

## Research

### Data collected by fruit body- and DNA-based survey methods yield consistent species-to-species association networks in wood-inhabiting fungal communities

Sonja Saine, Otso Ovaskainen, Panu Somervuo and Nerea Abrego

S. Saine (<https://orcid.org/0000-0002-4925-5878>) ✉ ([sonja.saine@helsinki.fi](mailto:sonja.saine@helsinki.fi)) and N. Abrego, Dept of Agricultural Sciences, Univ. of Helsinki, Finland. – O. Ovaskainen and P. Somervuo, Organismal and Evolutionary Biology Research Programme, Univ. of Helsinki, Finland.

Oikos

129: 1833–1843, 2020

doi: 10.1111/oik.07502

Subject Editor: Wolf Eiserhardt

Editor-in-Chief: Dries Bonte

Accepted 10 August 2020



Inferring interspecific interactions indirectly from community data is of central interest in community ecology. Data on species communities can be surveyed using different methods, each of which may differ in the amount and type of species detected, and thus produce varying information on interaction networks. Since fruit bodies reflect only a fraction of the wood-inhabiting fungal diversity, there is an ongoing debate in fungal ecology on whether fruit body-based surveys are a valid method for studying fungal community dynamics compared to surveys based on DNA metabarcoding. In this paper, we focus on species-to-species associations and ask whether the associations inferred from data collected by fruit-body surveys reflect the ones found from data collected by DNA-based surveys. We estimate and compare the association networks resulting from different survey methods using a joint species distribution model. We recorded both raw and residual associations that respectively do not and do correct for the influence of the abiotic predictors when estimating the species-to-species associations. The analyses of the DNA data yielded a larger number of species-to-species associations than the analyses of the fruit body-based data as expected. Yet, we estimated unique associations also from the fruit-body data. Our results show that the directions of estimated residual associations were consistent between the data types, whereas the raw associations were much less consistent, highlighting the need to account for the influence of relevant environmental covariates when estimating association networks. We conclude that even though DNA-based survey methods are more informative about the total number of interacting species, fruit-body surveys are also an adequate method for inferring association networks in wood-inhabiting fungi. Since the DNA and fruit-body data carry on complementary information on fungal communities, the most comprehensive insights are obtained by combining the two survey methods.

**Keywords:** co-occurrence, DNA metabarcoding, interspecific interaction, joint species distribution model, observational data, residual association

## Introduction

Interspecific interactions are often hard to observe directly in species-rich communities. Consequently, inferring interspecific interactions indirectly from community data is of central interest in community ecology (Kissling et al. 2012, Morales-Castilla et al. 2015, Dormann et al. 2018) and fundamental for predicting community dynamics under the global environmental change (Araújo and Luoto 2007, Gilman et al. 2010, Wisz et al. 2013, Urban et al. 2016). However, data on species communities can be surveyed using different methods, each of which may differ in the amount and type of species that are detected, and thus produce diverging information on interaction networks (Wirta et al. 2014, Zhu et al. 2019). Data based on direct observations of species are prone to introducing false negatives due to imperfect detection (Guillera-Arroita 2017), whereas false positives due to species misidentifications are typical of data derived from DNA-based methods (Ficetola et al. 2015). Such observation error associated with each type of data is one of the reasons for the ongoing debate on which data types are valid for drawing ecological inferences (Yoccoz et al. 2001, Tyre et al. 2003, Lahoz-Monfort et al. 2014, Guillera-Arroita 2017). Given the exponential rise of DNA-based community datasets in ecology (Bush et al. 2017), assessing how the survey method affects the inferred interaction networks is a premise for advancing research on community ecology.

