of the invertebrate cellulases agree with those of the species they were found in, we compared the phylogeny from oribatid mites and springtails with the reconstructed cellulase gene tree (Figure 1c). The two trees display in large parts agreeing branching patterns, which is expected for intrinsic cellulases. However, they clearly differ in other parts. To trace down the reason for these differences, we reconstructed the invertebrate cellulase tree this time considering all detected GH45 co-orthologues. This revealed a highly dynamic and complex evolutionary history of the invertebrate GH45 cellulase family involving numerous lineage-specific gene duplications and losses (Figure S3).

3.4 | Fungal and invertebrate GH45 cellulases are structurally conserved

In a last analysis, we investigated whether the detected invertebrate GH45 cellulases are likely functional. We compared the domain architectures and the predicted 3D protein structures between GH45 cellulases of the fungus *Rhizoctonia solani* and the springtail *Folsomia candida* to that of the fungus *Humicola insolens*, whose GH45 cellulase has been functionally and structurally characterized before (Davies et al., 1993). All three enzymes agree in their domain architecture and have highly similar 3D structures (Figure 1d). Moreover, we found the three catalytic residues in the substrate-binding active grove initially described by Davies et al. (1993) to be conserved across the proteins (Figure S4). Together, this strongly indicates that all three proteins can hydrolyse glycoside bonds.

4 | DISCUSSION

Most cellulases discovered to date in metazoan genomes belong to the GH45 family (Busch et al., 2019), endo-β-1,4-glucanases which hydrolyse cellulose, lichenin and cereal β-D-glucans (EC 3.2.1.4). Individual members of this family appear specific for degrading

xyloglucan (EC 3.2.1.151). However, Busch et al. (2019) have shown that the tested GH45 cellulases show no substrate specificity and break down both cellulose and xyloglucan. It was hypothesized that invertebrate GH45 cellulases were repeatedly obtained via HGT from fungal donors (Busch et al., 2019). Our search in 18,412 RefSeq genomes revealed that only 31 bacterial species harbour a GH45-type cellulase. This is in line with the earlier hypothesis about the evolutionary roots of this enzyme within fungi (Busch et al., 2019; Palomares-Rius et al., 2014). The taxonomic distribution of GH45 cellulases was non-uniform in metazoans, with most orthologues being found in arthropods. GH45 cellulases were reported in marine bivalves (Okmane et al., 2022) which feed on phytoplankton and on particulate matter. We could not evaluate these findings, as genomes of the respective species are not yet included in the NCBI RefSeq database. Our results confirm previous research which did not find evidence for GH45 presence in vertebrates (Chang & Lai, 2018). The screening of new soil invertebrate genomes uncovered novel GH45 cellulase presence patterns. GH45 genes were found in well over half of the investigated springtail and oribatid mite species.

Springtails form a basal hexapod group abundant across the globe, especially in cold regions (Potapov et al., 2023). They are known as fungal feeders, but also consume detritus and fresh plant materials (Potapov et al., 2020). While the orthologue search showed the presence of GH45 cellulase in springtail lineages Poduromorpha, Entomobryomorpha and Symphypleona, it is missing from the earliest branching Neelipleona. Neelipleona feed on detritus colonized by fungi and bacteria rather than on plant remains (Potapov et al., 2022), and thus may not require the activity of an endogenous cellulase. A single acquisition event post-dating the divergence of the Neelipleona may explain this observation. However, since Neelipleona are represented only by a few taxa, the conclusion on cellulase absence should be taken cautiously. Within Symphypleona and Entomobryomorpha, cellulases are consistently absent in one clade each. The absence of cellulase genes is unexpected in the lucerne flea *Sminthurus viridis* that feeds on live leaf tissue (Greenslade & Ireson, 1986), and in the families of Dicyrtomidae, Bourletiellidae and Sminthuridae which consume mainly fresh plant materials (Potapov et al., 2020, 2022). The lack of cellulase genes suggests that these taxa either outsource cellulose decomposition to their midgut microbiome, rely on other GH families or do not digest cellulose. Besides these exceptions, the presence of GH45 cellulase genes seems to be a common trait among springtails.

Mites are the most numerous arthropods on land (Rosenberg et al., 2023), with most representatives in soil ecosystems belonging to Oribatida. Oribatid mites can have diverse feeding strategies, but they mostly feed on leaf litter at different decomposition stages and on microorganisms (Maraun et al., 2023). Our results showed that GH45 cellulases are present all over the basal clades of Oribatida while they were missing in half of the Branchypylina, which diversified later in the course of oribatid evolution. One interesting finding is that GH45 cellulases are missing especially in sexually reproducing mites, while they are present in other sexually reproducing taxa. However, the correlation between the reproduction mode and the

presence of the GH45 cellulases is not statistically significant. An alternative hypothesis is that parthenogenetic oribatid mites tend to occupy lower trophic positions and typically function as primary decomposers, opposed to secondary decomposers feeding predominantly on microorganisms (Fischer et al., 2014). However, ecological interpretation of these patterns is difficult since we do not know if species without GH45 cellulase genes contain other classes of cellulases, digest cellulose with the help of their microbiome or are indeed incapable of cellulose digestion. In general, the complex pattern of GH45 presence is similar to the low phylogenetic conservatism of ecological traits in oribatids, such as feeding mode (Potapov et al., 2022).

