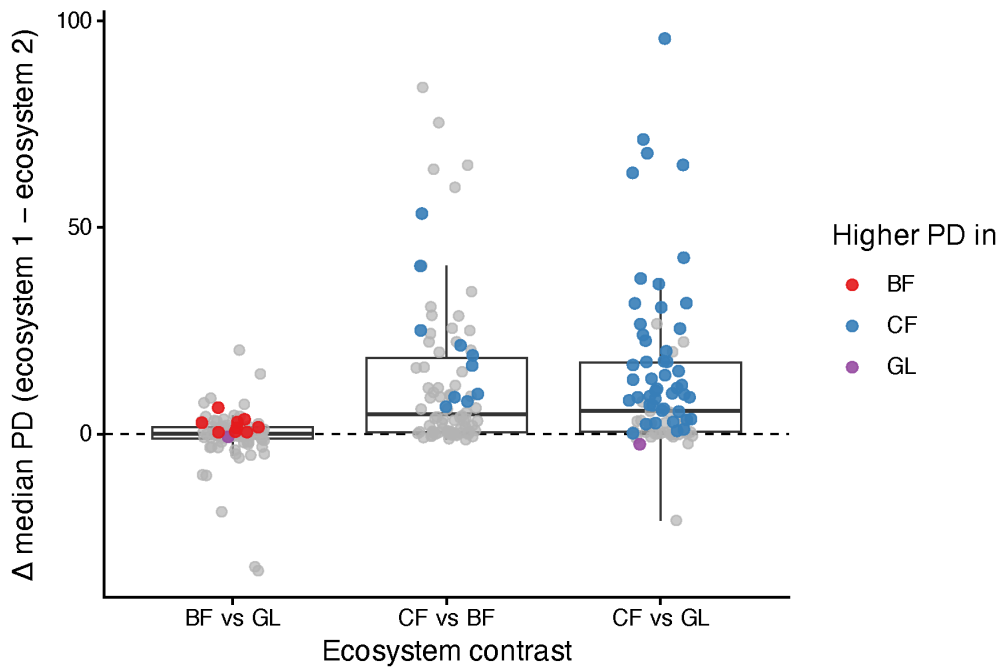


Supplementary Figures

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

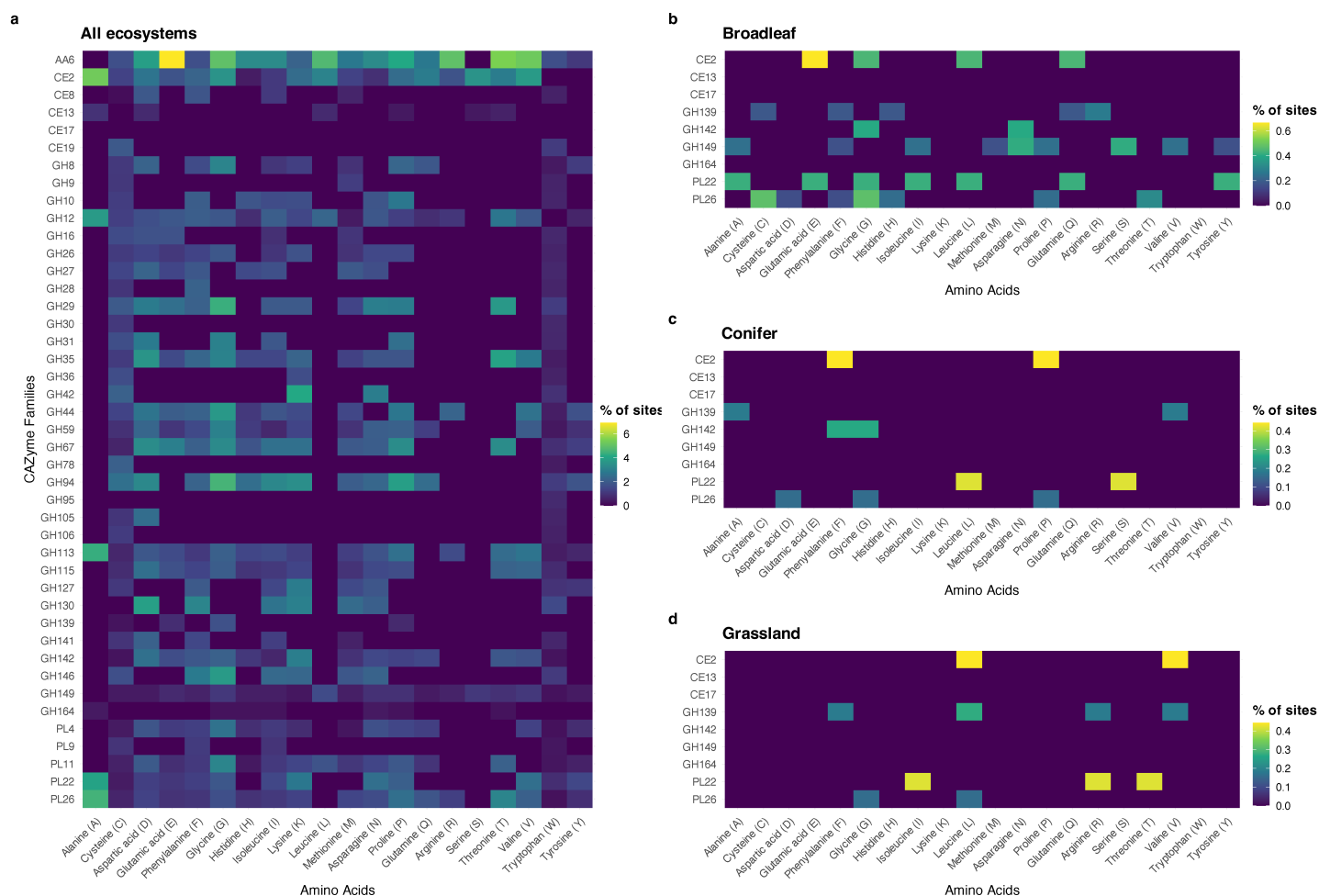


Supplementary Fig.1 | Ecosystem contrasts in CAZyme-family phylogenetic diversity.

The Δ median PD distribution across CAZyme families (one point per family), restricted to families represented in all three ecosystems is shown. Δ median PD is computed as first ecosystem minus second ecosystem; values < 0 indicate higher PD in the second ecosystem. Points are coloured by the ecosystem with higher median PD for that family. Dashed line indicates no difference ($\Delta = 0$). PD was computed as Faith's phylogenetic diversity from CAZyme sequence-cluster trees; pairwise ecosystem comparisons were tested with Wilcoxon tests and Benjamini–Hochberg correction.