In species-rich communities where observation of biotic interactions is challenging, interactions are often estimated from snapshot data using network and co-occurrence analyses (Steele et al. 2011, Barberán et al. 2012, Weiss et al. 2016). Because species can co-occur more than expected by random not only due to interactions but also due to shared habitat requirements, the development of methods allowing to disentangle between the environmentally constrained and non-environmentally constrained co-occurrences is gaining increasing interest in the ecological literature (Peres-Neto et al. 2001, Ovaskainen et al. 2010a, 2016, Barberán et al. 2012, Legendre and Legendre 2012, Blois et al. 2013, Pollock et al. 2014, Morueta-Holme et al. 2016, Zurell et al. 2018). Recently developed joint species distribution models with latent variable structures enable inferring the residual associations (i.e. the co-occurrence patterns remaining after accounting for the effects of environmental predictors) from datasets consisting of large amounts of species (Warton et al. 2015, Ovaskainen et al. 2017). Since the residual species-to-species associations can reflect both biotic interactions and shared responses to missing environmental variables, it is important to note that they should not be directly interpreted as interspecific interactions (Ovaskainen et al. 2010a, 2017, Dormann et al. 2018). However, compared to raw associations, residual association provide a closer approximation of species interactions.

Fungal species comprise megadiverse communities (Blackwell 2011) in which species interact in a number of ways, ranging from strong antagonisms to mutualism

(Boddy et al. 2008). In particular, wood-inhabiting fungal communities are highly structured by interspecific interactions. Laboratory experiments have shown that the community development of wood-inhabiting fungi is largely affected by biotic interactions (Holmer and Stenlid 1997, Heilmann-Clausen and Boddy 2005, Fukami et al. 2010, Hiscox et al. 2015a). Likewise, field studies based on snapshot fruit-body observations suggest that during the decay process of dead wood, the identity of a successor fungal species is affected by the identity and abundance of the preceding fungal species (Renvall 1995, Pouska et al. 2013, Ottosson et al. 2014, Norberg et al. 2019). In addition, environmental conditions may influence the outcome of interspecific interactions (Hiscox et al. 2016) and conversely, fungal interactions may influence ecosystem functions such as wood decomposition (Boddy 2001, Fukami et al. 2010, Hiscox et al. 2015b, Maynard et al. 2018).

To date, most studies exploring wood-inhabiting fungal communities have applied fruit-body surveys (Halme et al. 2012). The method is relatively fast and low-cost, and thus enables large scale comparative studies with conspicuous species (Halme et al. 2012, Runnel et al. 2015). Nevertheless, several fungal groups produce ephemeral fruit bodies during limited fruiting seasons, making the number and timing of surveys critical for species' detectability (Halme and Kotiaho 2012, Abrego et al. 2016, Purhonen et al. 2017). Moreover, some wood-inhabiting fungi are microscopic (Lumley et al. 2000) and some individuals may not produce fruit bodies at all (Moore et al. 2008). Therefore, only a fraction of species detected by DNA-based metabarcoding methods are detected by fruit body-based (Allmér et al. 2006, Ovaskainen et al. 2010b, 2013, Kubartová et al. 2012, Ottosson et al. 2015), generating the debate on whether fruit body-based surveys are a valid method for studying fungal community dynamics compared to DNA-based surveys (Allmér et al. 2006, Porter et al. 2008, Halme et al. 2012, van der Linde et al. 2012, Runnel et al. 2015, Frøslev et al. 2019). Since fruit body-based surveys are able to capture only the part of the fungal community that is in the reproductive phase, these methods are clearly less optimal than DNA-based surveys for assessing the fungal diversity in terms of the number of species present. While Ovaskainen et al. (2013) showed that the ecological signal on the effects of dead wood properties on wood-inhabiting fungal communities is highly consistent between data based on fruit-body observations and DNA metabarcoding, it remains unclear whether the data collected by these two survey methods carry the same (or complementary) information on species-to-species associations for those species that can be detected with both methods.