As expected from previous findings (Kirsch et al., 2014), we detected cellulases in Coleoptera. GH45 cellulases were completely absent in the genomes of Chilopoda and Diplopoda. The latter was surprising as Diplopoda are a key litter-feeding soil invertebrate group (Joly et al., 2020; Potapov et al., 2022). However, it was previously shown that the millipede *Telodeinopus aoutii* relies on its gut microbiome for lignocellulose degradation (Sardar et al., 2022). GH45 cellulases were also absent in Gamasina mites which are predators and therefore might not benefit from cellulose degradation. The first report of endogenous cellulases in Thysanoptera suggests that our analysis uncovers only the tip of the iceberg. We expect that taxonomically broad genome sequencing of eukaryotes promoted, for example, by the Earth BioGenome Project (Formenti et al., 2022; Lewin et al., 2022) will recover further animal groups in possession of enzymes targeting lignocellulose decomposition.

Taken together, our data suggest an early acquisition of a GH45 cellulase during the diversifications of both springtails and oribatids, instead of repeated horizontal transfer events. This implies that the possession of a GH45 cellulase is an ancestral trait in these groups. Similar to our results, cellulase acquisition was shown to be important for the diversification of herbivorous beetles (Kirsch et al., 2014). Differences in the GH45 cellulase gene tree from the oribatid and springtail species trees likely result from a highly dynamic evolution of the GH45 cellulase repertoire. Lineage-specific gene duplications and losses have partially disconnected the evolutionary history of the contemporary cellulase genes from the phylogeny of the species they are found in (Figure 1c; Figure S3). Lineage-specific duplications have been described for other cellulases (Shelomi et al., 2016; Shin et al., 2022), and differential duplicate loss has been shown to result in gene tree–species tree incongruencies (Parey et al., 2020). The presence of cellulases detected in thrips suggests that processes similar to those in Oribatida and Collembola might have been important also during the evolution of other arthropod groups.

We want to emphasize that the widespread presence of cellulase genes in soil invertebrates does not exclude that many species also rely on cellulolytic enzymes from microorganisms. Lignocellulose is a highly complex composite of cellulose, hemicellulose, lignin and pectin and its decomposition requires complex enzymatic cocktails (Cragg et al., 2015). For example, enzymes from the microbiome and the host work synergistically for lignocellulose degradation in the pill bug *Armadillidium vulgare* (Bredon et al., 2018). Wood decomposition

in termites is also a synergistic activity carried out together by insect and gut symbiont enzymes (Brune, 2014). In addition, transient microbiota, consumed together with lignocellulosic fragments, may also be secreting relevant enzymes. Decomposition performed by invertebrates in synergy with gut microorganisms would be a very flexible system capable of fast adaptation to environmental changes.

Cellulase presence in invertebrate genomes is not a proof of function. However, several lines of evidence point towards functionality. First, domain architecture of GH45 cellulases in fungi (*H. insolens* and *R. solani*) and the springtail *F. candida* are similar, and the fungal enzymes also have similar predicted 3D structures with the cellulases found by us (Figure 1d). Second, the catalytic residues that are important for the enzymatic function (Davies et al., 1993) are conserved in the two fungi and the springtail. Finally, orthologues with conserved domain architectures were retained over hundreds of millions of years of evolution in springtails and oribatids. This suggests that little change has occurred in the trophic niche and position of springtails and oribatid mites in soil food webs since their evolutionary origins. This view is further supported by the presence of both taxonomic groups in the first fossil soils (Schaefer & Caruso, 2019; Shear et al., 1984). Taken together, these are strong indications that GH45 cellulases in springtails and oribatids perform cellulose decomposition. Future work to experimentally evaluate the functional properties of soil invertebrate cellulases (Song et al., 2017) should consider all glycoside hydrolase genes, as gene duplication events may have led to substrate diversification (Busch et al., 2019; Shin et al., 2022), with duplicates being able to break down other polysaccharides like xyloglucan, mannans or xylan. While orthologues can be identified bioinformatically, functional properties need to be confirmed experimentally by expressing these enzymes heterologously, and test their substrate specificity on cellulose and hemicellulose polysaccharides.

Strong evidence exists that fungal GH45 cellulase genes have been repeatedly horizontally transferred to invertebrates. Several properties can explain why the initially foreign genes were integrated as an evolutionarily stable part of the recipient species' metabolic network. First, cellulases, as secreted, gut-acting enzymes (Fischer et al., 2013) do not depend on existing physiological pathways and their regulation for proper functioning in the recipient organism. Second, such enzymes likely reach their correct extracellular destination directly after the successful transfer of the cellulase gene, its incorporation into the genome and its translation, because the signals of protein export generally work independent of origin in most other taxa, even over vast evolutionary distances (Clérico et al., 2008). Third, the reaction catalysed by cellulases yields products that can serve as a beneficial fitness-relevant resource in any organism, because the necessary downstream pathways are ubiquitously present. Supporting these arguments, genes transferred horizontally are often secreted proteins (Savory et al., 2015; Undheim & Jenner, 2021).

