Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

From shared genes to staged decay: ecosystem-specific genomic potential for litter decomposition in soils

Kirsten Küsel

kirsten.kuesel@uni-jena.de

Aquatic Geomicrobiology, Institute of Biodiversity, Ecology and Evolution, Friedrich Schiller University Jena, Jena, Germany https://orcid.org/0000-0002-5396-0975

Bala Singavarapu

Aquatic Geomicrobiology, Institute of Biodiversity, Ecology and Evolution, Friedrich Schiller University Jena, Jena, Germany https://orcid.org/0000-0002-0147-895X

Zander Rainier Human

Institute of Microbiology of the Czech Academy of Sciences

Guillaume Patoine

German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstr. 4, 04103 Leipzig, Germany;Institute of Biology, Leipzig University, Puschstr. 4, 04103 Leipzig, Germany https://orcid.org/0000-0002-3748-6644

Anna Heintz-Buschart

University of Amsterdam - SILS https://orcid.org/0000-0002-9780-1933

Qicheng Bei

University of Southern California

Carl-Eric Wegner

Heinrich Heine University Dusseldorf

Nico Eisenhauer

Leipzig University https://orcid.org/0000-0002-0371-6720

François Buscot

Helmholtz Centre for Environmental Research https://orcid.org/0000-0002-2364-0006

Carlos A. Guerra

Universidade de Coimbra, Departamento de Geografia

Ika Djukic

Environment Agency Austria https://orcid.org/0000-0002-5144-9321

Narendrakumar Chaudhari

Institute of Biodiversity, Ecology and Evolution

Simone Cesarz

German Centre for Integrative Biodiversity Research (iDiv) https://orcid.org/0000-0003-2334-5119

Antonis Chatzinotas

Helmholtz Centre for Environmental Research https://orcid.org/0000-0002-0387-9802

Carol Adair

Rubenstein School of Environment and Natural Resources. University of Vermont. Burlington, Vermont, USA 05403

Bernd Ahrends

Northwest German Forest Research Institute, Grätzelstrasse 2, 37079 Göttingen, Germany https://orcid.org/0000-0001-7400-7297

Christopher Andrews

UK Centre for Ecology and Hydrology https://orcid.org/0000-0003-2428-272X

Andrijana Andrić

University of Novi Sad, BioSense Institute, Dr Zorana Đinđića 1, 21000 Novi Sad, Serbia

Milan Barna

Institute of Forest Ecology, Slovak Academy of Sciences https://orcid.org/0000-0002-8389-7997

Claus Beier

University of Copenhagen

Björn Berg

University of Helsinki https://orcid.org/0000-0002-6776-4158

Andreas Bohner

Agricultural Research and Education Centre Raumberg-Gumpenstein, Raumberg 38, Irdning-Donnersbachtal 8952, Austria

Cristina Branquinho

Faculdade de Ciências da Universidade de Lisboa https://orcid.org/0000-0001-8294-7924

Helge Bruelheide

Martin Luther University Halle-Wittenberg https://orcid.org/0000-0003-3135-0356

Rafaella Canessa

Martin Luther Universität Halle-Wittenberg https://orcid.org/0000-0002-6979-9880

Michele Carbognani

University of Parma https://orcid.org/0000-0001-7701-9859

Roberto Cazzolla Gatti

Biome Lab, Department of Biological, Geological, and Environmental Sciences

Chiling Chen

Taiwan Agricultural Research Institute

Casper Christiansen

Terrestrial Ecology Section, Department of Biology, University of Copenhagen, Denmark

Michael Danger

Université de Lorraine

Evgeny Davydov

Altai State University https://orcid.org/0000-0002-2316-8506

Christine Delire

CNRM, Université de Toulouse, Météo-France, CNRS https://orcid.org/0000-0002-6114-3211

Valter Di Cecco

Maiella Seed Bank - Maiella National Park, via Badia, 28, 67039 Sulmona (AQ), Italy

Luciano Di Martino

Maiella Seed Bank - Maiella National Park, via Badia, 28, 67039 Sulmona (AQ), Italy

Jan Dick

Markus Didion

Swiss Federal Institute for Forest, Snow and Landscape Research WSL, 8903 Birmensdorf, Switzerland https://orcid.org/0000-0003-0346-0646

Simon Drollinger

University of Göttingen https://orcid.org/0000-0002-4912-2629

Tsutomu Enoki

Kyushu University Forest, Kyushu University, 1-85 Kita 5, Ashoro, Hokkaido 089-3705 Japan https://orcid.org/0000-0003-3767-938X

István Fekete

Institute of Environmental and Natural Sciences, University of Nyíregyháza, Nyíregyháza, Hungary

Heike Feldhaar

Universität Bayreuth

Olga Ferlian

German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig https://orcid.org/0000-0002-2536-7592

Nina Filippova

Yugra State University

Veronika Fontana

Institute for Alpine Environment, Eurac Research https://orcid.org/0000-0002-0666-7434

Petra Fransson

Swedish University of Agricultural Sciences https://orcid.org/0000-0003-0842-9197

Martin Freitag

Institute of Landscape Ecology, University of Münster, Münster, Germany

Mark Frenzel

Department of Community Ecology, Helmholtz Centre for Environmental Research-UFZ

Karibu Fukuzawa

Nakagawa Experimental Forest, Field Science Center for Northern Biosphere, Hokkaido

Rosario Gavilán

Botany Unit. Dept Pharmacology, Pharmacognosy and Botany. Complutens University https://orcid.org/0000-0002-1022-445X

Aurora Gaxiola

Pontificia Universidad Católoca de Chile & Instituto de Ecología y Biodiversidad, Chile

Romain Georges

UMR 6553 Ecosystèmes Biodiversité, Evolution, Université de Rennes, LTSER France, Zone Atelier Armorique, 35042 Rennes, France

Stephan Glatzel

Department of Geography and Regional Research, Geoecology, University of Vienna https://orcid.org/0000-0002-2572-5484

Maria Glushkova

Forest Research Institute-Bulgarian Academy of Sciences

Grizelle González

USDA Forest Service, International Institute of Tropical Forestry https://orcid.org/0000-0003-3007-5540

Ken Green

National Parks and Wildlife Service

Peter Haase

Senckenberg Society for Nature Research - Leibniz Institution for Biodiversity and Earth System Research https://orcid.org/0000-0002-9340-0438

Frank Hagedorn

Swiss Federal Institute for Forest, Snow and Landscape Research (WSL) https://orcid.org/0000-0001-5218-7776

Ute Hamer

Institute of Landscape Ecology, University of Münster https://orcid.org/0000-0002-3845-3983

Rebecca Hewitt

Department of Environmental Studies, Amherst College, USA

Takuo Hishi

Shiiba Research Forest, Kyushu University, 949 Ohkawauchi, Shiiba Village, Miyazaki Prefecture 883-0402, Japan

Tsutom Hiura

Department of Ecosystem Studies, The University of Tokyo

Kazuhiko Hoshizaki

Department of Biological Environment, Akita Prefectural University https://orcid.org/0000-0001-7654-9129

Hervé Jactel

INRAE, University of Bordeaux https://orcid.org/0000-0002-8106-5310

Juan Jiménez

IPE/CSIC https://orcid.org/0000-0003-2398-0796

Heinke Jäger

Charles Darwin Foundation https://orcid.org/0000-0001-9065-4048

Róbert Kanka

Institute of Landscape Ecology, Slovak Academy of Sciences, Štefánikova 3, 81499 Bratislava, Slovakia

Sebastian Kepfer Rojas

University of Copenhagen https://orcid.org/0000-0002-1681-2877

Till Kleinebecker

Institute of Landscape Ecology and Resource Management, University of Gießen
https://orcid.org/0000-0003-1121-2861

Yuji Kominami

Department of Disaster Prevention, Meteorology and Hydrology, Forestry and Forest Products Research Institute, Tsukuba, Japan

Zsolt Kotroczó

Hungarian University of Agriculture and Life Sciences https://orcid.org/0000-0002-5637-2802

Kaie Kriiska

Institute of Ecology and Earth Sciences, University of Tartu https://orcid.org/0000-0002-4464-3433

Hiroko Kurokawa

Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan; Forestry and Forest Products Research Institute, Tsukuba 305-8687, Japan

Kate Lajtha

Department of Crop and Soil Sciences, Oregon State University, Corvallis OR 97331, USA

Klaus Larsen

Department of Geosciences and Natural Resource Management, University of Copenhagen
https://orcid.org/0000-0002-1421-6182

Anika Lehman

Freie Universität Berlin, Institute of Biology, Berlin 14195, Germany

Alfred Lochner

German Center for Integrative Biodiversity Research

John Loehr

Faculty of Biological and Environmental Sciences, Lammi Biological Station, University of Helsinki
https://orcid.org/0000-0002-6212-0273

Jörg Löffler

University of Bonn https://orcid.org/0000-0002-9320-6168

Stefan Löfgren

Swedish University of Agricultural Sciences, SLU https://orcid.org/0000-0001-7892-2708

Vincent Maire

Département des sciences de l'environnement, Université du Québec à Trois-Rivières, Canada

Florence Maunoury-Danger

LIEC, UMR7360 Université de Lorraine-CNRS

Henning Meesenburg

Northwest German Forest Research Institute https://orcid.org/0000-0002-3035-4737

Inara Melece

Institute of Biology, University of Latvia

Joël Merlet

Behaviour and Ecology of Wildlife, INRAE Toulouse-Occitanie, 24 chemin de Borderouge 31326 Auzeville-Tolosane

Umberto Morra di Cella

Environmental Sustainability and Climate Change Unit, Environmental Protection Agency of Aosta Valley (ARPA), 11020 Saint-Christophe, Italy

Ivika Ostonen

University of Tartu https://orcid.org/0000-0001-9043-6083

Alain Paquette

Centre for Forest Research, Université du Québec à Montréal https://orcid.org/0000-0003-1048-9674

William Parker

Ontario Forest Research Institute

Pablo Peri

INTA-UNPA-CONICET https://orcid.org/0000-0002-5398-4408

Alessandro Petraglia

University of Parma https://orcid.org/0000-0003-4632-2251

Veronika Piscová

Institute of Landscape Ecology, Slovak Academy of Sciences, Štefánikova 3, 81499 Bratislava, Slovakia

Quentin Ponette

Université catholique de Louvain https://orcid.org/0000-0002-2726-7392

Marie-Noëlle Pons

Université de Lorraine, CNRS, Laboratoire Réactions et Génie des Procédés, LTSER-Zone Atelier Bassin de la Moselle, Nancy, France

Anne Probst

Centre de Recherche sur la Biodiversité et l'Environnement (CRBE), Université de Toulouse, Toulouse INP, CNRS, IRD, Toulouse, France

Jean-Luc Probst

Centre de Recherche sur la Biodiversité et l'Environnement (CRBE), Université de Toulouse, Toulouse INP, CNRS, IRD, Toulouse, France

Mihai Puscas

Babeş-Bolyai University

Tshililo Ramaswiela

South African Environmental Observation Network (SAEON), Kimberley, South Africa, 8306

Peter Reich

University of Minnesota https://orcid.org/0000-0003-4424-662X

Matthias Rillig

Free University of Berlin https://orcid.org/0000-0003-3541-7853

Marcus Schaub

Swiss Federal Research Institute WSL https://orcid.org/0000-0002-0158-8892

Stefan Scheu

JF Blumenbach Institute of Zoology and Anthropology, University of Göttingen https://orcid.org/0000-0003-4350-9520

Anja Schmidt

Helmholtz Centre for Environmental Research – UFZ, Department of Community Ecology https://orcid.org/0000-0001-5339-219X

Inger Schmidt

University of Copenhagen https://orcid.org/0000-0003-3880-2060

Julia Seeber

Eurac Research https://orcid.org/0000-0003-0189-7377

Steffen Seitz

Osnabrück University https://orcid.org/0000-0003-4911-3906

Helena Serrano

CE3C, Centre for Ecology, Evolution and Environmental Changes https://orcid.org/0000-0003-3743-8084

Ana Sousa

ECOMARE, CESAM-Centre for Environmental and Marine Studies, Department of Biology,University of Aveiro https://orcid.org/0000-0003-0783-5177

Jutta Stadler

Helmholtz Centre for Environmental Research https://orcid.org/0000-0003-4456-8709

Angela Stanisci

University of Molise https://orcid.org/0000-0002-5302-0932

Artur Stefanski

Department of Forest Resources, University of Minnesota; College of Natural Resources, University of Wisconsin https://orcid.org/0000-0002-5412-1014