In this paper, our aim is to assess whether the species-to-species associations inferred from fruit body-based data reflect the ones inferred from DNA-based data. We hypothesize that DNA-based data allows estimating more associations compared to fruit body-based data, yet those associations detected from both data types are consistent. For comparing data collected by these survey methods, we use snapshot fruit-body

and sequence data published in Ovaskainen et al. (2013) and Mäkipää et al. (2017). The data consist of 100 Norwegian spruce *Picea abies* logs of similar size that were simultaneously sampled by fruit-body and DNA-based surveys, the latter based on ITS metabarcoding of wood samples. We estimate and compare the species-to-species associations using a joint species distribution model. The model partitions the variation in species occurrences and co-occurrences to that explained by the abiotic predictors and variation remaining after accounting for these abiotic predictors, namely residual associations (Ovaskainen et al. 2016, 2017).

## Material and methods

### Data collection

The data were collected from a protected, natural-like spruce dominated forest located in the municipality of Sipoo in Helsinki-Uusimaa Region, Finland (Ovaskainen et al. 2013). In this site, 100 large-sized logs (DBH 20–43 cm) were randomly selected. The selected logs varied in their decay stage (decay classes 1–4; see Hottola and Siitonen (2008) for class descriptions, and Supplementary Table 1 in Ovaskainen et al. (2013) for a table of decay stages), fall type (uprooted or broken) and ground contact but as minimally as possible in their size.

In November 2008, saw dust samples were taken from the 100 spruce logs for subsequent fungal DNA sequencing. Saw dust samples were taken separately from the basal and the middle part of the log with an electric drill. With strongly decayed logs, a sampling cylinder was used. The fruit-body surveys were carried out three times during the peak fruiting season. In September and November 2008, all logs were sampled twice for fruit bodies of all polypores and certain corticioid and hydroid basidiomycetes that were easily identified in the field. During a third survey in October 2009, all polypore, corticioid and hydroid basidiomycetes and some easily identified ascomycetes were sampled. If the species could not be reliably identified in the field, specimen samples were taken for later microscopic identification. During the latter two surveys, the abundance of species on a log was measured as the area covered by fruit bodies (cm<sup>2</sup>) for all polypores and those corticioids and hydroid species that were successfully identified in the field. For more detailed information on the sampling scheme, see Ovaskainen et al. (2013).

### DNA extraction and sequencing

The protocols for DNA extraction and sequencing are described in detail in Ovaskainen et al. (2013). Briefly, DNA extractions for the 200 saw dust samples (two for each log) were done using Power Soil DNA isolation kit (MoBio Laboratories, Inc., Carlsbad, CA, USA). Taq polymerase enzyme (Thermo Fisher Scientific, Waltham, MA, USA) with the primers ITS1 and ITS4 (White et al. 1990) was used to amplify the target DNA region. Real-time quantitative PCR

(qPCR) assays were conducted on the 18S ribosomal RNA gene. Samples with a DNA concentration  $\geq 0.05$  ng  $\mu\text{m}^{-1}$  were included in the qPCR analyses (163 out of 200 samples). CGX384 Real-Time PCR detection system was used to run the reactions. As a results, around 400 PCR fragments were obtained for the primers ITS1 and ITS2 to be used in the 454-sequencing.

For 297 samples with a DNA concentration  $> 1.43$  ng  $\mu\text{m}^{-1}$ , PCR reactions for ITS1 region were done with a primer pair ITS1F (Gardes and Bruns 1993) and ITS2 (White et al. 1990). The primer pair ITS4 and ITS3 (White et al. 1990) was used for ITS2 region. For 97 samples with a DNA concentration  $< 1.43$  ng  $\mu\text{m}^{-1}$ , a shorter form of ITS1F and ITS4 primers without the A and tag sequences was used for both ITS1 and ITS2 regions. Then, the PCR reactions were diluted and amplified again using ITS1F and ITS4 primers with tag and A sequences. Finally, sequencing was done using a Genome Sequencer FLX. In the case of six samples, no PCR fragments could be obtained.

We note that the DNA extraction and sequencing methods applied for Ovaskainen et al. (2013) represent the early stages of ITS metabarcoding techniques. Therefore, in these data the fungal community might be underrepresented, especially for those less abundant species requiring a deeper sequencing depth for being detected. However, Ovaskainen et al. (2013) showed a good correspondence between fruit-body and DNA abundance, suggesting a high signal-to-noise ratio in these data despite the limitations.