Supplementary Figures

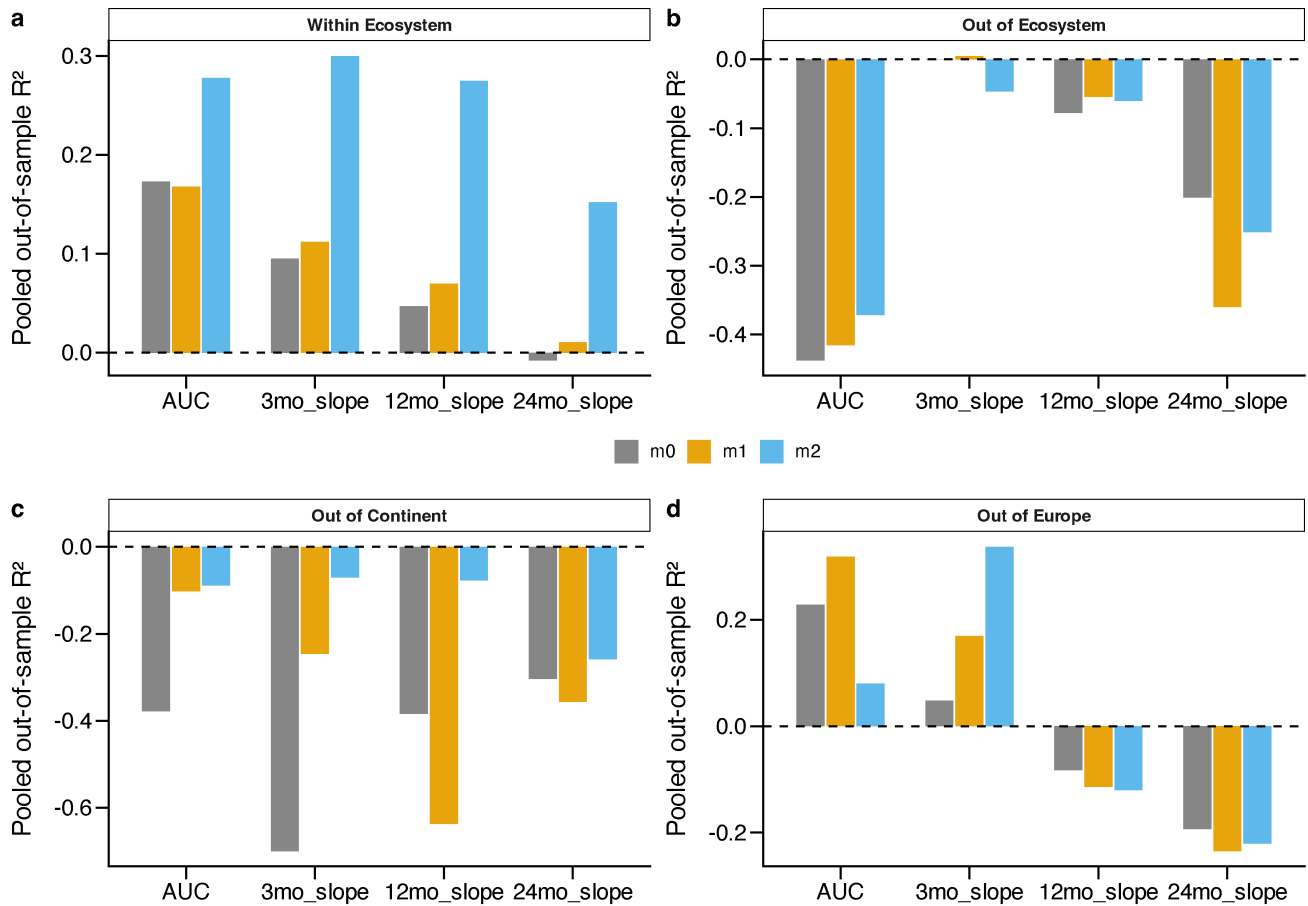
From shared genes to staged decay: ecosystem-specific genomic potential for litter decomposition in soils



Supplementary Fig.2 | Ecosystem-associated directional selection in CAZyme family protein sequences. Heatmaps summarise directional selection (FADE analysis, see methods) results for CAZyme family amino-acid alignments, reporting the percentage of alignment sites inferred to be under directional selection towards each amino acid (Bayes factor ≥ 100). (a) Pooled analysis across ecosystems for CAZyme families with ecosystem differences in phylogenetic diversity; rows are CAZyme families and columns are amino acids. (b–d) Ecosystem-targeted FADE fits for a representative subset of families, run separately for broadleaf forest (b), coniferous forest (c) and grassland (d). Colour intensity indicates the fraction of sites assigned to each amino acid within a family (scales shown at right); blank/low-intensity cells indicate that no sites in that family met the FADE threshold for that amino acid.