The horizontal acquisition of cellulases and of other plant cell wall-degrading enzymes likely was a key event driving the evolutionary emergence of herbivory in arthropods (Wybouw et al., 2016). It coincided, for example, with the massive radiation of Phytophaga,

the most species-rich clade of beetles (Busch et al., 2019), and in adaptation to lignocellulose-rich diets in crustaceans (King et al., 2010). The long-term evolutionary preservation of GH45 genes suggests that cellulases likely confer fitness benefits also to soil invertebrates. These benefits may come from direct use of plant carbohydrate resources, although some theories imply that soil invertebrates are limited rather by access to proteins, but not by access to carbohydrates. The ability to degrade complex polysaccharides may also provide access to more nutritious, protein-rich cytosols or microorganisms colonizing the inside of plant cells, such as saprotrophic fungi, which are considered as major dietary components of both collembolans and oribatids (Pollierer & Scheu, 2021). We expect that cellulase presence/absence patterns across large taxon collections integrated with phenotypic traits of the respective species and the trophic niches they inhabit (Maraun et al., 2023; Potapov et al., 2022) will provide insights into the functional ecology and evolution of soil invertebrates.

It was previously shown that individual invertebrate species are capable of degrading lignocellulose independent from their gut microbiome (King et al., 2010). The widespread presence of GH45 cellulases in springtails and oribatid mites suggests that such endogenous cellulolytic abilities are substantially more common in invertebrates than it is generally appreciated. In addition to bacteria and fungi, invertebrates should, therefore, be considered a third evolutionarily and ecologically distinct group with such capability. This has important consequences for our understanding of soil food webs and the soil carbon cycle. Fungi compared to bacteria are known to react differently to environmental change such as experimental warming (Melillo et al., 2017) or habitat degradation (Zhou et al., 2018). This results from key differences in life-history strategies, for example, growth rates or nutrient use (Jansson & Hofmockel, 2020). Their differential reaction to environmental change influences decomposition as distinct taxa determine the rate and biochemical pathways of organic matter processing (Crowther et al., 2019). For example, fungal-based food webs in soils, and the processes of C and N loss they govern, are more resistant against and are better adaptable to drought than bacterial food webs (de Vries et al., 2012). Fungi accordingly contribute more to litter decomposition than bacteria under drought conditions (Ullah et al., 2023). Soil invertebrates react differently to environmental change compared to microorganisms (Sünnemann et al., 2021). Given key differences in life-history strategies among soil invertebrates, bacteria and fungi, a comparison of endogenous cellulolytic capabilities of soil invertebrates and microorganisms is an important direction for future research. We hypothesize that global change has a more detrimental impact on decomposition performed by soil invertebrates, given their lower effective population sizes and adaptive elasticity (Lanfear et al., 2014; Pauls et al., 2013). It might be essential to consider all three systematic groups and their differences for a better integration of belowground processes into ecosystem models (Chertov et al., 2017; Deckmyn et al., 2020; Filser et al., 2016) including global carbon models (Friedlingstein et al., 2022), and for better predictions of soil carbon and nutrient cycling.

AUTHOR CONTRIBUTIONS

MB, IE, HM, SS, IS, AP, MP, YP and GC conceptualized the study. IE, HM, FA, MB, CS, JR, RL, KH and GC contributed methodology and materials. HM, FA and IE performed the analyses and visualized the results. IE supervised the analyses. MB and IE wrote the original draft of the manuscript. HM, GC, TH, KH, RL, YP, MP, AP, JR, IS, SS, CS, IE and MB reviewed and finalized the manuscript. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Genome data will be available via NCBI BioProject PRJNA758215.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

Genome data: Only publicly available genome data were used for this research. The gene sets originated from NCBI RefSeq or GenBank (see Table S1), while the genome assemblies of soil-living invertebrates were taken from Bioproject PRJNA758215.

Metadata: Metadata associated with this study are available via figshare (https://doi.org/10.6084/m9.figshare.c.7121512.v1).

BENEFIT-SHARING STATEMENT

The authors have no benefits to report.

ORCID

Hannah Muelbaier https://orcid.org/0009-0009-4651-8650
Yannick Pauchet https://orcid.org/0000-0002-2918-8946
Markus Pfenninger https://orcid.org/0000-0002-1547-7245
Clément Schneider https://orcid.org/0000-0003-3743-9319
Ingo Ebersberger https://orcid.org/0000-0001-8187-9253
Miklós Bálint https://orcid.org/0000-0003-0499-8536

REFERENCES

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V., & Notredame, C. (2006). Expresso: Automatic incorporation of structural information in multiple sequence alignments using 3D-coffee. *Nucleic Acids Research*, 34, W604–W608. https://doi.org/10.1093/nar/gkl092

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19, 455–477. https://doi.org/10.1089/cmb.2012.0021

Berg, M., Stoffer, M., & Heuvel, H. (2004). Feeding guilds in Collembola based on digestive enzymes. *Pedobiologia*, 48, 589–601. https://doi.org/10.1016/j.pedobi.2004.07.006

Berto, G. L., Velasco, J., Tasso Cabos Ribeiro, C., Zanphorlin, L. M., Noronha Domingues, M., Tyago Murakami, M., Polikarpov, I., de Oliveira, L. C., Ferraz, A., & Segato, F. (2019). Functional characterization and comparative analysis of two heterologous endoglucanases from diverging subfamilies of glycosyl hydrolase family 45. *Enzyme and Microbial Technology*, 120, 23–35. https://doi.org/10.1016/j.enzmictec.2018.09.005

Bharadwaj, V. S., Knott, B. C., Ståhlberg, J., Beckham, G. T., & Crowley, M. F. (2020). The hydrolysis mechanism of a GH45 cellulase and its potential relation to lytic transglycosylase and expansin function. *The Journal of Biological Chemistry*, 295, 4477–4487. https://doi.org/10.1074/jbc.RA119.011406