### Molecular species identification

For this paper, we renewed the original molecular species identifications done for Ovaskainen et al. (2013) using PROTAX-fungi (Abarenkov et al. 2018). PROTAX-fungi is a software performing probabilistic taxonomic placement of fungal ITS sequences to known and unknown fungal lineages. The tool is based on a taxonomic classification system of Index Fungorum + Species Fungorum (Royal Botanic Gardens Kew et al. 2019) that covers a total of 131 484 species. The reference sequences are obtained from UNITE database (Kõljalg et al. 2013). PROTAX-fungi uses a database that consists of 420 319 reference sequences out of which 217 663 are annotated at the species level (Abarenkov et al. 2018). In our sequence data used for the identifications, samples for the ITS1 and ITS2 regions and for the basal and the middle part of the log were treated separately (ca 400 samples altogether, four per log). We included here those taxonomic placements that could be assigned to the species level with at least 50% probability.

### Data preparation

We treated individual logs as sampling units, and thus we first pooled the data at the log level. For the data based on fruit-body observations, the area covered by fruit bodies of each species in the basal and the middle part of the logs were summed for each log, and then averaged over the surveys to

yield a measure of fruit-body abundance for those species that were measured in the field (Ovaskainen et al. 2013). In the DNA-based data, we pooled the ITS1 and ITS2 sequences as well as the basal and the middle part of the logs by summing the species-specific sequence counts for all species detected. The sequence counts were used as a proxy of DNA abundance. Thus, the resulting data consisted of occurrence (i.e. presence-absence) and abundance (i.e. fruit-body coverage and number of sequences mapped to the focal species) of each species as fruit bodies and DNA on each log. The species nomenclature in fruit body-based data was updated according to Index Fungorum (Royal Botanic Gardens Kew et al. 2019) to match the nomenclature in the DNA-based data. Sequencing depth (i.e. the total number of sequences in a sample) was used as an estimate for sampling effort in the DNA-based data (Tedersoo et al. 2014).

The original data based on fruit-body observations is available from Supplementary Table 1 in Ovaskainen et al. (2013): <[https://static-content.springer.com/esm/art%3A10.1038%2Fismej.2013.61/MediaObjects/41396\\_2013\\_BFismej201361\\_MOESM21\\_ESM.xls](https://static-content.springer.com/esm/art%3A10.1038%2Fismej.2013.61/MediaObjects/41396_2013_BFismej201361_MOESM21_ESM.xls)>. The raw sequence data (Ovaskainen et al. 2013, Mäkipää et al. 2017) is archived in NCBI (National Center for Biotechnology Information, Bethesda, MD, USA) BioProject ID PRJNA341814.

## Statistical methods

Since our aim was to test whether association networks estimated using each data type are consistent with each other, for the statistical analyses we selected only those fungal species that were detected both in fruit-body and DNA-based data. Namely, we did not estimate the associations among those species detected by only one of the survey methods, as those associations can obviously not be compared between datasets.

To obtain sufficient statistical power, only those species that occurred at least on 10 logs in at least one of the datasets (based on fruit-body observations or DNA) were included in the analyses. Such prevalence threshold was applied because if a given species occurs in a very small fraction of the sampling units, there is not sufficient statistical power to reliably estimate the species-to-species associations (Ovaskainen and Abrego 2020).

We conducted our analyses with the R package Hmsc (<[www.r-project.org](http://www.r-project.org)>, Tikhonov et al. 2020) which belongs to the class of joint species distribution models (Warton et al. 2015, Ovaskainen et al. 2017). Joint species distribution models allow inferring the species-to-species associations after controlling for the effects of the variables included in the model. The response matrix **Y** (see Ovaskainen et al. 2017 for notation) consisted of the occurrences and log-transformed abundances of each species as fruit bodies and DNA on the 100 sampling units (i.e. logs). Each species was thus included in the **Y** matrix in four columns representing the different types of data. We applied a hurdle approach that modelled the occurrences (presence-absences) with probit regression