Supplementary Figures

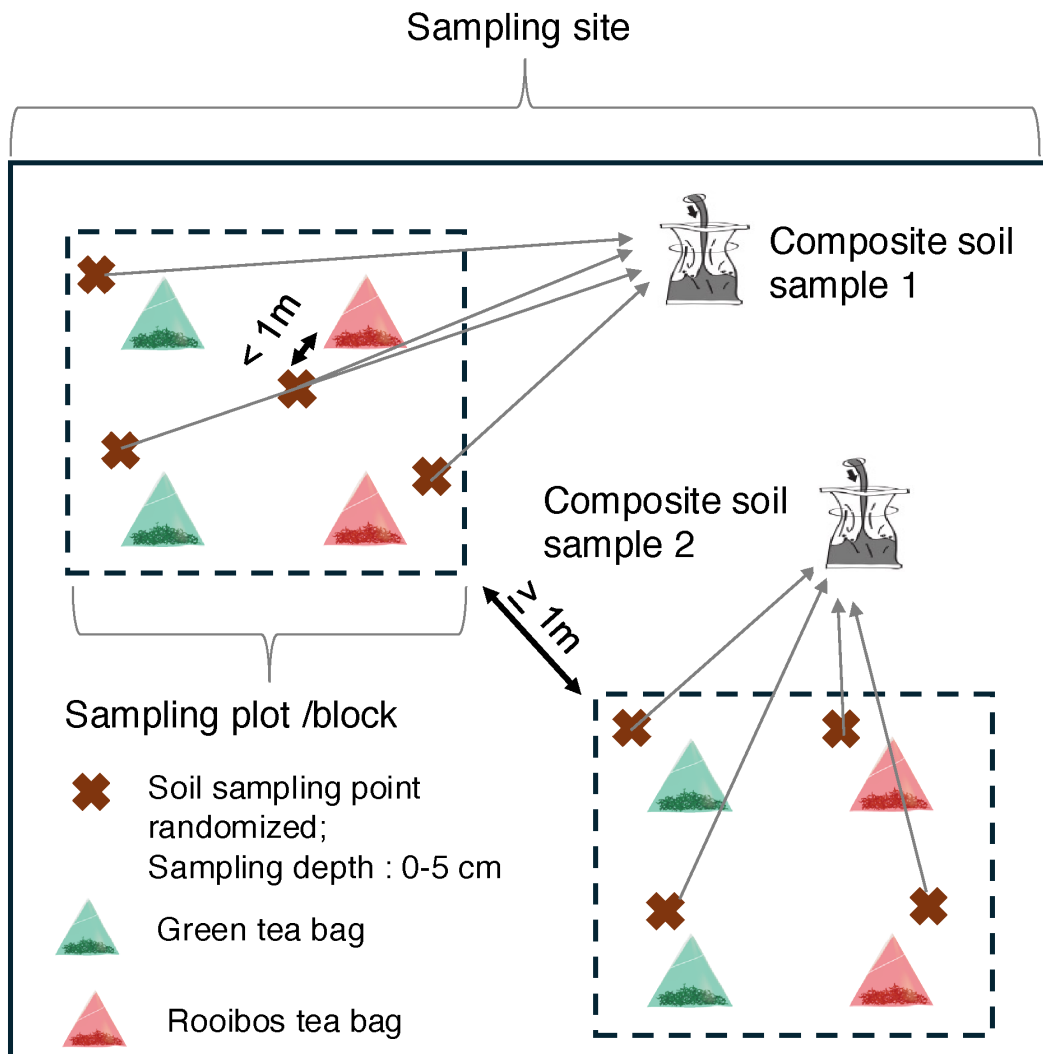
From shared genes to staged decay: ecosystem-specific genomic potential for litter decomposition in soils



Supplementary Fig.3 | Predictive performance of models for litter decomposition across geographic transfer regimes. Pooled out-of-sample R^2 for predicting tea decomposition metrics, AUC (area under the mass-loss curve) and decomposition phase-specific mass-loss slopes (3-month, 12-month and 24-month), under four validation regimes: (a) within ecosystem stratified K-fold cross-validation; (b) leave-one-ecosystem-out; (c) leave-one continent-out; and (d) out-of-Europe evaluation (models trained on European sites and tested on non-European sites). Bars show three model configurations: M0, environmental covariates and soil functional indicators only; M1, elastic-net regression with added CAZyme profiles; and M2, random-forest model with added CAZyme profiles. The dashed line indicates $R^2 = 0$; negative values indicate performance worse than predicting the mean of the training data for held-out observations.