Flurin Sutter

Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland

Jean-Paul Theurillat

Université de Genève https://orcid.org/0000-0002-1843-5809

Katja Tielbörger

University of Tübingen https://orcid.org/0009-0003-7767-1734

Naoko Tokuchi

Kyoto University

Marcello Tomaselli

University of Parma

Stacey Trevathan-Tackett

Deakin Marine Research and Innovation Centre, School of Life and Environmental Science, Deakin University

Stefan Trogisch
Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg https://orcid.org/0000-0002-1426-1012
Pavel Dan Turtureanu
Babeș-Bolyai University https://orcid.org/0000-0002-7422-3106
Tudor Ursu
Institute of Biological Research Cluj, branch of NIRDBS Bucharest https://orcid.org/0000-0002-4898-6345
Matthew Vadeboncoeur
University of New Hampshire https://orcid.org/0000-0002-8269-0708
Liesbeth van den Brink
UT https://orcid.org/0000-0003-0313-8147
Helga van der Merwe
South African Environmental Observation Network (SAEON), Arid Lands Node, PO Box 110040, Hadison Park, Kimberley, 8306, Northern Cape, South Africa https://orcid.org/0000-0002-7677-5123
Inge van Halder
University of Bordeaux, INRAE, BIOGECO, 69 route d'Arcachon 33610, Cestas, France
Kris Vandekerkhove
Research Institute for Nature and Forest https://orcid.org/0000-0003-1954-692X
Susanna Venn
Deakin University https://orcid.org/0000-0002-7433-0120
Kris Verheyen
Ghent University https://orcid.org/0000-0002-2067-9108
Arne Verstraeten
Research Institute for Nature and Forest (INBO) https://orcid.org/0000-0002-8813-1261
Jeyanny Vijayanathan
Forest Research Institute Malaysia (FRIM)
Pascal Vittoz
University of Lausanne https://orcid.org/0000-0003-4218-4517
Markus Wagner
Northwest German Forest Research Institute https://orcid.org/0009-0006-2600-9304
Martin Weih
Swedish University of Agricultural Sciences https://orcid.org/0000-0003-3823-9183
Laura Williams
University of Minnesota https://orcid.org/0000-0003-3555-4778
Georg Wohlfahrt
University of Innsbruck https://orcid.org/0000-0003-3080-6702
Dirk Wundram
Department of Geography, University of Bonn, Meckenheime Allee 166, 53115 Bonn, Germany

Franz Zehetner
University of Natural Resources and Applied Life Sciences
Miglena Zhiyanski
Forest Research Institute-Bulgarian Academy of Sciences, Sofia, Bulgaria

Article

Keywords:

Posted Date: August 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9962094/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Abstract

Plant litter decomposition governs how much carbon soils store and emit, yet the microbial traits that shape ecosystem-scale decay remain unresolved. Metagenomes can quantify genes encoding plant cell-wall-degrading enzymes, but it is unclear whether ecosystem differences in decay reflect distinct enzymatic repertoires, and whether these data improve prediction beyond climate and soil properties. We paired standardized green and rooibos tea-bag decomposition assays across 3–24 months with 295 soil metagenomes from 264 global sites. Using 196 European plots for primary inference, we built a stage-resolved catalogue of 17.6 million carbohydrate-active enzyme (CAZyme) genes. Forest microbiomes decomposed tea faster than grasslands, but this was not explained by greater CAZyme family richness. Instead, ecosystems differed in CAZyme abundance, subfamily and protein-sequence variation, and allocation across biochemical stages of plant cell-wall decay, with evidence of ecosystem-specific selection. CAZyme profiles added explanatory power for 24-month mass loss and improved within-ecosystem prediction but generalized poorly across ecosystems and continents. By showing that ecosystem differences in decomposition arise from the stage-specific distribution of shared enzymatic functions rather than their presence alone, this work shifts microbial trait inference beyond gene inventories and provides a mechanistic genomic framework for carbon-cycle modelling within defined environmental limits.

Introduction

Plant litter decomposition returns a large fraction of net primary production to soils and releases ~ 60 Pg C yr^{-1} as CO_2, making it a dominant terrestrial carbon flux and key control on nutrient release and soil formation[1–3]. Litter decomposition and standardised tea bag experiments show that rates vary widely across climates and ecosystems[4,5]. Even with common substrates, ecosystems such as forests and grasslands follow distinct mass-loss trajectories, implying that local soil conditions and decomposer communities modulate decay beyond climate and litter traits[3,4].

Soil microorganisms initiate litter decomposition through the secretion of extracellular enzymes that attack intact plant polymers. This extracellular depolymerization represents the first biochemical stage of plant cell-wall decomposition and is a key biological regulator of nutrient cycling and ecosystem process rates[6,7].

Plant cell walls comprise cellulose, hemicelluloses, pectins and lignin, whose breakdown proceeds through biochemical stages converting complex polymers into partially depolymerized polymers, oligomers and ultimately monomers[8]. Carbohydrate-active enzymes (CAZymes) catalyse these steps, and variation in their amino acid sequences can alter enzyme specificity and sensitivity[9]. Thus, CAZyme gene repertoires and sequence variants within CAZyme families (subfamilies) link microbial communities to biochemical-stage-specific decomposition and realised litter decay[10,11]. Determining whether ecosystems differ in this genomic potential and whether it improves prediction of decomposition beyond climate and soils is critical for projections of ecosystem carbon turnover.

Metagenomic surveys suggest substantial overlap in broad functional gene repertoires among soils, while regional studies report climate- and land-use-associated shifts in CAZyme family composition[12–14]. However, CAZymes are often analysed at coarse functional resolution or within single biomes and are rarely organised along biochemical stages of plant cell-wall decomposition, limiting inference about which decomposition steps differ among ecosystems. Meanwhile, tea bag experiments have quantified global controls on mass loss, including strong substrate-quality and climate effects, and have documented ecosystem-dependent patterns[3,5,15]. Yet microbial information in these experiments has largely been limited to community composition or bulk enzyme activity[16,17], leaving unclear whether cross-ecosystem differences in decomposition are reflected in CAZyme gene repertoires, particularly at subfamily resolution across decomposition stages, and whether CAZyme profiles improve prediction beyond environmental and spatial covariates.

To address these gaps, we paired green (labile), and rooibos (more recalcitrant) tea bags incubated for up to 24 months with shotgun metagenomes from soils sampled at the 24-month endpoint across 264 sites (295 samples) worldwide. We used these materials to span a broad litter-chemistry gradient rather than to test tea-type effects. We focus inference on 196 European grassland, broadleaf-forest and coniferous-forest soils, and use the full dataset for predictive modelling and cross-ecosystem transfer tests. We compiled tea mass-loss data at 3, 12 and 24 months with soil physicochemical variables and functional indicators, grouped CAZymes by target plant-cell-wall substrates and biochemical decomposition stage (Fig. 1) using the Database for automated Carbohydrate-active enzyme Annotation (dbCAN) and enzyme-commission (EC) evidence[18,19], quantified their contribution to mass loss relative to environmental and spatial covariates and tested CAZyme-informed models for generalisation.

We hypothesised that (1) soil microbial CAZyme composition differs systematically among ecosystems and decomposition stages, mirroring variation in tea litter mass-loss trajectories; and that (2) CAZyme profiles explain variation in tea litter mass loss beyond environmental and spatial covariates, so that models including CAZyme profiles outperform those based only on environmental and soil functional indicators.

Results

Forest soils show greater tea litter decomposition than grasslands

In the primary European dataset (n = 196 plots; Methods), tea litter mass loss was higher in forest soils than in grasslands, whereas broadleaf and coniferous forests did not differ significantly (Fig. 2). For green tea, mass loss was higher in both forest types than in grasslands at 3, 12 and 24 months, although the conifer–grassland contrast at 3 months was only marginal (Fig. 2a,b). For rooibos, ecosystem differences followed the same direction but were weaker, with a single robust contrast at 12 months, when coniferous forests showed greater mass loss than grasslands (Fig. 2c,d). Standardized forest–

grassland differences (Hedges' g) were generally small to moderate, with the largest differences for green tea at 12–24 months and for the conifer–grassland contrast in rooibos at 12 months (Fig. 2b,d). Overall, forest soils consistently showed modestly greater tea litter mass loss than grasslands across both tea types and time.

Global soil metagenomic resource reveals a substrate- and stage-resolved CAZyme catalogue for plant litter decomposition

Shotgun soil metagenomes from tea bag decomposition sites generated a large, bacteria-dominated resource containing millions of CAZyme sequences that we organized by target plant cell-wall substrates (hereafter, substrates) and biochemical decomposition stages (hereafter, stages) (Fig. 1; Supplementary Tables 1–7). The taxonomic classification of assembled metagenomic contigs was dominated by Bacteria (66.99%), whereas smaller fractions were assigned to Metazoa (9.21%), Fungi (2.80%), Viridiplantae (0.91%), Archaea (0.17%) and Viruses (0.05%); 3.93% were classified only as Eukaryota, and 15.92% remained unclassified. In the primary European dataset, high-quality assemblies comprised ~ 4.3 Tbp from 29 billion quality-filtered paired-end reads, yielding 47 million contigs ≥ 1 kb and 348 million coding sequences (CDSs), including 17.6 million CAZyme annotations spanning 379 families and 11,865 subfamilies. To characterize the broader resource used for predictive modelling, we also annotated the full global dataset (n = 295; 6.7 Tbp total sequence), including 70 non-European sites that expanded the catalogue to 400 families and 13,218 subfamilies (Supplementary Table 8).

To quantify decomposition potential at finer biochemical resolution, we curated a stage-resolved CAZyme reference linking families to major substrates (cellulose, hemicelluloses, pectin and lignin) and to initial, mid, last or broad stages of decomposition (Fig. 1; Supplementary Tables 1–7). We defined 93 parent CAZyme families and, using subfamily-level EC mapping, expanded these to 159 stage-specific entries spanning substrates and stages. This staged catalogue enabled comparisons of CAZyme gene repertoires across ecosystems by substrate and stage and provided the framework for testing whether substrate- and stage-specific CAZyme profiles explain variation in tea litter mass loss beyond environmental and spatial covariates.

CAZyme gene abundances, rather than richness, differ by substrate and ecosystem

CAZyme family richness assigned to each substrate was broadly similar across ecosystems (medians: cellulose 26–27 families, xylan 36–37, lignin 8), with only minor differences for hemicellulose-related categories (xyloglucan and mixed-linkage β-glucans; Δmedian ≤ 1 family) (Fig. 3a). In contrast, relative abundance within the CAZyme pool differed among ecosystems, indicating quantitative shifts in genomic repertoire rather than gains or losses of substrate-targeting families (Fig. 3b; Supplementary Tables 9–10): conifer forests showed higher cellulose- and hemicellulose-associated CAZymes than grasslands (Cliff's δ = 0.55 for cellulose, 0.67 for mannans and 0.66 for xyloglucan). Broadleaf forests

showed higher lignin-associated CAZymes than grasslands (Cliff's δ = 0.47), whereas grasslands showed higher xylan-associated CAZymes than broadleaf forests (Cliff's δ = 0.25). Together, these patterns indicate a largely shared substrate-targeting repertoire across ecosystems, with ecosystem differences driven primarily by quantitative shifts in CAZyme gene abundances.

CAZyme-associated taxonomic composition varies mainly across stages

The phylum-level composition of CAZyme-annotated genes varied strongly across decomposition stages, with smaller but consistent differences among ecosystems (Fig. 4a). Across stages and ecosystems, Actinomycetota, Acidobacteriota and Pseudomonadota contributed most to the CAZyme gene pool. Stage structure was evident in all ecosystems: the mid-stage CAZyme pool showed an increased contribution from Acidobacteriota, while Pseudomonadota declined from initial to last stages. Ecosystem differences were secondary to stage effects: Acidobacteriota contributed relatively more CAZyme genes in conifer forests, whereas Pseudomonadota contributed relatively more to broadleaf forests. Actinomycetota were broadly represented across ecosystems, with a higher contribution in grasslands; Verrucomicrobiota contributed more in broadleaf forests and Planctomycetota more in conifer forests across stages (Fig. 4a). Overall, these patterns show that the microbial communities contributing most to CAZyme-encoded decomposition potential shift across ecosystems by stage, with the strongest shift at the mid stage.