Birikmen, M., Bohnsack, K. E., Tran, V., Somayaji, S., Bohnsack, M. T., & Ebersberger, I. (2021). Tracing eukaryotic ribosome biogenesis factors into the archaeal domain sheds light on the evolution of functional complexity. *Frontiers in Microbiology*, 12, 2598. https://doi.org/10.3389/fmicb.2021.739000

Bradford, M. A., Veen, G. F., Bonis, A., Bradford, E. M., Classen, A. T., Cornelissen, J., Crowther, T. W., De Long, J. R., Freschet, G. T., Kardol, P., Manrubia-Freixa, M., Maynard, D. S., Newman, G. S., Logtestijn, R. S. P., Viketoft, M., Wardle, D. A., Wieder, W. R., Wood, S. A., & van der Putten, W. H. (2017). A test of the hierarchical model of litter decomposition. *Nature Ecology & Evolution*, 1, 1836–1845. https://doi.org/10.1038/s41559-017-0367-4

Bredon, M., Dittmer, J., Noël, C., Moumen, B., & Bouchon, D. (2018). Lignocellulose degradation at the holobiont level: Teamwork in a keystone soil invertebrate. *Microbiome*, 6, 162. https://doi.org/10.1186/s40168-018-0536-y

Briones, M. J. I. (2018). The serendipitous value of soil Fauna in ecosystem functioning: The unexplained explained. *Frontiers in Environmental Science*, 6, 1–11. https://doi.org/10.3389/fenvs.2018.00149

Brune, A. (2014). Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews. Microbiology*, 12, 168–180. https://doi.org/10.1038/nrmicro3182

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12, 59–60. https://doi.org/10.1038/nmeth.3176

Busch, A., Danchin, E. G. J., & Pauchet, Y. (2019). Functional diversification of horizontally acquired glycoside hydrolase family 45 (GH45) proteins in Phytophaga beetles. *BMC Evolutionary Biology*, 19, 100. https://doi.org/10.1186/s12862-019-1429-9

Challis, R., Richards, E., Rajan, J., Cochrane, G., & Blaxter, M. (2020). BlobToolKit—Interactive quality assessment of genome assemblies. *G3: Genes, Genomes, Genetics*, 10, 1361–1374. https://doi.org/10.1534/g3.119.400908

Chang, W. H., & Lai, A. G. (2018). Mixed evolutionary origins of endogenous biomass-depolymerizing enzymes in animals. *BMC Genomics*, 19, 483. https://doi.org/10.1186/s12864-018-4861-0

Chertov, O., Komarov, A., Shaw, C., Bykhovets, S., Frolov, P., Shanin, V., Grabarnik, P., Priputina, I., Zubkova, E., & Shashkov, M. (2017). Romul_Hum—A model of soil organic matter formation coupling with soil biota activity. II. Parameterisation of the soil food web biota activity. *Ecological Modelling*, 345, 125–139. https://doi.org/10.1016/j.ecolmodel.2016.10.024

Clérico, E. M., Maki, J. L., & Gierasch, L. M. (2008). Use of synthetic signal sequences to explore the protein export machinery. *Peptide Science*, 90, 307–319. https://doi.org/10.1002/bip.20856

Collins, G., Schneider, C., Boštjančić, L. L., Burkhardt, U., Christian, A., Decker, P., Ebersberger, I., Hohberg, K., Lecompte, O., Merges, D., Muelbaier, H., Romann, J., Römbke, J., Rutz, C., Schmelz, R.,

Schmidt, A., Theissinger, K., Veres, R., Lehmitz, R., … Bálint, M. (2023). The MetaInvert soil invertebrate genome resource provides insights into below-ground biodiversity and evolution. *Communications Biology*, 6(1), Article 1. https://doi.org/10.1038/s42003-023-05621-4

Cragg, S. M., Beckham, G. T., Bruce, N. C., Bugg, T. D., Distel, D. L., Dupree, P., Etxabe, A. G., Goodell, B. S., Jellison, J., McGeehan, J. E., McQueen-Mason, S. J., Schnorr, K., Walton, P. H., Watts, J. E., & Zimmer, M. (2015). Lignocellulose degradation mechanisms across the tree of life. *Current Opinion in Chemical Biology*, 29, 108–119. https://doi.org/10.1016/j.cbpa.2015.10.018

Crowther, T. W., van den Hoogen, J., Wan, J., Mayes, M. A., Keiser, A. D., Mo, L., Averill, C., & Maynard, D. S. (2019). The global soil community and its influence on biogeochemistry. *Science*, 365, eaav0550. https://doi.org/10.1126/science.aav0550

Davies, G. J., Dodson, G. G., Hubbard, R. E., Tolley, S. P., Dauter, Z., Wilson, K. S., Hjort, C., Mikkelsen, J. M., Rasmussen, G., & Schülein, M. (1993). Structure and function of endoglucanase V. *Nature*, 365, 362–364. https://doi.org/10.1038/365362a0

de Vries, F. T., Liiri, M. E., Bjørnlund, L., Bowker, M. A., Christensen, S., Setälä, H. M., & Bardgett, R. D. (2012). Land use alters the resistance and resilience of soil food webs to drought. *Nature Climate Change*, 2, 276–280. https://doi.org/10.1038/nclimate1368

