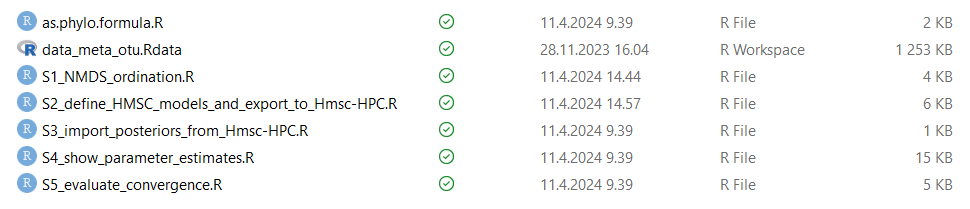
Data and scripts for manuscript Saine et al. “*Species introductions cause phylogenetically structured successional pathways in natural wood-inhabiting fungal communities*”

DOI: 10.5281/zenodo.13819389

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



**Scripts S1–S5**

* The script ***S1\_NMDS\_ordination.R*** reads the input data, runs the NMDS analyses of community data and plots the results.
* The script ***S2\_define\_HMSC\_models\_and\_export\_to\_Hmsc-HPC.R*** reads the input data, defines (but does not fit) the HMSC models described in the manuscript, and exports them to the Hmsc-HPC where the models are fitted. This script uses also the script ***as.phylo.formula.R*** to generate a taxonomic tree for the OTU-level model when defining the models.
* The script ***S3\_import\_posteriors\_from\_Hmsc\_HPC.R*** imports and saves the fitted HMSC models from the Hmsc-HPC.
* The script ***S4\_show\_parameter\_estimates.R*** shows the parameter estimates of the fitted HMSC models, including the variance partitioning and the beta parameters.
* The script ***S5\_evaluate\_convergence.R*** shows results for the MCMC convergence of the fitted HMSC models.

**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.