Stage-resolved differential abundance highlights ecosystem-dependent CAZyme investment across decomposition steps

To identify which CAZyme families differed among ecosystems within each stage, we tested for differential abundance across pairwise ecosystem contrasts (Fig. 4b). Across all stage-by-contrast tests, we detected 104 significant differences (adjusted $P < 0.05$), concentrated in broad-acting families (43) and initial-stage families (26), with additional differences in mid (18) and last (17) stages. The largest number of enrichments occurred in conifer forests for broad-stage families (24 significant enrichments) and in broadleaf forests for initial-stage families (13), whereas grasslands showed the strongest enrichment in last-stage families (7). Functionally, broadleaf forests enrichments were biased toward oxidative and de-esterifying functions in initial steps, conifer forests enrichments were biased toward mid-stage depolymerization families (including core cellulose/hemicellulose hydrolases), and grasslands enrichments were biased toward terminal hemicellulose/xylan-processing families. Together, these results indicate that ecosystems share a largely common CAZyme family repertoire but differ in the stage- and substrate-dependent allocation of CAZyme family abundances.

Ecosystem divergence emerges within shared CAZyme families across subfamilies, phylogenies and protein

sequences

To test whether ecosystem signatures persist within the CAZyme families shared across ecosystems, we resolved families to subfamilies and assessed divergence in abundance, phylogenetic breadth and protein sequence. At subfamily resolution, enrichment patterns were strongly structured by substrate and decomposition stage (Fig. 5). Across all substrate–stage combinations, conifer forests contained the largest number of subfamily enrichments (580), followed by grasslands (386) and broadleaf forests (355) (Fig. 5a). Ecosystem-unique enrichments were lesser in broadleaf forests (59) but common in conifer forests (292) and grasslands (304) (Fig. 5b). These unique signals were concentrated in hemicellulose- and cellulose-associated decomposition stages (for example, unique subfamilies in xylan: 15, 67, 89; xyloglucan: 11, 59, 56; mannans: 10, 55, 41 for broadleaf, conifer and grasslands, respectively; totals across stages). Overall, broadleaf forests showed predominant enrichment in lignin-broad stage, conifer forests in mannan-, pectin- and xyloglucan-associated stages, and grasslands in xylan, cellulose, lignin and mix-β-glucans with enrichments concentrated at the initial or final stages.

Ecosystem divergence within shared families was also evident at the phylogenetic level. For CAZyme families represented in all three ecosystems, conifer forests exhibited higher phylogenetic diversity than grasslands for 48 families and higher than broadleaf forests for 10 (Supplementary Fig. 1), indicating greater within-family evolutionary breadth in conifer-associated communities. At the protein sequence level, directional selection analyses further supported ecosystem-associated divergence within CAZyme families. Across ecosystems, we detected strong evidence of directional selection in 42 families (Bayes factor ≥ 100), reaching ~ 6.9% of sites in AA6, ~ 5.4% in CE2, ~ 5.3% in GH94 and ~ 4.8% in PL26 (Supplementary Fig. 2). Frequently selected residues included glycine, aspartic acid, phenylalanine, and proline, which are important residues in catalytic and binding domains. Ecosystem-targeted analysis revealed contrasting residue preferences across ecosystems: broadleaf forests showed pronounced shifts in pectin-related families (PL22/PL26) and in CE2 and several GH families toward polar residues such as glutamine and asparagine, whereas conifer forests showed shifts toward aromatic or hydrophobic residues (for example, phenylalanine in CE2) along with increased proline substitutions (for example, in PL26). In contrast, grasslands showed shifts toward aliphatic residues (leucine, isoleucine, and valine) across multiple CAZyme families.

Together, subfamily-level specialisation, differences in within-family phylogenetic diversity, and ecosystem-associated directional shifts in CAZyme protein sequences indicate that ecosystem divergence in genomic potential for litter decomposition manifests at multiple molecular scales.

CAZymes explain unique variation in litter mass loss and improve within-ecosystem prediction, but generalise poorly across ecosystems

To test whether ecosystem-structured CAZyme repertoires translate into realised decomposition outcomes and add information beyond environmental context, we partitioned explained variation in 24-month tea mass loss among environmental (soil physicochemical and climatic variables), space, soil

functional indicators (microbial respiration and biomass), and CAZyme profiles (Fig. 6). Across tea types, the predictor sets explained a moderate fraction of variation, with stronger explanatory power for rooibos than for green tea (Fig. 6). For green tea, the model explained 32% of 24-month mass-loss variation, dominated by a unique environmental fraction (21%). CAZyme profiles contributed an additional 6% uniquely, plus 4% shared with other predictor sets. For rooibos, the model explained 44%, again led by the environment (18% unique), while CAZymes accounted for 1% unique and 6% shared fractions (Fig. 6). Together, the results indicate a non-redundant component of explained variation in tea mass loss attributable to CAZyme repertoires.

We next assessed whether CAZyme information improves out-of-sample prediction of decomposition trajectories using the full global dataset (n = 295; Methods). Predictors in baseline models (m0) comprised environmental variables and soil functional indicators. We compared these with models additionally including CAZyme profiles using elastic-net regression (m1) and random forests (m2), evaluated under four validation regimes: within-ecosystem, leave-one-ecosystem-out, leave-one-continent-out, and out of Europe (Supplementary Fig. 3). Within ecosystems, CAZymes improved predictive performance most consistently for random forests: pooled out-of-sample R^2 increased from 0.17 to 0.28 for AUC, 0.09 to 0.30 for the 3-month slope, 0.05 to 0.27 for the 12-month slope, and from ~ 0 to 0.15 for the 24-month slope. Elastic-net models showed modest gains (R^2 ~ 0.11, 0.07, 0.01 for 3-, 12-, and 24-month slopes) and did not improve AUC. However, under biogeographic extrapolation, pooled out-of-sample R^2 was often ≤ 0 despite occasional relative improvements, indicating limited transferability across ecosystems and continents (Supplementary Fig. 3; Supplementary Tables 11–12).

Discussion

Here we show that ecosystem differences in litter decomposition are not explained by differences in the inventory of plant cell-wall-degradation genes, but by how a largely shared CAZyme repertoire is allocated across the biochemical stages of decomposition. This allocation offers a mechanistic basis to incorporate microbial traits into decomposition models and shows that gene–rate relationships are strongly context dependent. We asked whether ecosystems that differ in litter decomposition activity also differ in how microbial genomic potential for degradation of plant cell-wall substrates is allocated across biochemical stages, and whether that information improves prediction of decomposition beyond environment and space. Despite ecosystem differences in tea mass loss, decomposition-related CAZyme family richness was similar, whereas stage-resolved CAZyme profiles differed and explained a small unique fraction of 24-month mass-loss variation. In predictive models, these profiles improved within-ecosystem prediction of decomposition metrics but added little power under cross-ecosystem or cross-continent extrapolation. Together, these results support our first hypothesis that ecosystems differ systematically in CAZyme profiles in ways that align with tea-litter mass-loss trajectories and partially support our second hypothesis that CAZymes add predictive information beyond environmental and spatial controls.

Tea bag assays provide a standardized decomposition sequence that can scaffold interpretation of stage-resolved genomic potential. However, the sampling intervals used here (3, 12 and 24 months) represent operational observation points rather than discrete biochemical stages, as decomposition processes overlap and vary among ecosystems. Green tea has a larger water-soluble fraction than rooibos (~ 0.49 versus ~ 0.21 g g^{-1} of initial mass), so early (~ 3-month) mass loss is dominated by leaching, with microbial activity becoming more important as soluble pools decline[5,20]. Losses between ~ 3 and 12 months increasingly reflect microbial depolymerisation of cellulose and hemicelluloses, whereas later changes (12–24 months) are constrained by lignified and physically protected material[3,20]. Forest soils lost more green-tea mass than grasslands at all three decomposition periods (3, 12 and 24 months), whereas ecosystem contrasts for rooibos were weaker and more time specific. Because the earliest phase is leaching-dominated, this pattern suggests that ecosystem differentiation emerges mainly during the 3–12-month transition, when enzymatic depolymerisation of plant cell-wall substrates dominates mass loss and aligns with higher relative abundances of CAZymes targeting cellulose and hemicelluloses and strong ecosystem differences in those profiles. Within ecosystems, CAZyme-augmented models most strongly improved prediction of the 3- and 12-month slopes and the overall mass-loss curve (AUC), with relatively small gains for the 24-month slope, indicating that genomic information is most informative during enzymatically dominated stages. Although tea bags provide standardised litter spanning a labile-to-more-recalcitrant chemistry gradient, native litter differs markedly among ecosystems: broadleaf forests receive relatively nutrient-rich, less lignified leaves; conifer forests receive more lignified, structurally defended needles that accumulate on acidic forest floors; and grasslands combine forb and graminoid litter with substantial below-ground root inputs[21–23]. Our results, including greater green-tea mass loss in forests than in grasslands and ecosystem-specific CAZyme allocation, are consistent with decomposer adaptation to dominant native-litter polymers. Thus, tea bags enable standardised comparison, but native-litter decomposition should be inferred cautiously because resident litters differ in chemistry, structure and placement.

Across ecosystems, soil microbiomes harboured hundreds of CAZyme families and thousands of subfamilies, yet the family richness of decomposition-related CAZymes was mostly similar in broadleaf forests, conifer forests and grasslands. This overlap points to a near-universal repertoire for substrate decomposition and indicates that functional gene repertoires can remain conserved despite large variation in microbial taxa, consistent with other global analyses[12,13,24]. Similar decoupling between taxonomy and function has been reported in aquatic systems, where microbial assemblages in oceans and aquifers maintain stable functional profiles for organic-matter degradation and nutrient cycling despite highly variable species composition[25–27]. Our findings extend this pattern to terrestrial litter decomposition and show that deeper resolution of CAZyme family- and subfamily-level abundances yields mechanistic insight: a broadly shared pool of plant cell-wall-degrading genes underpins decomposition across ecosystems, whereas differences in decomposition trajectories emerge primarily from quantitative reallocation of this repertoire rather than from which CAZyme families are present.

CAZyme gene repertoires were strongly substrate- and stage-specific: conifer forests showed stronger mid-stage investment in cellulose and hemicellulose depolymerases, consistent with greater tea mass loss between 3 and 12 months, whereas grasslands were enriched in terminal hemicellulose/xylan-processing families, providing a basis for sustained utilisation of hemicellulosic residues as decomposition progresses and accessible pools decline. Broadleaf forests also showed higher lignin-associated CAZyme investment than grasslands, consistent with greater capacity for oxidative steps linked to lignified fractions even when bulk mass loss differs only modestly between forest types. Subfamily-resolution analyses revealed ecosystem-unique, substrate- and stage-specific CAZyme variants, indicating divergence in fine-scale trait combinations (for example, kinetics, substrate range and sensitivity to local constraints) and supporting conceptual work that functional similarity declines when multiple functions and fine-scale traits are considered[28]. Consistent with this stage structuring, the taxonomic composition of CAZyme-encoding communities shifted more across biochemical stages than across ecosystems, with ecosystem context modulating these patterns (for example, relatively higher abundances of Acidobacteriota in conifer forests, Pseudomonadota in broadleaf forests, and Actinomycetota in grasslands). Stage effects nonetheless dominated, indicating that progression along the plant cell-wall decomposition pathway is a primary ecological filter shaping microbial gene abundances and reflecting successional shifts from fast-growing, resource-responsive taxa early to taxa favoured under lower substrate availability and more complex residues later[29,30].

We show that metagenome-derived CAZyme repertoires can be interrogated beyond coarse “enzyme classes” by resolving genes to CAZy families and subfamilies, within-family phylogenetic diversity and protein sequence, thereby capturing diversification that maps onto catalytic specificity and substrate affinity. Conventional bulk soil enzyme assays, such as β-glucosidase activity, instead quantify potential activity of broad enzyme classes with standardised proxy substrates and integrate across isoenzymes, regulatory state and spatial microhabitats, limiting their capacity to detect sequence-resolved functional potential[31,32]. Consistent with this complementarity, our variation-partitioning and predictive analyses indicate that CAZyme profiles should complement, rather than replace, enzyme assays in enzyme-explicit and microbe-explicit decomposition models. Environment remained the dominant predictor of 24-month mass loss, consistent with syntheses showing that climate, litter traits and soil properties largely control decomposition, yet within-family phylogenetic breadth and directional-selection signatures indicate that local environments fine-tune CAZyme variants, generating molecular-level ecosystem divergence that contributes to ecosystem-specific decomposition trajectories[1,3,11]. CAZyme augmentation improved within-ecosystem predictions, particularly in random-forest models relative to elastic-net models, revealing the non-linear, interaction-rich relationships expected when genomic potential is filtered by environmental constraints, whereas predictive power during extrapolation across ecosystems or continents was often near or below zero, even when CAZymes produced occasional relative improvements, underscoring strong context dependence in gene–rate relationships[29,33,34]. Our results therefore suggest that CAZymes can be included to improve litter-decomposition models, provided their context dependence is explicitly acknowledged and linked to ecosystem-specific environmental conditions and sequence-level CAZyme adaptation[16,35]. Specifically, CAZyme profiles are

best used as stage-resolved indicators of ecosystem-specific depolymerisation capacity rather than as universal rate constants. In empirical and hybrid multipool models, CAZyme indices could help parameterise the relative contribution of labile, intermediate and lignified pools, or scale substrate-specific modifiers for cellulose-, hemicellulose-, pectin- and lignin-rich material[36]. In enzyme- or microbe-explicit models, they could constrain enzyme allocation, relative catalytic potential or carbon routing into microbial biomass versus stabilised soil organic matter[37]. Improving model portability across ecosystems and biogeographic regions will require denser, more balanced metagenomic sampling of underrepresented ecosystems, such as tropical forests, wetlands, arid lands and tundra, to separate region-specific associations and link CAZyme sequence adaptation more directly to functional consequences.