Deckmyn, G., Flores, O., Mayer, M., Domene, X., Schnepf, A., Kuka, K., Looy, K. V., Rasse, D. P., Briones, M. J. I., Barot, S., Berg, M., Vanguelova, E., Ostonen, I., Vereecken, H., Suz, L. M., Frey, B., Frossard, A., Tiunov, A., Frouz, J., … Yuste, J. C. (2020). KEYLINK: Towards a more integrative soil representation for inclusion in ecosystem scale models. I. Review and model concept. *PeerJ*, 8, e9750. https://doi.org/10.7717/peerj.9750

Dosch, J., Bergmann, H., Tran, V., & Ebersberger, I. (2023). FAS: Assessing the similarity between proteins using multi-layered feature architectures. *Bioinformatics*, 39, btad226. https://doi.org/10.1093/bioinformatics/btad226

Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B., & Terrapon, N. (2022). The carbohydrate-active enzyme database: Functions and literature. *Nucleic Acids Research*, 50, D571–D577. https://doi.org/10.1093/nar/gkab1045

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792–1797. https://doi.org/10.1093/nar/gkh340

FAO, ITPS, GSBI, CBD, EC. (2020). *State of knowledge of soil biodiversity—Status, challenges and potentialities, report 2020*. FAO. https://doi.org/10.4060/cb1928en

Filser, J., Faber, J. H., Tiunov, A. V., Brussaard, L., Frouz, J., De Deyn, G., Uvarov, A. V., Berg, M. P., Lavelle, P., Loreau, M., Wall, D. H., Querner, P., Eijsackers, H., & Jiménez, J. J. (2016). Soil fauna: Key to new carbon models. *The Soil*, 2, 565–582. https://doi.org/10.5194/soil-2-565-2016

Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., Bateman, A., & Eddy, S. R. (2015). HMMER web server: 2015 update. *Nucleic Acids Research*, 43, W30–W38. https://doi.org/10.1093/nar/gkv397

Fischer, B. M., Meyer, E., & Maraun, M. (2014). Positive correlation of trophic level and proportion of sexual taxa of oribatid mites (Acari: Oribatida) in alpine soil systems. *Experimental & Applied Acarology*, 63, 465–479. https://doi.org/10.1007/s10493-014-9801-3

Fischer, R., Ostafe, R., & Twyman, R. M. (2013). Cellulases from insects. In A. Vilcinskas (Ed.), *Yellow biotechnology II: Insect biotechnology in plant protection and industry, advances in biochemical engineering/biotechnology* (pp. 51–64). Springer. https://doi.org/10.1007/10_2013_206

Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., Ciofi, C., Crottini, A., Godoy, J. A., Höglund, J., Malukiewicz, J., Mouton, A., Oomen, R. A., Paez, S., Palsbøll, P. J., Pampoulie, C., Ruiz-López, M. J., Svardal, H., Theofanopoulou, C., … Zammit, G. (2022). The era of reference genomes in conservation genomics. *Trends in Ecology & Evolution*, 37, 197–202. https://doi.org/10.1016/j.tree.2021.11.008

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Alkama, R., … Zheng, B. (2022). Global carbon budget 2022. *Earth System Science Data*, 14, 4811–4900. https://doi.org/10.5194/essd-14-4811-2022

García-Palacios, P., Maestre, F. T., Kattge, J., & Wall, D. H. (2013). Climate and litter quality differently modulate the effects of soil fauna on litter decomposition across biomes. *Ecology Letters*, 16, 1045–1053. https://doi.org/10.1111/ele.12137

Greenslade, P., & Ireson, J. E. (1986). Collembola of the southern Australian culture steppe and Urban environments: A review of their Pest status and key to identification. *Australian Journal of Entomology*, 25, 273–291. https://doi.org/10.1111/j.1440-6055.1986.tb01115.x

Griffiths, H. M., Ashton, L. A., Evans, T. A., Parr, C. L., & Eggleton, P. (2019). Termites can decompose more than half of deadwood in tropical rainforest. *Current Biology*, 29, R118–R119. https://doi.org/10.1016/j.cub.2019.01.012

Griffiths, H. M., Ashton, L. A., Parr, C. L., & Eggleton, P. (2021). The impact of invertebrate decomposers on plants and soil. *The New Phytologist*, 231, 2142–2149. https://doi.org/10.1111/nph.17553

Han, Z., Sieriebriennikov, B., Susoy, V., Lo, W.-S., Igreja, C., Dong, C., Berasategui, A., Witte, H., & Sommer, R. J. (2022). Horizontally acquired Cellulases assist the expansion of dietary range in Pristionchus Nematodes. *Molecular Biology and Evolution*, 39, msab370. https://doi.org/10.1093/molbev/msab370

Hong, S. M., Sung, H. S., Kang, M. H., Kim, C.-G., Lee, Y.-H., Kim, D.-J., Lee, J. M., & Kusakabe, T. (2014). Characterization of Cryptopygus antarcticus Endo-β-1,4-Glucanase from Bombyx mori expression systems. *Molecular Biotechnology*, 56, 878–889. https://doi.org/10.1007/s12033-014-9767-8

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of Molecular Graphics*, 14, 33–38. https://doi.org/10.1016/0263-7855(96)00018-5

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., & Tappu, R. (2016). MEGAN Community edition—Interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Computational Biology*, 12, e1004957. https://doi.org/10.1371/journal.pcbi.1004957

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23, 254–267. https://doi.org/10.1093/molbev/msj030