and log-transformed abundances conditional on presence with linear regression. Fruit-body and DNA abundances were standardized to zero mean and unit variance and considered conditional on presence, i.e. considered as missing data for cases in which the species were absent. Decay stage (categorical variable with the four levels), ground contact (present or absent), fall type (uprooted or broken) and log-transformed sequencing depth were included as the sampling unit level explanatory variables in the matrix **X**. To estimate the species-to-species association networks, we included community-level random effects at the scale of sampling unit. The data type was included as a categorical variable with four levels (fruit-body occurrence, fruit-body abundance, DNA occurrence or DNA abundance) in the species trait matrix **T**. To disentangle between the raw species associations (i.e. those arising mainly from the shared log characteristics) and the residual associations (i.e. the ones remaining after controlling for log characteristics), we fitted two model variants. The first model variant with raw associations included sequencing depth as the only covariate (model 1), and the second one with residual associations included sequencing depth, decay stage, ground contact and fall type as covariates (model 2).

To sample the posterior distribution, we applied the default priors in the R-package Hmsc (Tikhonov et al. 2020) and performed 1 500 000 Markov chain Monte Carlo (MCMC) iterations with four chains. The first 500 000 iterations were removed as burn-in and the remaining ones were thinned by 1000, resulting in 1000 posterior samples per chain, and thus 4000 posterior samples in total. Before fitting the final model, we explored the rate of MCMC convergence by fitting otherwise identical models but with 1500 (burn-in 500, thin 1), 15 000 iterations (burn-in 5000, thin 10) and 150 000 iterations (burn-in 50 000, thin 100). We followed Tikhonov et al. (2020) by evaluating MCMC convergence by computing the effective sample size and the potential scale reduction factor for the model parameters related to species-level ( $\beta$ ) and community-level ( $\gamma$ ) responses to environmental covariates, residual variation in species responses to environmental covariates ( $V$ ) and the species association matrices ( $\Omega$ ).

We assessed the explanatory power by computing for each species and each data type the AUC for the occurrence data and  $R^2$  for the abundance data. Then, we averaged for each data type the AUC values over all species and the  $R^2$  values over those species for which abundance data was available for at least five logs. To examine the relative importance of environmental covariates and random effects (i.e. the associations) for the species communities, we performed variance partitioning among the explanatory variables.

To compare the species-to-species association networks identified from the fruit-body and DNA-based data, we classified the estimated associations as positive, negative or neutral, depending whether the posterior probability for the association was greater than 0.75, smaller than 0.25 or between 0.25 and 0.75, respectively. We note that the neutral case includes both those species pairs for which there is no



association in reality, as well as those for which there may be an association in reality but the data do not have the sufficient statistical power for identifying it. We computed the numbers of species pairs for which the fruit-body and DNA-based data showed consistent results (both positive, both negative or both neutral), contradictory results (one positive and the other one negative), and for which the results were not consistent nor contradictory (one positive or negative and the other one neutral).

## Results

Altogether, 142 species were observed in the fruit body-based surveys. The DNA sequencing resulted in a total of 1 349 965 sequences, out of which 18.5% (249 153) were assigned to a species level with at least 50% probability, yielding 413 species. Fruit-body and DNA-based data shared 80 species, out of which 37 species occurred at least 10 times in either fruit-body or DNA-based data and were consequently included in the HMSC-analyses. Out of these 37 target species, fruit-body abundance was available for 16 species. On average, each log hosted 6.2 target species as fruit bodies (ranging from 0 to 15) and 6.7 species as DNA (ranging from 1 to 15). The mean number of occurrences for the target species was 16.7 for fruiting species (ranging from 1 to 73) and 18.2 for species detected as DNA (ranging from 1 to 61). For a list of the 37 species included in the analyses and information on their prevalence and abundance, see Supplementary material Appendix 1 Table A1.