Although this research provides key insights into links between soil microbial genomic potential for decomposition and tea mass loss, it has limitations. A limitation is that fungal genes were underdetected, which might bias decomposition-related functional interpretation toward bacterial processes in our metagenomic analyses. Soil metagenomes were sampled once, at 24 months, providing a snapshot of genomic potential rather than time-resolved expression or enzyme activities, and longitudinal studies show temporal decoupling between gene abundance, transcription and realised process rates[30,38,39]. Our mapping of CAZyme families and subfamilies to substrates and stages relies on current CAZy and dbCAN annotations and necessarily simplifies enzyme multifunctionality, although it is grounded in curated EC assignments[10,18]. Future work combining standardised and native litters, multi-timepoint metagenomes and metatranscriptomes, direct measurements of extracellular enzyme activities and metabolite profiles, and improved biochemical characterisation of CAZymes will be required to resolve how genomic potential, expression and environmental context jointly govern progression through biochemical stages of litter decomposition.

Together, these analyses reveal ecosystem-specific, stage-structured differences in genomic potential for plant cell-wall degradation that mirror variation in litter decomposition trajectories. By coupling stage-resolved CAZyme repertoires with time-resolved tea mass loss, we provide a mechanistic framework for incorporating microbial genomic traits into decomposition models and for designing microbial trait–informed ecosystem experiments. Improving forecasts of terrestrial carbon cycling will require moving beyond gene inventories to determine how conserved enzymatic functions are allocated across substrates and stages, how this allocation varies among ecosystems, and when it translates into measurable differences in litter decay and carbon–climate feedbacks.

Methods

Study system, tea bag incubation and soil sampling

Sampling was conducted within the International Soil Biogeography Consortium (iSBio; [https://home.uni-leipzig.de/idiv/isbio/]) in collaboration with the TeaComposition initiative ([https://www.teacomposition.org/]). In TeaComposition, litter decomposition was quantified over a 3-

year period (2016–2019) at 336 sites spanning nine biomes[40] using standardized tea bags (woven nylon, 250-µm mesh) containing green tea (Camellia sinensis) and rooibos tea (Aspalathus linearis)[5]. Tea bags were oven-dried (70°C, 48 h) and weighed prior to burial, then incubated during the local summer season (both hemispheres) in the upper 0–5 cm of soil. Four bags per tea type were retrieved after 3, 12, 24 and 36 months, cleaned, re-dried (70°C, 48 h) and re-weighed to calculate mass loss[3].

For the present study, soil sampling was performed at the 24-month tea bag harvest (summer 2018–2019). At each site, we established two sampling plots (Supplementary Fig. 4). Within each plot, four soil cores were collected from randomly selected points (0–5 cm depth) and pooled to generate a plot-level composite soil sample[41,42]. Composite samples were subdivided for molecular (DNA sequencing), physicochemical and soil functional analyses and transported to the respective laboratories under appropriate conditions. After quality filtering and sequencing, 295 plot-level composite samples from 264 sites yielded high-quality soil metagenomes; remaining samples were excluded primarily due to loss during transport and/or failure of sequencing quality control[42]. The resulting metagenomes spanning worldwide were paired with tea mass-loss data from the same plots at 3, 12 and 24 months (Supplementary Fig. 5). Tea litter mass loss was successfully quantified for 90%, 90% and 76% of samples at the 3-, 12- and 24-month harvests, respectively.

Ecosystem classification

Each site was assigned to a land-cover class using the European Space Agency Climate Change Initiative (ESA CCI) global land-cover map for 2015 (300-m resolution), which classifies pixels under the FAO Land Cover Classification System[43]. For analysis, ESA CCI classes were reclassified into eight broad categories: coniferous forests (CF), broadleaf forests (BF), tropical forests (TF), other forests (OF), grasslands (GL), shrublands (SL), croplands (CL) and sparse/bare lands (BL). Land-cover was assigned using local site descriptions and the closest matching ESA CCI category. Throughout, we treat these land-cover classes as operational terrestrial ecosystem types (sensu CBD/IUCN)[44,45] and refer to them as “ecosystems”.

Environmental and edaphic variables

Environmental covariates included geographic coordinates, climate variables and soil physicochemical properties for each sampling location. Latitude and longitude were recorded in the field, and elevation was extracted from WorldClim 2 (1-km resolution)[46]. Climate variables including mean annual temperature (MAT), temperature seasonality, mean annual precipitation (MAP) and precipitation seasonality were obtained from CHELSA (1-km resolution)[47].

Soil properties were measured from the fresh composite samples collected at the 24-month harvest, or from additional samples from the same sites where material was insufficient. Where values were missing, they were obtained from a global soil information database. Soil physical variables included texture fractions (clay, silt and sand), aggregate-stability metrics (mean weight diameter (MWD) and water-stable aggregates (WSA)) and soil water content (SWC). Soil chemical variables included pH, soil

organic carbon (SOC), total nitrogen (TN) and the carbon-to-nitrogen ratio (C:N). The MWD and WSA metrics were measured on soil samples standardized to a particle size of < 4 mm according to the protocol by Kemper & Rosenau[48]. Soil pH was measured using a pH meter at a sample-to-solution ratio of 1:5 (g mL^{-1}) in 1 M KCl. After grinding, soil organic carbon (SOC) and total nitrogen (TN) were determined using an elemental analyzer (NA 1110, CE Instruments, Rodano, Milan, Italy), and C:N ratios were subsequently calculated[42].

Soil microbial respiration and biomass

Soil basal respiration and microbial biomass carbon were measured as soil functional indicators[49]. Fresh subsamples from the composite samples were stored at 4°C after arrival at the laboratory, sieved, and analysed within 14 days. Basal respiration was measured using O2-microcompensation apparatus[50], and microbial biomass carbon by substrate-induced respiration[51]. For both measurements, 7 g of fresh soil was sieved to 2 mm to remove stones, roots and coarse debris. Basal respiration, representing the activity of metabolically active microorganisms at sampling, was recorded for 24 h, and the mean of the final 10 h stable phase was used for calculation. Microbial biomass carbon was then estimated after addition of 24 mg glucose and water adjusted to medium water-holding capacity. Biomass was calculated from the three lowest consecutive respiration values within the first 10 h after substrate addition, assuming no microbial growth during this period, and using a conversion factor of 38[52].

DNA extraction and shotgun metagenomic sequencing

For shotgun metagenomics, we analysed plot-level composite soils, targeting one plot per site; however, a small subset of sites had an additional plot sequenced, yielding 295 soil metagenomes from 264 sites with matched tea decomposition data. Soil subsamples for molecular analyses were preserved in RNAlater (Thermo Fisher Scientific, Darmstadt, Germany) in sterile microcentrifuge tubes and stored at − 20°C until processing[41,42]. DNA was extracted from composite soils using the DNeasy PowerSoil Kit (QIAGEN, Hilden, Germany) following the manufacturer's protocol. Libraries were prepared using the Illumina Nextera Flex kit and sequenced as paired-end reads (2 × 150 bp) on an Illumina NovaSeq platform at the NGS Competence Center Tübingen (NCCT).

Bioinformatic analyses

Raw reads were quality-filtered and sequencing adapters and low-quality bases were trimmed with the *Read_QC* module of metaWRAP v1.3 using default settings[53]. Processed reads from each sample were assembled individually with MEGAHIT v1.1.1-2 from metaWRAP[54]. Assembled contigs were taxonomically classified at the nucleotide level using Kraken2 against NCBI nt and RefSeq (Bacteria, Archaea, Viruses, Fungi, and Protozoa) to partition contigs into prokaryotic vs eukaryotic sequence sets[55]. Contigs < 1,000 bp were excluded to reduce false positives in gene prediction. Open reading frames (ORFs) were predicted from contigs identified as prokaryotic with Prodigal v2.6.3[56]. Eukaryotic contigs were annotated with MetaEuk v6.a5d39d9[57] using UniRef90 as the reference database[58].

Taxonomic assignments for predicted coding DNA sequences (CDSs) from both prokaryotic and eukaryotic contigs were then made at the protein level using MMseqs2 with the '2bLCA' setting against the NCBI nr database, to obtain conservative lowest-common-ancestor gene-level assignments based on protein homology[59]. CDS abundances were quantified by mapping reads with Salmon v1.10.1using the '--meta' flag, using a database consisting of predicted prokaryotic and eukaryotic CDS sequences as reference, to obtain read counts, effective gene lengths, and TPM values[60].

All CDSs coding for carbohydrate-active enzyme (CAZyme) families and subfamilies and the assignment of CAZy families to putative target-substrates were made with dbCAN3[19], which uses a combination of HMMER, DIAMOND, and dbCAN-sub[19,61,62]. We retained CAZyme hits supported by ≥ 2 of these assignment methods (majority-vote rule) as suggested by the dbCAN3 developers, implemented via a custom Python script.

For the identification of CAZymes acting on substrates derived from plant cell walls, we used a reference list from dbCAN-seq (Supplementary Table S2 in[63]), which was subsequently manually verified, and then grouped into different stages of substrate decomposition pathways, viz., initial, mid, and last, based on EC numbers and information from CAZypedia consortium[64] and the MetaCyc Database[65]. To further refine the assignment of CAZyme families into different stages of decomposition, we used dbCAN-sub[19] via a custom Python script to make subfamily-level substrate annotations (accessed on 14-05-2025). A substantial fraction of the CAZyme family-level assignments were not assigned into any category and were thus assigned as "broad" stage category. This can be further defined as CAZymes that lacking a single confirmed substrate specificity or that participating in multiple substrates or stages of decomposition. For lignin and mixed β-glucans, available resources did not support confident mid-stage assignments, and consequently, no mid-stage category was used for these substrates.

To generate CAZyme family and subfamily abundance matrices, CAZyme assignments on annotated genes were derived as above, filtered to bacteria, fungi, and archaea, and quantified as TPM values using Salmon with the same settings as before, using custom Python scripts.

To assess the amino acid level variation, the catalytic domains of sequences encoding CAZymes were extracted using HMMER, using the sequence coordinate ranges obtained from dbCAN3[19], clustered at 99% identity with USEARCH[66]and then aligned with MAFFT v7[67]. Subsequently, approximate maximum-likelihood trees were inferred with FastTree 2[68]. Directional selection between ecosystem groups (e.g., forest vs. grassland) was tested in a Bayesian framework using FADE (the FUBAR Approach to Directional Evolution) implemented in HyPhy, which detects branch-specific directional amino-acid preferences on rooted trees[69].

Statistical analyses

All statistical analyses were conducted in R v.4.4.1[70] using base functions and contributed packages as detailed below. Because the dataset was geographically imbalanced, with 225 of 295 metagenomes

originating from Europe and the most represented ecosystem types being grasslands (GL; n = 101), broadleaf forests (BF; n = 75), and conifer forests (CF; n = 61), we performed an *a priori* power analysis to determine adequate replication for robust inference. Assuming a medium-large effect size (Cohen's f = 0.30), α = 0.05 and statistical power of 0.80, we restricted primary inferential analyses to European sites within GL, BF, and CF ecosystem types unless otherwise stated. Descriptive statistics of metagenome assemblies were calculated using both the European subset and the full global dataset, where indicated in the main text.

For inferential analyses, metagenome assemblies were screened for outlier samples and assessed for homogeneity of variance, and only assemblies (n = 196) comprising contigs ≥ 1 kb were retained for downstream analyses. The full global metagenome dataset (n = 295) comprising the primary European dataset (n = 196), additional European samples (n = 29; total European n = 225), and samples from other continents and ecosystem types (n = 70), was used exclusively for predictive modelling and not for inferential tests. Where multiple hypothesis tests were performed, *P* values were adjusted using the Benjamini–Hochberg false discovery rate procedure[71].