Jansson, J. K., & Hofmockel, K. S. (2020). Soil microbiomes and climate change. *Nature Reviews. Microbiology*, 18, 35–46. https://doi.org/10.1038/s41579-019-0265-7

Joly, F.-X., Coq, S., Coulis, M., David, J.-F., Hättenschwiler, S., Mueller, C. W., Prater, I., & Subke, J.-A. (2020). Detritivore conversion of litter into faeces accelerates organic matter turnover. *Communications Biology*, 3, 1–9. https://doi.org/10.1038/s42003-020-01392-4

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772–780. https://doi.org/10.1093/molbev/mst010

Kern, M., McGeehan, J. E., Streeter, S. D., Martin, R. N. A., Besser, K., Elias, L., Eborall, W., Malyon, G. P., Payne, C. M., Himmel, M. E.,

Schnorr, K., Beckham, G. T., Cragg, S. M., Bruce, N. C., & McQueen-Mason, S. J. (2013). Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance. *Proceedings of the National Academy of Sciences*, *110*, 10189–10194. https://doi.org/10.1073/pnas.1301502110

King, A. J., Cragg, S. M., Li, Y., Dymond, J., Guille, M. J., Bowles, D. J., Bruce, N. C., Graham, I. A., & McQueen-Mason, S. J. (2010). Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes. *Proceedings of the National Academy of Sciences*, *107*, 5345–5350. https://doi.org/10.1073/pnas.0914228107

Kirsch, R., Gramzow, L., Theißen, G., Siegfried, B. D., Ffrench-Constant, R. H., Heckel, D. G., & Pauchet, Y. (2014). Horizontal gene transfer and functional diversification of plant cell wall degrading polygalacturonases: Key events in the evolution of herbivory in beetles. *Insect Biochemistry and Molecular Biology*, *52*, 33–50. https://doi.org/10.1016/j.ibmb.2014.06.008

Kück, P., & Longo, G. C. (2014). FASconCAT-G: Extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in Zoology*, *11*, 81. https://doi.org/10.1186/s12983-014-0081-x

Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., Li, M., Stecher, G., & Hedges, S. B. (2022). TimeTree 5: An expanded resource for species divergence times. *Molecular Biology and Evolution*, *39*, msac174. https://doi.org/10.1093/molbev/msac174

Lanfear, R., Kokko, H., & Eyre-Walker, A. (2014). Population size and the rate of evolution. *Trends in Ecology & Evolution*, *29*, 33–41. https://doi.org/10.1016/j.tree.2013.09.009

Letunic, I., & Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*, W293–W296. https://doi.org/10.1093/nar/gkab301

Levy Karin, E., Mirdita, M., & Söding, J. (2020). MetaEuk—Sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. *Microbiome*, *8*, 48. https://doi.org/10.1186/s40168-020-00808-x

Lewin, H. A., Richards, S., Lieberman Aiden, E., Allende, M. L., Archibald, J. M., Bálint, M., Barker, K. B., Baumgartner, B., Belov, K., Bertorelle, G., Blaxter, M. L., Cai, J., Caperello, N. D., Carlson, K., Castilla-Rubio, J. C., Chaw, S.-M., Chen, L., Childers, A. K., Coddington, J. A., … Zhang, G. (2022). The earth BioGenome project 2020: Starting the clock. *Proceedings of the National Academy of Sciences of the United States of America*, *119*, e2115635118. https://doi.org/10.1073/pnas.2115635118

Louca, S., & Doebeli, M. (2018). Efficient comparative phylogenetics on large trees. *Bioinformatics*, *34*, 1053–1055. https://doi.org/10.1093/bioinformatics/btx701

Maraun, M., Thomas, T., Fast, E., Treibert, N., Caruso, T., Schaefer, I., Lu, J.-Z., & Scheu, S. (2023). New perspectives on soil animal trophic ecology through the lens of C and N stable isotope ratios of oribatid mites. *Soil Biology and Biochemistry*, *177*, 108890. https://doi.org/10.1016/j.soilbio.2022.108890

Melillo, J. M., Frey, S. D., DeAngelis, K. M., Werner, W. J., Bernard, M. J., Bowles, F. P., Pold, G., Knorr, M. A., & Grandy, A. S. (2017). Long-term pattern and magnitude of soil carbon feedback to the climate system in a warming world. *Science*, *358*, 101–105. https://doi.org/10.1126/science.aan2874

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*, 268–274. https://doi.org/10.1093/molbev/msu300

Nozaki, M., Miura, C., Tozawa, Y., & Miura, T. (2009). The contribution of endogenous cellulase to the cellulose digestion in the gut of earthworm (Pheretima hilgendorfi: Megascolecidae). *Soil Biology and Biochemistry*, *41*, 762–769. https://doi.org/10.1016/j.soilbio.2009.01.016

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*, D733–D745. https://doi.org/10.1093/nar/gkv1189

Okmane, L., Nestor, G., Jakobsson, E., Xu, B., Igarashi, K., Sandgren, M., Kleywegt, G. J., & Ståhlberg, J. (2022). Glucomannan and beta-glucan degradation by *Mytilus edulis* Cel45A: Crystal structure and activity comparison with GH45 subfamily A. *B and C. Carbohydrate Polymers*, *277*, 118771. https://doi.org/10.1016/j.carbpol.2021.118771