The MCMC convergence was satisfactory, as the effective sample sizes were close to the theoretical optimum of 4000, and the potential scale reduction factors were close to the theoretical optimum of one (Supplementary material Appendix 1 Fig. A1). Measured by the AUC statistic, the fitted model with residual associations explained on average 82% (respectively, 84%) of the variation in the occurrence of fruiting (respectively, present as DNA) species. Measured by  $R^2$ , the fitted model with residual associations explained on average 27% (respectively, 29%) of the variation in species abundance (conditional on presence) of fruiting (respectively, present as DNA) species. Decay stage was the most important abiotic predictor, explaining 47% of the variance (on average over all species and data types). Ground contact, fall type and sequencing depth explained on average 9, 11 and 12% of the variance, respectively. The 22% of the variance belonged to the random effects at the log level. As expected, including the environmental covariates in the model improved the model fit (Supplementary material Appendix 1 Table A2) and decreased the proportion of variance explained by the random effect (Supplementary material Appendix 1 Table A3).

The total number of detected species-to-species associations was higher in the DNA than fruit body-based data (Fig. 1, 2). However, the directions of estimated species-to-species associations were generally consistent between the data types, especially in the case of residual associations (Fig. 1b) that showed no contrary associations. In other

words, species pairs showing positive (respectively, negative) associations in the fruit body-based data showed positive (respectively, negative) associations also in the DNA-based data (Fig. 1). In the occurrence data, a total of 666 residual species-to-species associations were detected. Out of these, 61.5% were consistent (both positive 11.4%, both negative 8.1% or both neutral 42.0%) between fruit-body and DNA-based data (Supplementary material Appendix 1 Table A4). No contrary associations were found (i.e. where one data type would support a positive association and the other one negative) (Supplementary material Appendix 1 Table A4). The remaining 38.4% represent cases in which the association was detected only in one data type (i.e. one data type giving support for a positive or negative association, and the other one for a neutral association) (Supplementary material Appendix 1 Table A4, Fig. 2). Such unique, positively and negatively associated species pairs were detected with both methods but the number of such pairs was larger for DNA-based data (Fig. 2a–b). Out of all positive and negative residual associations in the occurrence data, 33.7% were detected by both methods (34.9% for raw associations), 18.7% only by fruit-body surveys (23.1% for raw associations) and 47.6% only by molecular surveys (41.9% for raw associations) (Fig. 2a–b). Considering the type of detected association, both fruit-body and DNA-based data recorded more positive than negative associations in the occurrence data (Fig. 1, 2).

Compared to residual associations, raw associations in the occurrence data recorded a somewhat lower proportion (52.4%) of consistent associations (both positive 17.7%, both negative 7.2% or both neutral 27.5%) (Supplementary material Appendix 1 Table A4). They showed also some contradictory associations (1.2%), and more associations that were not consistent nor contradictory (46.5%) (Supplementary material Appendix 1 Table A4). In the case of abundance data (conditional on presence), only few associations were recorded in general (Fig. 1, 2c–d), making the comparison between fruit-body and DNA-based data less informative (Supplementary material Appendix 1 Table A4).

## Discussion

In this study, we demonstrate that even if data derived from fruit body-based surveys are particularly prone to introducing false negatives compared to data derived from DNA-based surveys, these data types are valid for inferring species-to-species association networks in wood-inhabiting fungal communities. Our results show that the directions of estimated residual species-to-species associations were consistent between the data derived from direct fruit-body observations and DNA-based surveys. However, there were differences in the amount of association links detected by each of the survey methods, and consistency of the links varied depending on whether the associations were estimated from models that accounted or did not account for the effects of environmental conditions. Below, we will explain each finding in turn.