Mass loss data for green and rooibos tea were first assessed for normality. To test for differences among ecosystems at each decomposition period (3,12 and 24 months) we applied Welch's ANOVA to mass loss separately for each tea type and time period, followed by Games-Howell post hoc tests for pairwise ecosystem comparisons, implemented in the 'rstatix' package[72]. Effect sizes for pairwise contrasts were calculated as Hedges' g with 95% confidence intervals.

CAZyme family richness was computed per sample, and ecosystem differences in richness were evaluated using pairwise Wilcoxon rank-sum tests. TPM proportions were used to derive CAZyme family abundances associated with plant-substrate decomposition, and ecosystem differences in these abundances per substrate were tested with pairwise Wilcoxon rank-sum tests. Length-scaled TPM counts were obtained from Salmon output using the 'tximport' package[73], and differential abundance of CAZyme families between ecosystems across pathway stages was analysed with 'DESeq2'[74]. To identify CAZyme subfamilies significantly enriched in each ecosystem across pathway stages, one-sided Wilcoxon rank-sum tests were applied to CAZyme subfamily TPM abundance matrices.

For phylogenetic diversity analyses, phylogenetic trees and CAZyme sequence-cluster presence-absence community matrices generated by the bioinformatic pipeline were imported into R, and Faith's phylogenetic diversity (PD) was calculated[75] using the 'picante' package[76]. Ecosystem differences in PD were assessed with pairwise Wilcoxon rank-sum tests.

Variation partitioning of tea litter mass loss was conducted separately for green and rooibos tea using the 'vegan' package[77], with four predictor blocks: environment, space, soil functional indicators, and CAZyme profiles. The environmental block comprised soil physicochemical and climatic variables, standardized after removal of collinear predictors (variance inflation factor > 10). The CAZyme block was derived from stage-total abundances for the broad, initial, mid and last decomposition stages; these

variables were centred and scaled and summarized by principal component analysis, and the first four principal components were used. The spatial block was constructed from positive dbMEM eigenfunctions derived from site coordinates projected in metres, and informative spatial eigenfunctions were stepwise selected. Soil functional indicators comprised standardized basal respiration and microbial biomass. For predictive modelling, we derived four decomposition metrics from tea mass-loss trajectories: the mean monthly loss rates over 0–3, 3–12 and 12–24 months, and the normalized 0–24 month area under the mass-loss curve (AUC), averaged across green and rooibos tea. To avoid artefactual declines across harvests, mass loss within each tea type was constrained to be non-decreasing before metric calculation. Baseline models included environmental and soil functional predictors; CAZyme-augmented models additionally included stage-total CAZyme abundance estimates for the broad, initial, mid and last decomposition stages. Predictor preprocessing was fitted within each training set to avoid information leakage: environmental variables were winsorized and standardized, whereas CAZyme totals were log1p-transformed and standardized; for elastic-net models, the CAZyme block was further reduced by principal component analysis. We fitted gaussian elastic-net and random-forest models[78,79] and evaluated them using out-of-sample R^2 and RMSE under four validation regimes: repeated within-ecosystem K-fold cross-validation, leave-one-ecosystem-out, leave-one-continent-out, and training on European samples with testing on non-European samples. Within-ecosystem and leave-one-ecosystem-out analyses used the European dataset, whereas continent-level transfer analyses used the full global dataset.

Declarations

Data availability

Metagenomic reads generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) under accession code PRJNA1045969.

Code availability

The code used for statistical analysis and figure generation, together with the data required to reproduce the reported analyses and figures, is available in Zenodo at 10.5281/zenodo.19484720

Competing interests

The authors declare no competing interests.

Author Contributions

K.Kü., B.S., Z.H., C.-E.W, and A.H.-B. conceived the research and developed the methodological framework. K.Kü., A.H.-B., C.-E.W., F.B. and N.E. designed the study and planned the analytical and experimental approach. B.S., Q.B., and N.K.C. performed the bioinformatic analyses. B.S., Z.H., and G.P.

performed the data curation. B.S. performed the data analyses. B.S. wrote the first draft of the manuscript, and all authors contributed to reviewing and editing the final manuscript. K.Kü., C.-E.W., F.B., A.H.-B., N.E., I.D., A.Le., A.Lo., M.C.R., S.Sc., and G.P. conducted laboratory analyses and data compilation. Site coordination, soil sampling, and tea bag decomposition experiments were carried out by C.Ad., B.A., C.An., A.A., M.B., C.Be., B.B., A.B., C.Br., H.B., R.C., M.C., R.C.G., C.C., C.T.C., M.Da., E.A.D., C.D., V.D.C., L.D.M., J.D., M.Di., S.D., T.E., I.F., H.F., O.F., N.F., V.F., P.F., M.Frei., M.Fren., K.F., R.G.G., A.G., R.G., S.G., M.G., G.G., K.G., P.H., F.H., U.H., R.H., T.His., T.Hiu., K.H., H.Ja., J.J.J., H.Jä., R.K., S.K.-R., T.K., Y.K., Z.K., K.Kr., H.K., K.L., K.S.L., J.Lo., J.Lö., S.L., V.M., F.M.-D., H.M., I.M., J.M., U.M.d.C., I.O., A.Pa., W.P., P.P., A.Pe., V.P., Q.P., M.-N.P., A.Pr., J.-L.P., M.P., T.R., P.R., M.S., A.Sc., I.K.S., J.Se., S.Se., H.C.S., A.I.S., J.St., A.Sta., A.Ste., F.S., J.-P.T., K.T., N.T., M.T., S.T.-T., S.T., P.D.T., T.U., M.V., L.v.d.B., H.v.d.M., I.v.H., K.Va., S.V., K.Ve., A.V., J.V., P.V., M.Wa., M.We., L.J.W., G.W., D.W., F.Z., M.Z.

Acknowledgements

We thank all the people who helped in sampling and experiment processes and volunteering researchers from the networks of ILTER (https://www.ilter.network), eLTER (https://elter-ri.eu), iSBio (https://home.uni-leipzig.de/idiv/isbio) and TeaComposition initiative (https://www.teacomposition.org). We would like to thank Uta Nuesse-Hahne, Susanna Boening-Klein and Anja Zeuner for their help in laboratory analyses. We acknowledge the support of Bart Muys, Thomas Spiegelberger, Hideaki Shiabta, Mioko Ataka, André-Jean Francez and Amélie Saillard for facilitating filed work. We also thank all the landowners for the access to their lands. We acknowledge funding to K.Kü, BS from the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena- Leipzig of the German Research Foundation (FZT118). Sequencing was financed by the German Research Foundation (DFG, BU 941/32 – 1 | EI 862/27 – 1 | HE 8266/4 – 1 | KU 1367/14 – 1 | WE 6579/2 – 1 to F.B., N.E., A.H.-B., K.Kü., C.-E.W.). Sequence data analysis was performed at the High-Performance Computing (HPC) Cluster EVE, a joint effort of both the Helmholtz Centre for Environmental Research-UFZ and the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, whose administrators are thanked for excellent support. ZRH acknowledges financial support received through the SFB1127 ChemBioSys (project number 239748522), which is funded by the Deutsche Forschungsgemeinschaft which was awarded to KKü and CEW. G.W. acknowledges partial financial support by the Austrian Science Fund (FWF) through grants 10.55776/P31669, 10.55776/P35737 and 10.55776/PAT2661823. The FORBIO-Gedinne experiment is maintained with partial support by the Walloon forest service (SPW – DNF) in the frame of the 5-yr research programme ‘Plan quinquennal de recherche et de vulgarisation forestières’. H.J. and F.Z. acknowledge the Galapagos National Park Directorate for export sample permit Nr. 006-2021-EXP-CM-FAU-DBI/MAAE under genetic permit Nr. MAE-DNB-CM-2016-0043. A.S. acknowledges partial support by the following the National Science Foundation ASCEND Biology Integration Institute (NSF-DBI-2021898) and the US Department of Energy, Office of Science, and Office of Biological and Environmental Research award number DE-FG02-07ER64456; Minnesota Agricultural Experiment Station MN-42-030 and MN-42-060; and the College of Food, Agricultural and Natural Resources Sciences and Wilderness Research Foundation, University of Minnesota. LvdB, RC and MaB thank CONAF and the park

rangers from Contulmo, Nahuelbuta, La Campana and Pan de Azúcar national parks, and the people from Quebrada de Talca private reserve for their collaborations on-site, as well as all the students that helped out, and the German Research Foundation (DFG) priority research program SPP-1803 “EarthShape: Earth Surface Shaping by Biota” (TI 338/14 − 1&2, BA 3843/6 − 1). H.K. and Ka.H. acknowledges the support by the JSPS KAKENHI (26281049, 17H03835, 20H04380 to H.K.; 19H02999, 23K02256 to Ka.H.). We would like to thank the Research Center for Alpine Ecosystems (CREA Mont-Blanc) and the French national nature reserve of Ristolas - Mont-Viso for their work in the field. S.L. was supported by the Integrated Monitoring Programme contract 221-22-006 financed by the Swedish Environmental Protection Agency. PF was supported by the Swedish Research Council FORMAS (2020–01100) and the Swedish Research Council (2024–04821). ST acknowledges the management team of the BEF-China research platform for their support. MB acknowledges financial support by the Slovak Scientific Grant Agency VEGA (Grant No. 2/0088/26). RK acknowledges financial support by the Slovak Scientific Grant Agency VEGA (Grant No. 2/0074/25). G.G. was supported by the grants DEB 0620910, 1831952 and 2425484 from the National Science Foundation as part of the Luquillo LTER. The U.S. Forest Service (Department of Agriculture) and University of Puerto Rico gave additional support. A.G thanks CONAF for permission to work in Alerce Andino (CONAF 340688) and to Estación Biológica Senda Darwin, and PIA FB-210006 and FONDECYT 1201643. CB, HCS, and MS-R, acknowledge FCT support to cE3c through UID/BIA/00329/2013, UID/BIA/00329/2019, UIDB/00329/2020, and UID/00329/2025 (https://doi.org/10.54499/UID/00329/2025). KK and IO were supported by the Estonian Research Council grants PRG916, PRG2614 and the CoE FutureScapes (TK232U3). AIS acknowledges FCT – Fundação para a Ciência e a Tecnologia I.P., for support to CESAM-Centro de Estudos do Ambiente e do Mar, under the projects UID/50017/2025 (doi.org/10.54499/UID/50017/2025) and LA/P/0094/2020 (doi.org/10.54499/LA/P/0094/2020) and PORBIOTA— Portuguese E-Infrastructure for Information and Research on Biodiversity (POCI-01-0145-FEDER-022127), financed by FCT through PIDAC national funds and co-funded by the FEDER. TH was partly supported by the grants 21H05316, 21H02226 from JSPS and Forest GX project from SMBC. PDT acknowledges financial support from the RECOVER project (PN-IV-P6-6.1-CoEx-2024-0124). Zs.K. was supported by the Research Excellence Programme of the Hungarian University of Agriculture and Life Sciences. M.D. receives financial support from the Swiss Federal Office of the Environment. The work of UH and M Frei has been partly funded within the German Research Foundation (DFG) Priority Program 1374 ‘Infrastructure-Biodiversity-Exploratories’ (HA 4597/6 − 3).

