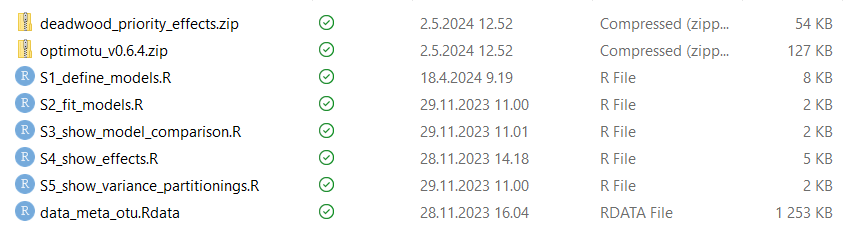
Data, scripts, and pipeline for bioinformatic analyses for manuscript Saine et al. **“Idiosyncratic responses to biotic and environmental filters in wood-inhabiting fungal communities”**

DOI: 10.5281/zenodo.11108552

The results presented in the manuscript can be reproduced by using the input data and running the scripts numbered as S1-S5.



**Bioinformatics pipeline**

To generate the input community data, a bioinformatic pipeline was applied using OptimOTU v0.6.4. Zip-file ***deadwood\_priority\_effects.zip*** includes the archive for the pipeline, and zip-file ***optimotu\_v0.6.4.zip*** includes OptimOTU v0.6.4.

**Scripts S1–S5**

Each script saves its results to the disk, and the following scripts use the results saved by the earlier scripts as input.

* The script ***S1\_define\_models.R*** reads the input data, defines (but does not fit) the set of HMSC models described in the manuscript, and saves these files.
* The script ***S2\_fit\_models.R*** reads the models defined by S1, fits them with MCMC methods, performs cross-validation, and saves the fitted models and cross-validation results to disk.
* The script ***S3\_show\_model\_comparison.R*** reads the cross-validation results from script S2 and shows a model comparison based on them.
* The script ***S4\_show\_effects.R*** shows the parameter estimates (beta-parameters) of the models fitted with script S2.
* The script***S5\_show\_variance\_partitionings.R***computes variance partitioning of the models fitted with script S2.

**Input data**

The file ***data\_meta\_otu.RData*** contains all the necessary input data. The file includes objects:

* ***meta***

Metadata separately for each sample. *SampleID* is the unique ID for each sample, consisting of sampling year, site code, and the running log ID; *SamplingYear* is the year when the data was collected; *SiteNew* is the name of the site; *SiteCode* is an ID for a site; *RunningLogID* is an unique ID for each study log; *CoordN* and *CoordE* are the North and East coordinates for each log in ETRS-TM35FIN coordinate system; *Date2019–2021* is the date of sampling separately for each sampling year; *LogType* indicates whether the log was natural (broken (*B*) or uprooted (*U*)) or felled (*R*); *Size* is the size of the log in centimeters; *DecayStage2019* and *DecayStage2021* are the decay stages of the log measured in 2019 and 2021, respectively; *GroundContact* is the ground contact of the log (%); *BarkCover* is the bark cover of the log (%); *CanopyOpenness* is the proportion of open canopy above the log; *InocSpecies* is the name of the species inoculated in the log; *nStrains* is the number of target species’ strains that were inoculated in the log.

* ***otu.table.plausible***and***otu.table.reliable***

Community matrix (OTU x sampling units) showing the number of sequence reads for each OTU in each sample. Tables are provided separately for OTUs identified with a ≥50% probability threshold (plausible) and a ≥90% probability threshold (reliable).

* ***read.counts.plausible*** and ***read.counts.reliable***

Number of sequence reads per sample after each step in the bioinformatic pipeline. Tables are provided separately for OTUs identified with a ≥50% probability threshold (plausible) and a ≥90% probability threshold (reliable).

* ***taxonomy.plausible*** and ***taxonomy.reliable***

Taxonomic assignment for each OTU with a ≥50% probability threshold (plausible) and a ≥90% probability threshold (reliable). *ref\_seq\_id* is a unique identifier for a most common sequence for the species-level OTU; *nsample* is the total number of sampling units the OTU was detected in; *nread* is the total number of sequence reads for the OTU; *kingdom, phylum, class, order, family, genus, species* is the taxonomic assignment for the OTU.