Palomares-Rius, J. E., Hirooka, Y., Tsai, I. J., Masuya, H., Hino, A., Kanzaki, N., Jones, J. T., & Kikuchi, T. (2014). Distribution and evolution of glycoside hydrolase family 45 cellulases in nematodes and fungi. *BMC Evolutionary Biology*, *14*, 69. https://doi.org/10.1186/1471-2148-14-69

Parey, E., Louis, A., Cabau, C., Guiguen, Y., Roest Crollius, H., & Berthelot, C. (2020). Synteny-guided resolution of gene trees clarifies the functional impact of whole-genome duplications. *Molecular Biology and Evolution*, *37*, 3324–3337. https://doi.org/10.1093/molbev/msaa149

Pauls, S. U., Nowak, C., Bálint, M., & Pfenninger, M. (2013). The impact of global climate change on genetic diversity within populations and species. *Molecular Ecology*, *22*, 925–946. https://doi.org/10.1111/mec.12152

Pausas, J. G., & Bond, W. J. (2020). On the three major recycling pathways in terrestrial ecosystems. *Trends in Ecology & Evolution*, *35*, 767–775. https://doi.org/10.1016/j.tree.2020.04.004

Phillips, H. R. P., Guerra, C. A., Bartz, M. L. C., Briones, M. J. I., Brown, G., Crowther, T. W., Ferlian, O., Gongalsky, K. B., Hoogen, J. v. d., Krebs, J., Orgiazzi, A., Routh, D., Schwarz, B., Bach, E. M., Bennett, J. M., Brose, U., Decaëns, T., König-Ries, B., Loreau, M., … Eisenhauer, N. (2019). Global distribution of earthworm diversity. *Science*, *366*, 480–485. https://doi.org/10.1126/science.aax4851

Pollierer, M. M., & Scheu, S. (2021). Stable isotopes of amino acids indicate that soil decomposer microarthropods predominantly feed on saprotrophic fungi. *Ecosphere*, *12*, e03425. https://doi.org/10.1002/ecs2.3425

Post, W. M., Peng, T.-H., Emanuel, W. R., King, A. W., Dale, V. H., & DeAngelis, D. L. (1990). The global carbon cycle. *American Scientist*, *78*, 310–326.

Potapov, A., Bellini, B. C., Chown, S. L., Deharveng, L., Janssens, F., Kováč, L., Kuznetsova, N., Ponge, J.-F., Potapov, M., & Querner, P. (2020). Towards a global synthesis of Collembola knowledge: Challenges and potential solutions. *Soil Organisms*, *92*, 161–188.

Potapov, A. M., Beaulieu, F., Birkhofer, K., Bluhm, S. L., Degtyarev, M. I., Devetter, M., Goncharov, A. A., Gongalsky, K. B., Klarner, B., Korobushkin, D. I., Liebke, D. F., Maraun, M., Mc Donnell, R. J., Pollierer, M. M., Schaefer, I., Shrubovych, J., Semenyuk, I. I., Sendra, A., Tuma, J., … Scheu, S. (2022). Feeding habits and multifunctional classification of soil-associated consumers from protists to vertebrates. *Biological Reviews*, *97*, 1057–1117. https://doi.org/10.1111/brv.12832

Potapov, A. M., Guerra, C. A., van den Hoogen, J., Babenko, A., Bellini, B. C., Berg, M. P., Chown, S. L., Deharveng, L., Kováč, Ľ., Kuznetsova, N. A., Ponge, J.-F., Potapov, M. B., Russell, D. J., Alexandre, D., Alatalo, J. M., Arbea, J. I., Bandyopadhyaya, I., Bernava, V., Bokhorst, S., … Scheu, S. (2023). Globally invariant metabolism but density-diversity mismatch in springtails. *Nature Communications*, *14*, 674. https://doi.org/10.1038/s41467-023-36216-6

Pryszcz, L. P., & Gabaldón, T. (2016). Redundans: An assembly pipeline for highly heterozygous genomes. *Nucleic Acids Research*, *44*, e113. https://doi.org/10.1093/nar/gkw294

Revell, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Roberts, E., Eargle, J., Wright, D., & Luthey-Schulten, Z. (2006). MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics*, 7, 382. https://doi.org/10.1186/1471-2105-7-382

Rosenberg, Y., Bar-On, Y. M., Fromm, A., Ostikar, M., Shoshany, A., Giz, O., & Milo, R. (2023). The global biomass and number of terrestrial arthropods. *Science Advances*, 9, eabq4049. https://doi.org/10.1126/sciadv.abq4049

Russell, R. B., & Barton, G. J. (1992). Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels. *Proteins*, 14, 309–323. https://doi.org/10.1002/prot.340140216

Sardar, P., Šustr, V., Chroňáková, A., & Lorenc, F. (2022). Metatranscriptomic holobiont analysis of carbohydrate-active enzymes in the millipede Telodeinopus aoutii (Diplopoda, Spirostreptida). *Frontiers in Ecology and Evolution*, 10, 1–27. https://doi.org/10.3389/fevo.2022.931986

Savory, F., Leonard, G., & Richards, T. A. (2015). The role of horizontal gene transfer in the evolution of the oomycetes. *PLoS Pathogens*, 11, e1004805. https://doi.org/10.1371/journal.ppat.1004805

Schaefer, I., & Caruso, T. (2019). Oribatid mites show that soil food web complexity and close aboveground-belowground linkages emerged in the early Paleozoic. *Communications Biology*, 2, 1–8. https://doi.org/10.1038/s42003-019-0628-7