References

1. Aerts, R., Climate: leaf litter chemistry and leaf litter decomposition in terrestrial ecosystems: a triangular relationship. *Oikos* (1997). https://doi.org/10.2307/3546886 doi:10.2307/3546886
2. Manzoni, S.: Analytical models of soil and litter decomposition: solutions for mass loss and time-dependent decay rates. Soil. Biol. Biochem. (2012). https://doi.org/10.1016/j.soilbio.2012.02.029
3. Djukic, I., et al.: Early stage litter decomposition across biomes. Sci. Total Environ. **628–629**, 1369–1394 (2018)

4. Wu, Q., et al.: Substrate and climate determine terrestrial litter decomposition. *Proceedings of the National Academy of Sciences* 122, e2420664122 (2025)
5. Keuskamp, J.: Tea Bag Index: a novel approach to collect uniform decomposition data across ecosystems. Methods Ecol. Evol. (2013). https://doi.org/10.1111/2041-210X.12097
6. Allison, S.D., Martiny, J.B.H.: Resistance, resilience, and redundancy in microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* (2008). https://doi.org/10.1073/pnas.0801925105 doi:10.1073/pnas.0801925105
7. Schimel, J.P., Schaeffer, S.M.: Microbial control over carbon cycling in soil. Front. Microbiol. (2012). https://doi.org/10.3389/fmicb.2012.00348
8. Boerjan, W.: Lignin biosynthesis. Annu. Rev. Plant. Biol. (2003). https://doi.org/10.1146/annurev.arplant.54.031902.134938
9. Xu, J., Zhang, H., Zheng, J., Dovoedo, P., Yin, Y.: eCAMI: simultaneous classification and motif identification for enzyme annotation. Bioinformatics. **36**, 2068–2075 (2020)
10. Lombard, V.: The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. (2014). https://doi.org/10.1093/nar/gkt1178
11. Joly, F.-X., Scherer-Lorenzen, M., Hättenschwiler, S.: Resolving the intricate role of climate in litter decomposition. Nat. Ecol. Evol. **7**, 214–223 (2023)
12. Chen, H.: Functional Redundancy in Soil Microbial Community Based on Metagenomics Across the Globe. Front. Microbiol. (2022). https://doi.org/10.3389/fmicb.2022.878978
13. Louca, S.: Function and functional redundancy in microbial systems. Nat. Ecol. Evol. (2018). https://doi.org/10.1038/s41559-018-0519-1
14. Cha, X., et al.: Metagenomic insights into the carbon decomposition of plant and microbial biomass in forests across biomes. Soil. Ecol. Lett. **7**, 250300 (2025)
15. Sarneel, J.M., et al.: Reading tea leaves worldwide: Decoupled drivers of initial litter decomposition mass-loss rate and stabilization. Ecol. Lett. **27**, (2024)
16. Bradford, M.A., et al.: A test of the hierarchical model of litter decomposition. Nat. Ecol. Evol. **1**, 1836–1845 (2017)
17. Glassman, S.I., et al.: Decomposition responses to climate depend on microbial community composition. *Proceedings of the National Academy of Sciences* 115, 11994–11999 (2018)
18. Drula, E.: The carbohydrate-active enzyme database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. (2022). https://doi.org/10.1093/nar/gkab1045
19. Zheng, J., et al.: dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res. **51**, W115–W121 (2023)
20. Berg, B., McClaugherty, C.: *Plant Litter: Decomposition, Humus Formation, Carbon Sequestration*. *Springer (book)* (2020). 10.1007/978-3-030-59631-6
21. Augusto, L., et al.: Influences of evergreen gymnosperm and deciduous angiosperm tree species on the functioning of temperate and boreal forests. Biol. Rev. **90**, 444–466 (2015)

22. Freschet, G.T., et al.: Linking litter decomposition of above-and below-ground organs to plant–soil feedbacks worldwide. J. Ecol. **101**, 943–952 (2013)
23. Didion, M., et al.: Towards harmonizing leaf litter decomposition studies using standard tea bags—a field study and model application. Forests. **7**, 167 (2016)
24. Bao, Y., et al.: Functional redundancy across space and time in litter-degrading fungal communities. J. Sustainable Agric. Environ. **3**, e12086 (2024)
25. Louca, S., Parfrey, L.W., Doebeli, M.: Decoupling function and taxonomy in the global ocean microbiome. Sci. (1979). **353**, 1272–1277 (2016)
26. Sunagawa, S.: Structure and function of the global ocean microbiome. *Science ()*. (1979). https://doi.org/10.1126/science.1261359 (2015) doi:10.1126/science.1261359
27. Taubert, M., Stähly, J., Kolb, S., Küsel, K.: Divergent microbial communities in groundwater and overlying soils exhibit functional redundancy for plant-polysaccharide degradation. PLoS One. **14**, e0212937 (2019)
28. Ramond, P., Galand, P.E., Logares, R.: Microbial functional diversity and redundancy: moving forward. FEMS Microbiol. Rev. **49**, fuae031 (2025)
29. Fierer, N.: Embracing the unknown: disentangling the complexities of the soil microbiome. Nat. Rev. Microbiol. (2017). https://doi.org/10.1038/nrmicro.2017.87
30. López-Mondéjar, R., et al.: Metagenomics and stable isotope probing reveal the complementary contribution of fungal and bacterial communities in the recycling of dead biomass in forest soil. Soil. Biol. Biochem. **148**, 107875 (2020)
31. Burns, R.G., et al.: Soil enzymes in a changing environment: current knowledge and future directions. Soil. Biol. Biochem. **58**, 216–234 (2013)
32. Fanin, N., et al.: Soil enzymes in response to climate warming: Mechanisms and feedbacks. Funct. Ecol. **36**, 1378–1395 (2022)
33. Rocca, J.D., et al.: Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed. ISME J. **9**, 1693–1699 (2015)
34. Graham, E.: Microbes as Engines of Ecosystem Function: When Does Community Structure Enhance Predictions of Ecosystem Processes? Front. Microbiol. (2016). https://doi.org/10.3389/fmicb.2016.00214
35. Keiser, A.D., Bradford, M.A.: Climate masks decomposer influence in a cross-site litter decomposition study. Soil. Biol. Biochem. (2017). https://doi.org/10.1016/j.soilbio.2016.12.022
36. Adair, E.C., et al.: Simple three-pool model accurately describes patterns of long-term litter decomposition in diverse climates. Glob Chang. Biol. **14**, 2636–2660 (2008)
37. Wieder, W.R., Grandy, A.S., Kallenbach, C.M., Bonan, G.B.: Integrating microbial physiology and physio-chemical principles in soils with the MIcrobial-MIneral Carbon Stabilization (MIMICS) model. *Biogeosciences* 11, 3899–3917 (2014)

38. Prosser, J.I.: Dispersing misconceptions and identifying opportunities for the use of'omics' in soil microbial ecology. Nat. Rev. Microbiol. **13**, 439–446 (2015)
39. Blazewicz, S.: Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. ISME J. (2013). https://doi.org/10.1038/ismej.2013.102
40. Walter, H., Breckle, S.W.: Ecological Systems of the Geobiosphere: 2 Tropical and Subtropical Zonobiomes, vol. 2. Springer Science & Business Media (2013)
41. Heintz-Buschart, A., et al.: Microbial diversity-ecosystem function relationships across environmental gradients. Res. Ideas Outcomes. **6**, e52217 (2020)
42. Lu, Y., et al.: Landscape effects on global soil pathogenic fungal diversity across spatial scales. Nat. Commun. (2026)
43. Agency, E.S.: *Land Cover CCI Product User Guide Version 2.0*. (2017). https://climate.esa.int/media/documents/CCI_Land_Cover_PUG_v2.0.pdf
44. IUCN: *Guidelines for the Application of IUCN Red List of Ecosystems Categories and Criteria, Version 2.0*. https://portals.iucn.org/library/sites/library/files/documents/2024-021-En.pdf (2024). 10.2305/CJDF9122
45. Diversity, C.: on B. *Convention on Biological Diversity: Article 2. Use of Terms*. (1992). https://www.cbd.int/convention/articles/default.shtml?a=cbd-02
46. Fick, S.E., Hijmans, R.J.: WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. (2017). https://doi.org/10.1002/joc.5086
47. Karger, D.N., et al.: CHELSA climatologies at high resolution for the earth's land surface areas (Version 1.0). (2016)
48. Kemper, W.D., Rosenau, R.C.: Aggregate stability and size distribution. Methods soil. analysis: Part. 1 Phys. mineralogical methods. **5**, 425–442 (1986)
49. Eisenhauer, N.: Plant diversity effects on soil microorganisms support the singular hypothesis. *Ecology* (2010). https://doi.org/10.1890/08-2338.1 doi:10.1890/08-2338.1
50. Scheu, S.: Automated measurement of the respiratory response of soil microcompartments: active microbial biomass in earthworm faeces. Soil. Biol. Biochem. **24**, 1113–1118 (1992)
51. Anderson, J.P.E., Domsch, K.H.: A physiological method for the quantitative measurement of microbial biomass in soils. Soil. Biol. Biochem. **10**, 215–221 (1978)
52. Beck, T., et al.: An inter-laboratory comparison of ten different ways of measuring soil microbial biomass C. Soil. Biol. Biochem. **29**, 1023–1032 (1997)
53. Uritskiy, G.V., DiRuggiero, J., Taylor, J.: MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6, (2018)
54. Li, D.: MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* (2015). https://doi.org/10.1093/bioinformatics/btv033 doi:10.1093/bioinformatics/btv033

55. Wood, D.: Improved metagenomic analysis with Kraken 2. Genome Biol. (2019). https://doi.org/10.1186/s13059-019-1891-0
56. Hyatt, D.: Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. (2010). https://doi.org/10.1186/1471-2105-11-119
57. Levy Karin, E., MetaEuk—sensitive: high-throughput gene discovery and annotation for large-scale eukaryotic metagenomics. *Microbiome* (2020). https://doi.org/10.1186/s40168-020-00808-x doi:10.1186/s40168-020-00808-x
58. Suzek, B.: UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* (2015). https://doi.org/10.1093/bioinformatics/btu739 doi:10.1093/bioinformatics/btu739
59. Steinegger, M., Söding, J.: MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat. Biotechnol. (2017). https://doi.org/10.1038/nbt.3988
60. Patro, R.: Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods. (2017). https://doi.org/10.1038/nmeth.4197
61. Eddy, S.R., Accelerated Profile, H.M.M., Searches: PLoS Comput. Biol. (2011). https://doi.org/10.1371/journal.pcbi.1002195
62. Buchfink, B.: Fast and sensitive protein alignment using DIAMOND. Nat. Methods. (2015). https://doi.org/10.1038/nmeth.3176
63. Zheng, J., et al.: dbCAN-seq update: CAZyme gene clusters and substrates in microbiomes. Nucleic Acids Res. **51**, D557–D563 (2023)
64. Consortium, C.: Ten years of CAZypedia: a living encyclopedia of carbohydrate-active enzymes. Glycobiology. **28**, 3–8 (2018)
65. Caspi, R.: The MetaCyc database of metabolic pathways and enzymes. Nucleic Acids Res. (2020). https://doi.org/10.1093/nar/gkz862
66. Edgar, R.C.: Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* (2010). https://doi.org/10.1093/bioinformatics/btq461 doi:10.1093/bioinformatics/btq461
67. Katoh, K., Standley, D.M.: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. (2013). https://doi.org/10.1093/molbev/mst010
68. Price, M.: FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS One. (2010). https://doi.org/10.1371/journal.pone.0009490
69. Murrell, B.F.U.B.A.R.: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. Mol. Biol. Evol. (2013). https://doi.org/10.1093/molbev/mst030
70. Team, R.C.: R: A Language and Environment for Statistical Computing. Preprint at (2024). https://www.R-project.org/
71. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc.: Ser. B (Methodol.). (1995). https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

72. Kassambara, A.: Pipe-Friendly Framework for Basic Statistical Tests (R package). *CRAN* Preprint at (2023). https://CRAN.R-project.org/package=rstatix
73. Soneson, C.: Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* (2015). https://doi.org/10.12688/f1000research.7563.1 doi:10.12688/f1000research.7563.1
74. Love, M.: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. (2014). https://doi.org/10.1186/s13059-014-0550-8
75. Faith, D.P.: Conservation evaluation and phylogenetic diversity. Biol. Conserv. (1992). https://doi.org/10.1016/0006-3207(92)91201-3
76. Kembel, S., Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* (2010). https://doi.org/10.1093/bioinformatics/btq166 doi:10.1093/bioinformatics/btq166
77. Oksanen, J.: vegan: Community Ecology Package (R package). *CRAN* Preprint at (2017). https://CRAN.R-project.org/package=vegan
78. Breiman, L., Random: forests. *Mach. Learn.* (2001). https://doi.org/10.1023/A:1010933404324 doi:10.1023/A:1010933404324
79. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. J. R Stat. Soc. Ser. B Stat. Methodol. (2005)