Supplementary Figures

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

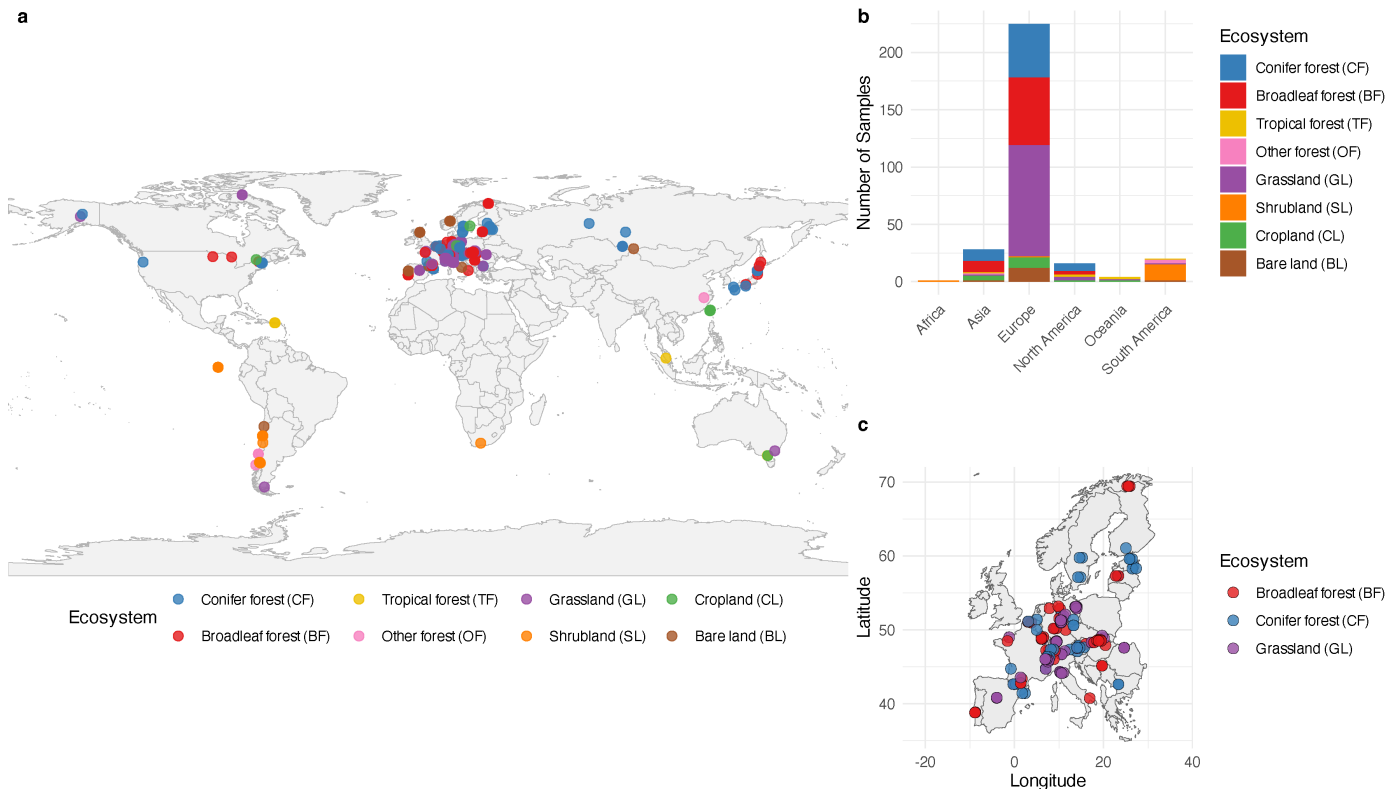


Supplementary Fig.4 | Sampling design for tea-bag incubation and soil collection.

Within each sampling site, at least two sampling plots (blocks) were established, separated by more than 1 m. Green and rooibos tea bags were buried within each block, and soil sampling points were randomly selected within the block (sampling depth, 0 to 5 cm). Soil cores from all sampling points within a block were pooled to form a composite soil sample (composite soil samples 1 and 2) for metagenomic analyses and for measurements of soil physicochemical properties and soil functional indicators.

Supplementary Figures

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



Supplementary Fig. 5 | Global distribution of soil metagenome sampling sites. (a)

Locations of soil metagenome sampling sites worldwide. Points are coloured by ecosystem type, classified from ESA CCI land-cover data and local site descriptions: coniferous forest (CF), broadleaf forest (BF), tropical forest (TF), other forest (OF), grassland (GL), shrubland (SL), cropland (CL) and bare land (BL). (b) Number of soil metagenome samples per continent, grouped by ecosystem type. (c) Locations of European soil metagenome sampling sites used in the primary analyses, coloured by ecosystem type.