Shear, W. A., Bonamo, P. M., Grierson, J. D., Rolfe, W. D. I., Smith, E. L., & Norton, R. A. (1984). Early land animals in North America: Evidence from Devonian age arthropods from Gilboa, New York. *Science*, 224, 492–494. https://doi.org/10.1126/science.224.4648.492

Shelomi, M., Heckel, D. G., & Pauchet, Y. (2016). Ancestral gene duplication enabled the evolution of multifunctional cellulases in stick insects (Phasmatodea). *Insect Biochemistry and Molecular Biology*, 71, 1–11. https://doi.org/10.1016/j.ibmb.2016.02.003

Shelomi, M., Watanabe, H., & Arakawa, G. (2014). Endogenous cellulase enzymes in the stick insect (Phasmatodea) gut. *Journal of Insect Physiology*, 60, 25–30. https://doi.org/10.1016/j.jinsphys.2013.10.007

Shimodaira, H., & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, 16, 1114–1116. https://doi.org/10.1093/oxfordjournals.molbev.a026201

Shin, N. R., Doucet, D., & Pauchet, Y. (2022). Duplication of horizontally acquired GH5_2 enzymes played a central role in the evolution of Longhorned beetles. *Molecular Biology and Evolution*, 39, msac128. https://doi.org/10.1093/molbev/msac128

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Song, J. M., Hong, S. K., An, Y. J., Kang, M. H., Hong, K. H., Lee, Y.-H., & Cha, S.-S. (2017). Genetic and structural characterization of a Thermo-tolerant, cold-active, and acidic Endo-β-1,4-glucanase from Antarctic springtail, Cryptopygus antarcticus. *Journal of Agricultural and Food Chemistry*, 65, 1630–1640. https://doi.org/10.1021/acs.jafc.6b05037

Steenwyk, J. L., Iii, T. J. B., Li, Y., Shen, X.-X., & Rokas, A. (2020). ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biology*, 18, e3001007. https://doi.org/10.1371/journal.pbio.3001007

Steinegger, M., & Salzberg, S. L. (2020). Terminating contamination: Large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biology*, 21, 115. https://doi.org/10.1186/s13059-020-02023-1

Sünnemann, M., Siebert, J., Reitz, T., Schädler, M., Yin, R., & Eisenhauer, N. (2021). Combined effects of land-use type and climate change on soil microbial activity and invertebrate decomposer activity. *Agriculture, Ecosystems and Environment*, 318, 107490. https://doi.org/10.1016/j.agee.2021.107490

Tran, N.-V., Greshake Tzovaras, B., & Ebersberger, I. (2018). PhyloProfile: Dynamic visualization and exploration of multi-layered phylogenetic profiles. *Bioinformatics*, 34, 3041–3043. https://doi.org/10.1093/bioinformatics/bty225

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., Pritzel, A., Figurnov, M., Ronneberger, O., Bates, R., Kohl, S. A. A., ... Hassabis, D. (2021). Highly accurate protein structure prediction for the human proteome. *Nature*, 596, 590–596. https://doi.org/10.1038/s41586-021-03828-1

Ullah, M. R., Carrillo, Y., & Dijkstra, F. A. (2023). Relative contributions of fungi and bacteria to litter decomposition under low and high soil moisture in an Australian grassland. *Applied Soil Ecology*, 182, 104737. https://doi.org/10.1016/j.apsoil.2022.104737

Undheim, E. A. B., & Jenner, R. A. (2021). Phylogenetic analyses suggest centipede venom arsenals were repeatedly stocked by horizontal gene transfer. *Nature Communications*, 12, 818. https://doi.org/10.1038/s41467-021-21093-8

van den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D. A., de Goede, R. G. M., Adams, B. J., Ahmad, W., Andriuzzi, W. S., Bardgett, R. D., Bonkowski, M., Campos-Herrera, R., Cares, J. E., Caruso, T., de Brito, C. L., Chen, X., Costa, S. R., Creamer, R., ... Crowther, T. W. (2019). Soil nematode abundance and functional group composition at a global scale. *Nature*, 572, 194–198. https://doi.org/10.1038/s41586-019-1418-6

Watanabe, H., Noda, H., Tokuda, G., & Lo, N. (1998). A cellulase gene of termite origin. *Nature*, 394, 330–331. https://doi.org/10.1038/28527

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with kraken 2. *Genome Biology*, 20, 257. https://doi.org/10.1186/s13059-019-1891-0

Wybouw, N., Pauchet, Y., Heckel, D. G., & Van Leeuwen, T. (2016). Horizontal gene transfer contributes to the evolution of arthropod herbivory. *Genome Biology and Evolution*, 8, 1785–1801. https://doi.org/10.1093/gbe/evw119

Zhou, Z., Wang, C., & Luo, Y. (2018). Effects of forest degradation on microbial communities and soil carbon cycling: A global meta-analysis. *Global Ecology and Biogeography*, 27, 110–124. https://doi.org/10.1111/geb.12663

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Muelbaier, H., Arthen, F., Collins, G., Hickler, T., Hohberg, K., Lehmitz, R., Pauchet, Y., Pfenninger, M., Potapov, A., Romahn, J., Schaefer, I., Scheu, S., Schneider, C., Ebersberger, I., & Bálint, M. (2024). Genomic evidence for the widespread presence of GH45 cellulases among soil invertebrates. *Molecular Ecology*, 00, e17351. https://doi.org/10.1111/mec.17351