Figures

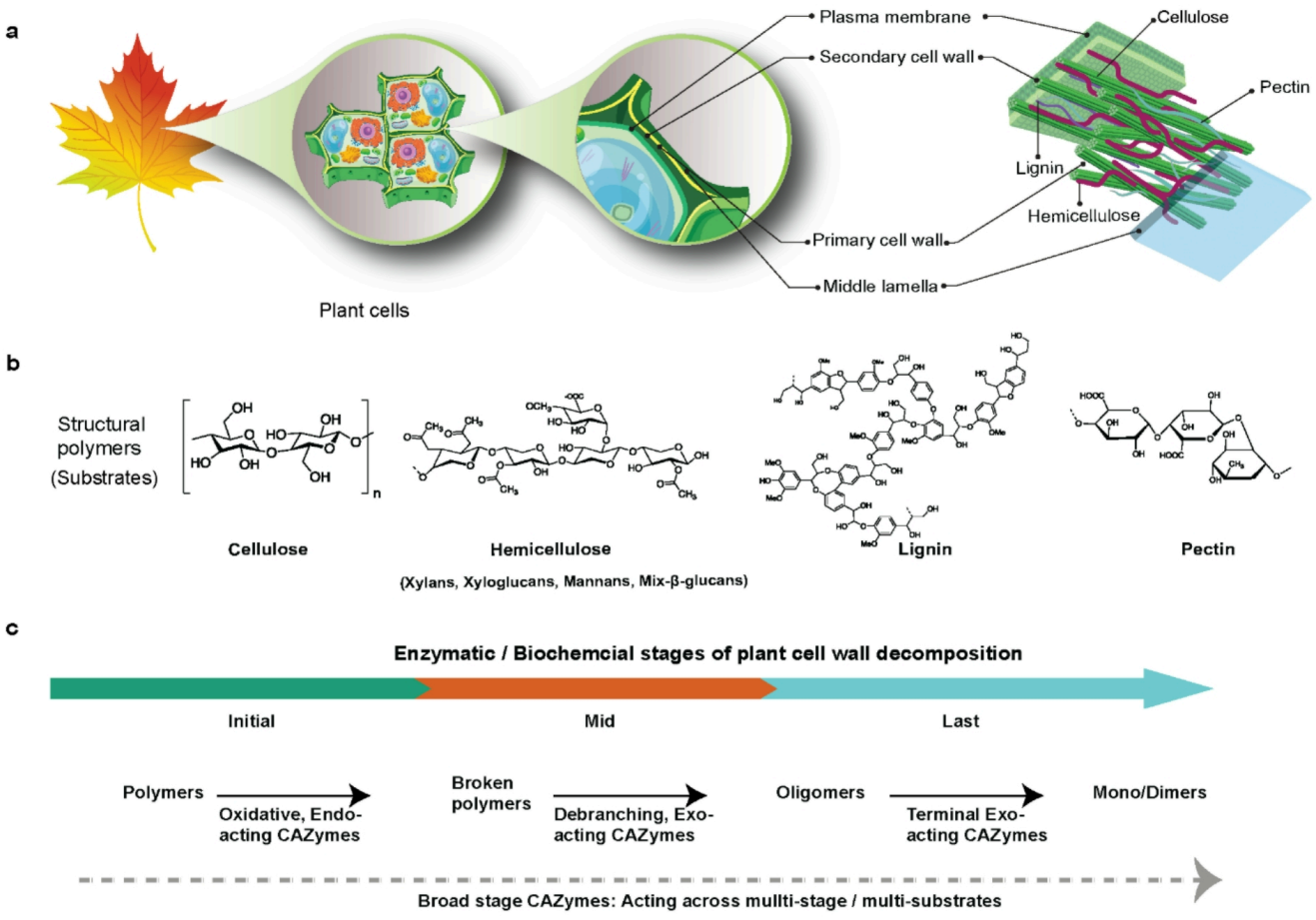


Figure 1

Conceptual overview of plant cell-wall components and biochemical stages of enzymatic decomposition.

(a) Schematic of plant leaf cells and the organization of the primary and secondary cell walls, middle lamella and plasma membrane, highlighting the interwoven matrix of cellulose, hemicellulose, lignin and pectin.

(b) Major structural polymers of the plant cell wall that serve as decomposition substrates: cellulose, hemicelluloses (including xylans, xyloglucans, mannans and mixed-linkage β-glucans), lignin and pectin.

(c) Biochemical stages of plant cell-wall polymer decomposition, from initial attack on intact polymers to the production of broken polymers, oligomers and ultimately mono- and dimers. Predominant CAZyme activities are indicated for each stage (oxidative and endo-acting CAZymes in the initial stage; debranching and exo-acting CAZymes in the mid stage; terminal exo-acting CAZymes in the last stage), along with a broad group of CAZymes with multi-substrate or multi-stage activity (grey dashed arrow).

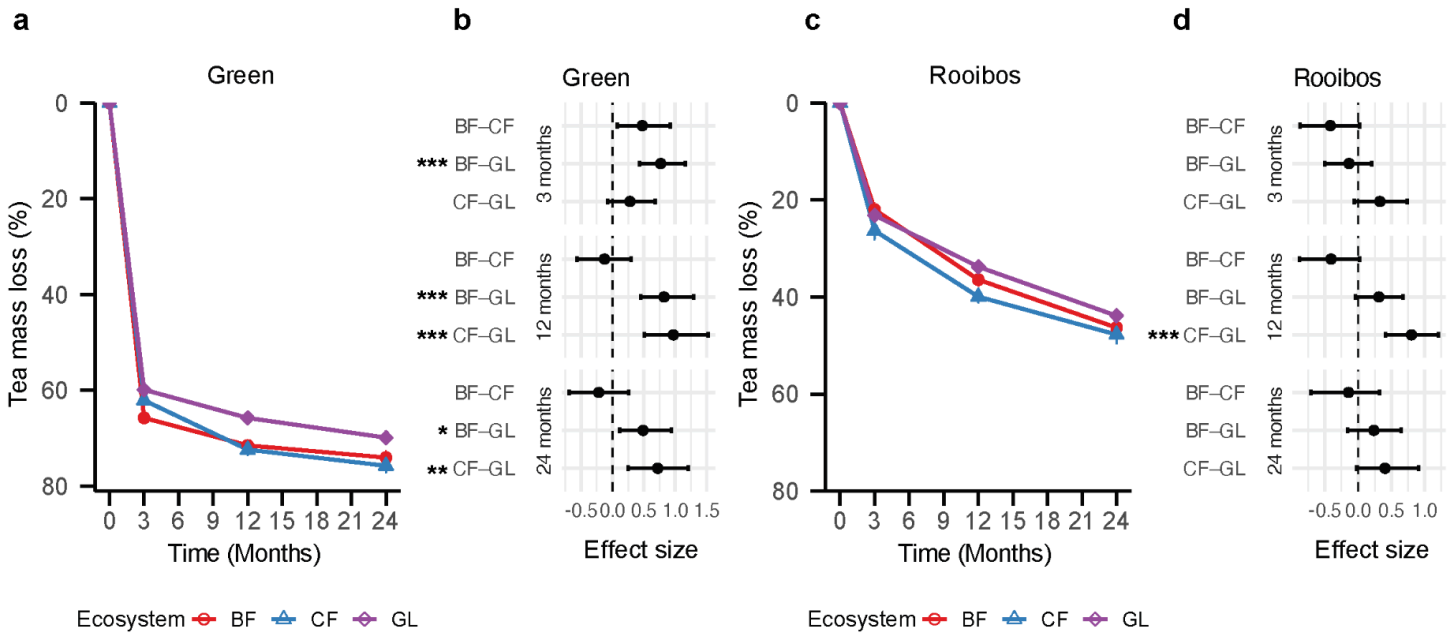


Figure 2

Temporal dynamics of tea litter mass loss and ecosystem contrasts.

(a),(c) Mean mass loss (%) of standardized green (a) and rooibos (c) tea bags after 3, 12 and 24 months of incubation in broadleaf forest (BF), coniferous forest (CF) and grassland (GL) ecosystems (primary European dataset). Lines connect ecosystem means; error bars show ±1 s.e. (b),(d) Standardized effect sizes for pairwise differences in mass loss between ecosystem types (BF–CF, BF–GL and CF–GL) for green (b) and rooibos (d) tea at each time point; horizontal bars show the associated uncertainty (see Methods), and the dashed line indicates no difference (effect size = 0). Asterisks denote significant contrasts ($P < 0.05$, $P < 0.01$, $P < 0.001$); non-significant contrasts are not labelled.

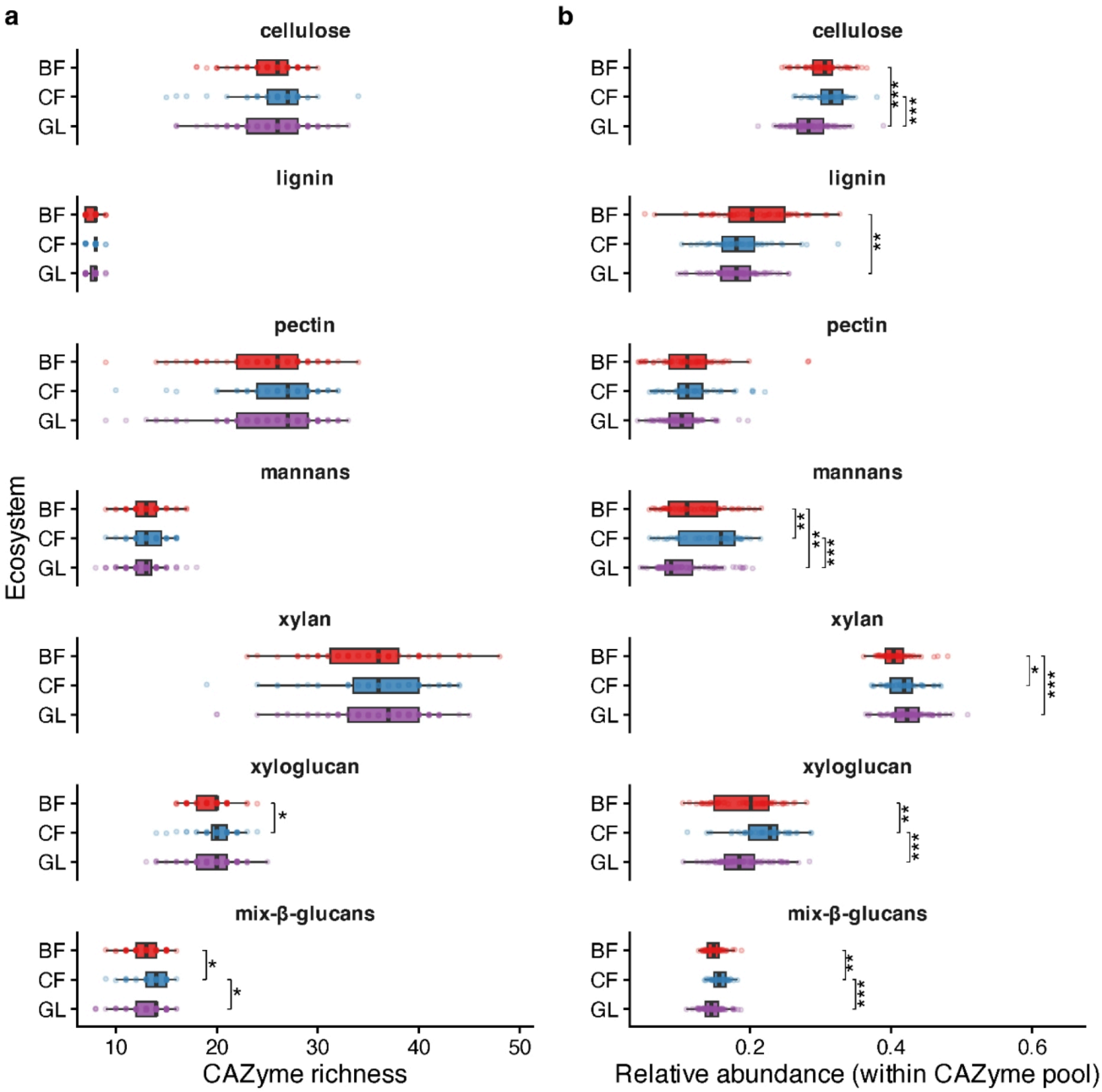


Figure 3

Substrate-associated CAZyme richness and relative abundance across ecosystems.

(a) CAZyme family richness for families assigned to plant cell-wall substrate categories (cellulose, lignin, pectin, mannans, xylan, xyloglucan and mixed-linkage β-glucans) in broadleaf forest (BF), coniferous forest (CF) and grassland (GL) ecosystems (primary European dataset). (b) Relative abundance of the same substrate-associated CAZyme families, expressed as relative abundance within the CAZyme pool (TPM re-normalized within CAZymes for each sample). Boxes show the median and interquartile range,

with whiskers extending to 1.5× IQR. Brackets and asterisks indicate significant pairwise ecosystem differences based on Wilcoxon rank-sum tests with Benjamini–Hochberg correction (P < 0.05, P < 0.01, P < 0.001).

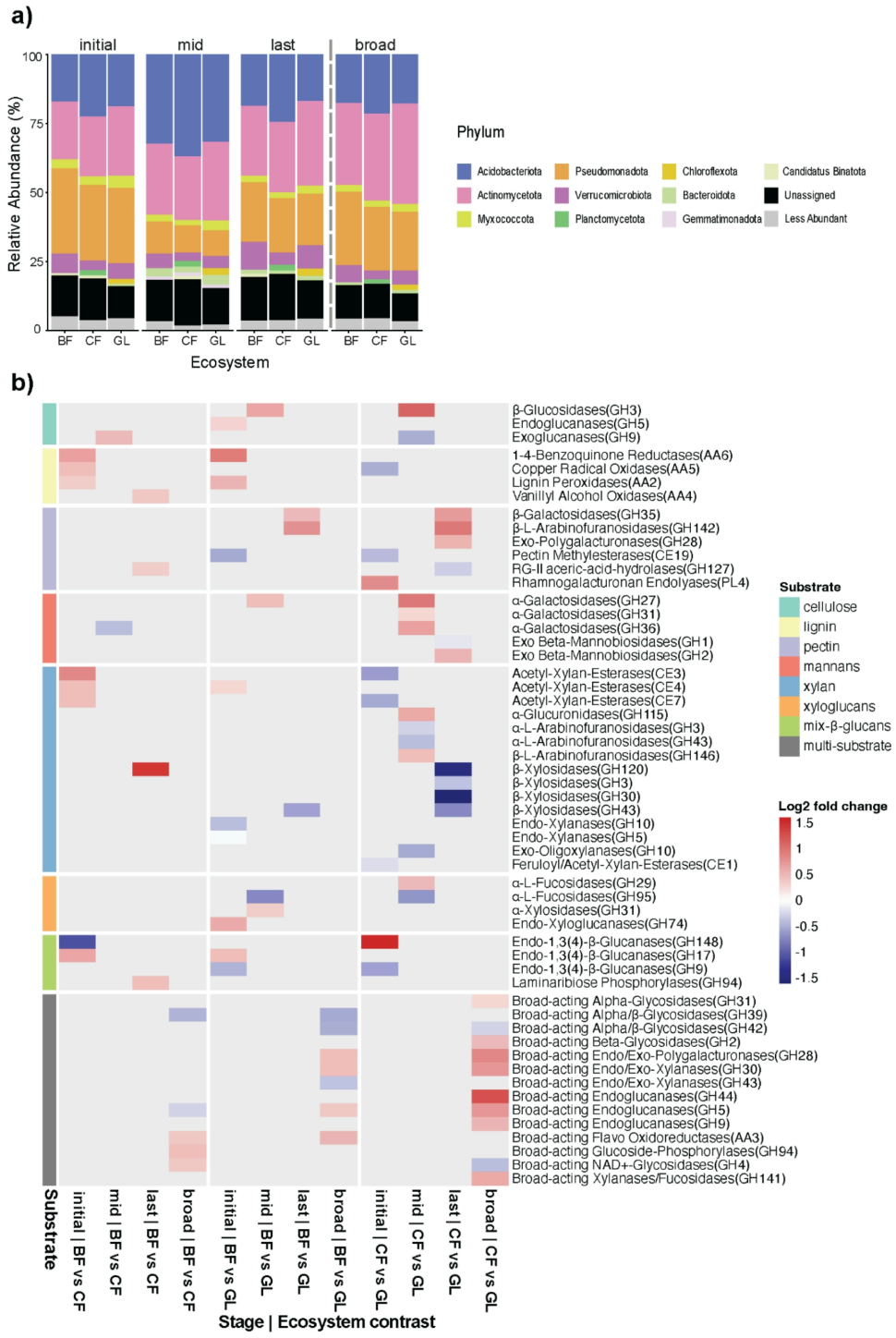


Figure 4

Taxonomic composition of stage-resolved CAZyme pools and stage-specific differential abundance of CAZyme families among ecosystems.

(a) Relative contribution of microbial phyla to CAZyme-annotated genes assigned to the initial, mid, last and broad stages in broadleaf forest (BF), coniferous forest (CF) and grassland (GL) ecosystems. Bars show proportional contributions within each stage–ecosystem combination; low-abundance phyla are grouped as “Less abundant”, and “Unassigned” denotes genes lacking phylum-level assignment.

(b) Heat map of CAZyme families showing significant differential abundance between ecosystems within each stage (DESeq2; adjusted $P < 0.05$). Columns show pairwise ecosystem contrasts (BF vs CF, BF vs GL and CF vs GL) within each stage; rows show CAZyme families (labels indicate family name/EC-style description and CAZy family) grouped by primary substrate class and annotated stage. Cell colours indicate shrunken log2 fold change (enrichment in the first ecosystem of each contrast is positive and enrichment in the second is negative). Grey cells indicate comparisons that were not significant or not applicable/not tested for that family–stage combination (for example, multi-substrate/broad families are only shown for the broad stage).

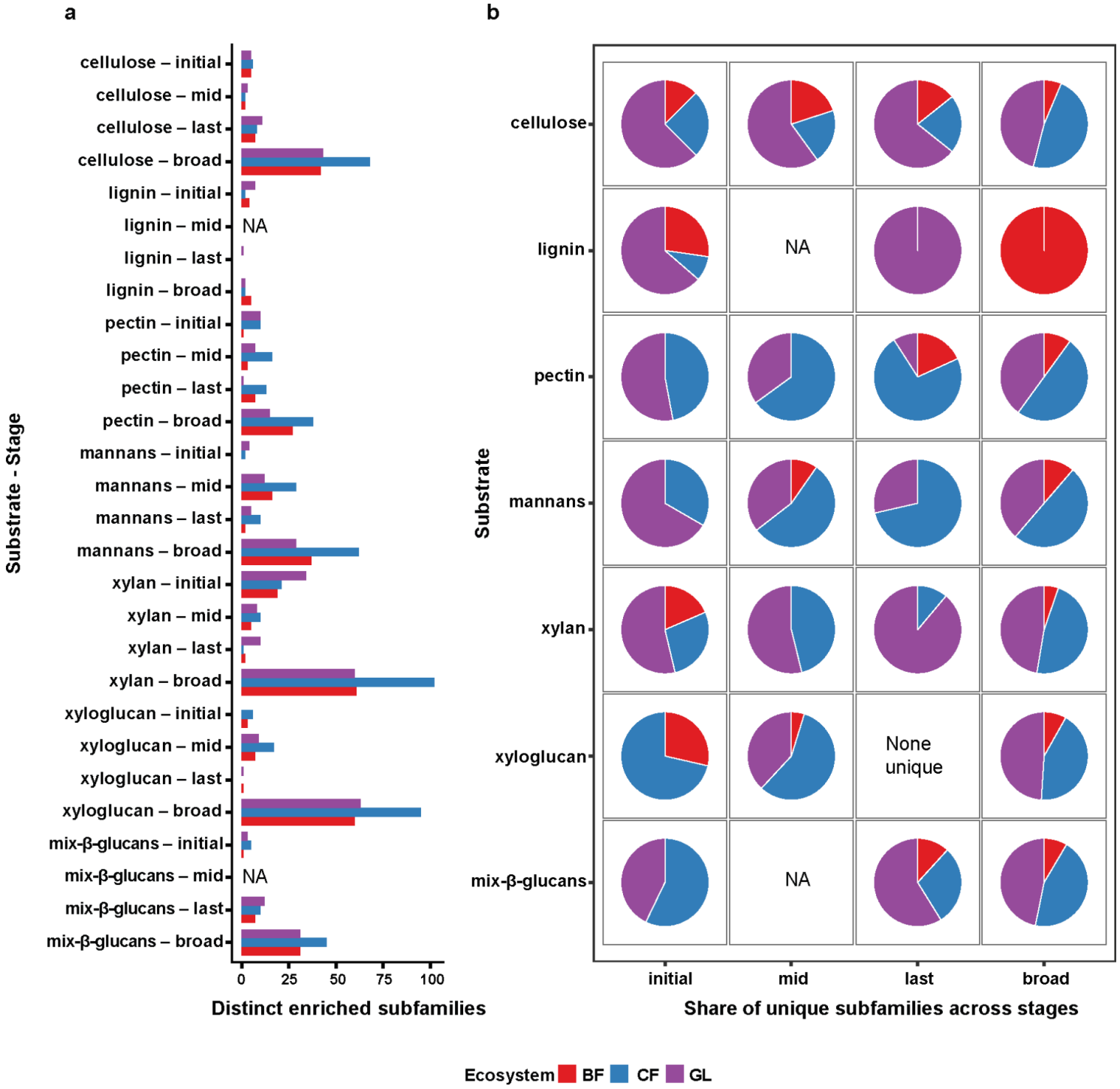


Figure 5

Differential enrichment of CAZyme subfamilies across substrates, decomposition stages and ecosystems.

(a) Number of distinct CAZyme subfamilies significantly enriched in broadleaf forests, coniferous forests and grasslands for each substrate–stage combination (initial, mid, last and broad). Counts treat each subfamily once per combination. Enrichment was defined as higher CAZyme relative abundance in one ecosystem compared with another (one-sided Wilcoxon tests with Benjamini–Hochberg correction; q <

0.05). NA indicates substrate–stage combinations not represented in the curated stage catalogue and therefore not tested.

(b) For each substrate–stage combination, pie charts show the fraction of ecosystem-unique enriched subfamilies (enriched in only one ecosystem and not in the other two), partitioned by ecosystem. None unique indicates combinations where enriched subfamilies were detected but none were ecosystem-specific.

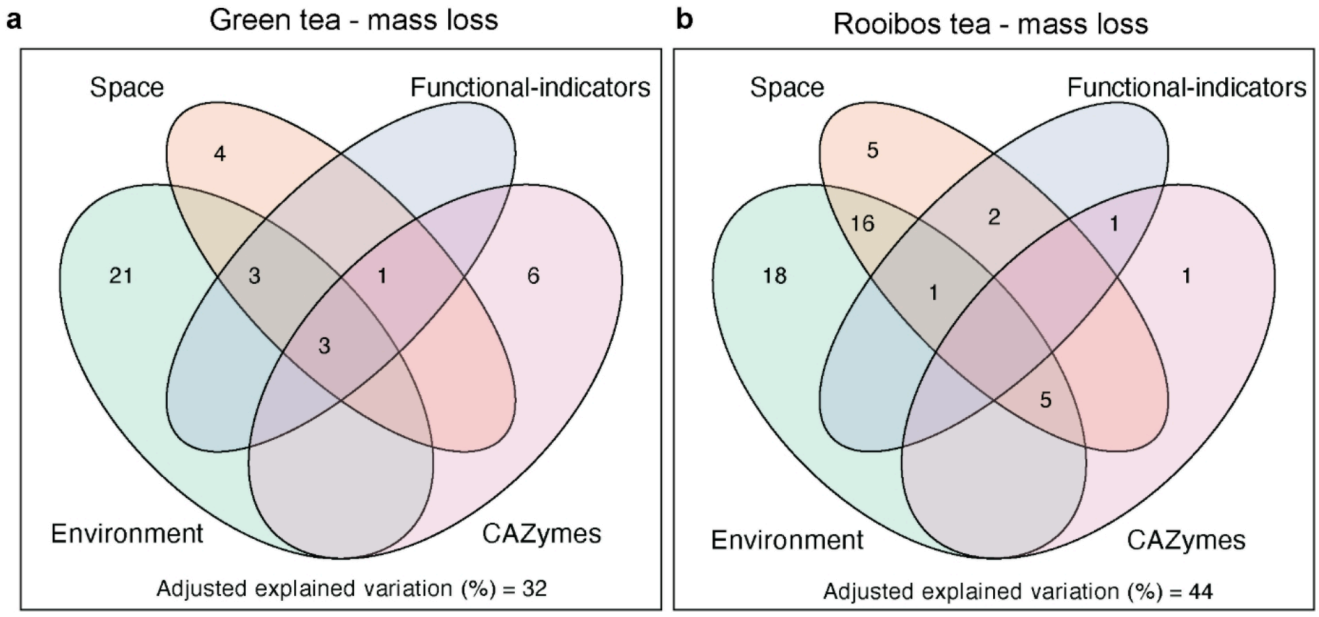


Figure 6

Variation partitioning of tea litter mass loss at 24 months.

Venn diagrams show the adjusted fraction of explained variation in 24-month tea litter mass loss partitioned among four predictor sets: environment (soil physicochemical and climatic variables), space (spatial structure modelled with dbMEM eigenfunctions derived from site coordinates), soil functional indicators (microbial respiration and microbial biomass), and CAZyme profiles. (a) Green tea. (b) Rooibos tea. Numbers denote the proportion of variation attributed uniquely to each predictor set or jointly to their overlaps. Values < 1 are not shown.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables17.xlsx

- SupplementaryTables1112.xlsx
- Supplementaryfigures.pdf
- SupplementaryTable8.xlsx
- SupplementaryTables910.xlsx