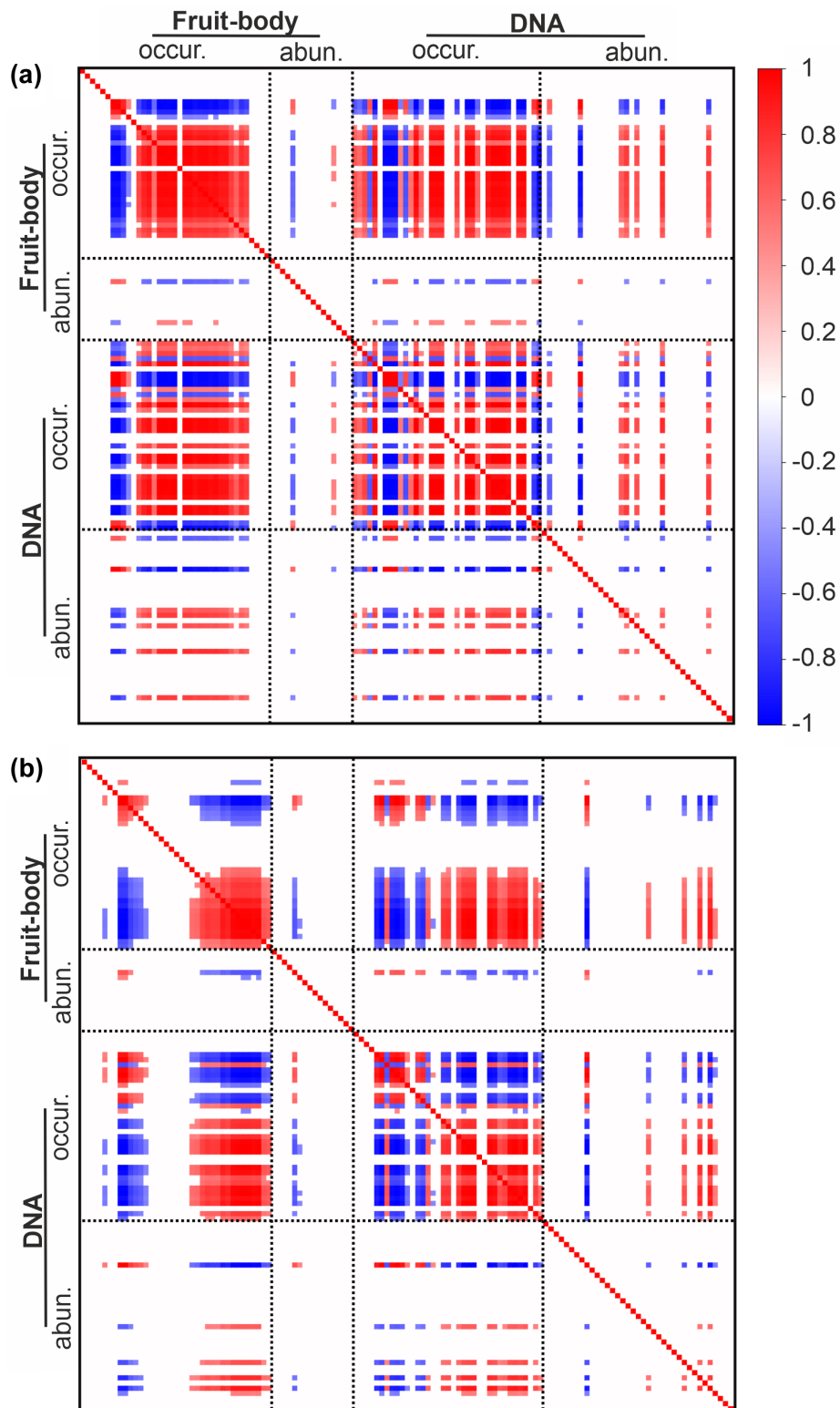


Figure 1. The estimated species-to-species associations measured as (a) raw and (b) residual correlations, separately for each data type (species occurrences (occur.) and abundances (abun.) in data based on fruit-body observations and DNA metabarcoding). The matrices show species pairs with positive (red) and negative (blue) associations with at least 75% posterior probability. The color shade indicates the strength of the correlation (the darker the shade, the stronger the correlation). The remaining cases, i.e. neutral associations, are shown in white. Each data type contains the same set of 37 species (except for fruit-body abundance data including 16 out of these species) in a corresponding order (for species names, see Supplementary material Appendix 1 Table A1). The order is set to highlight positively and negatively associated groups.

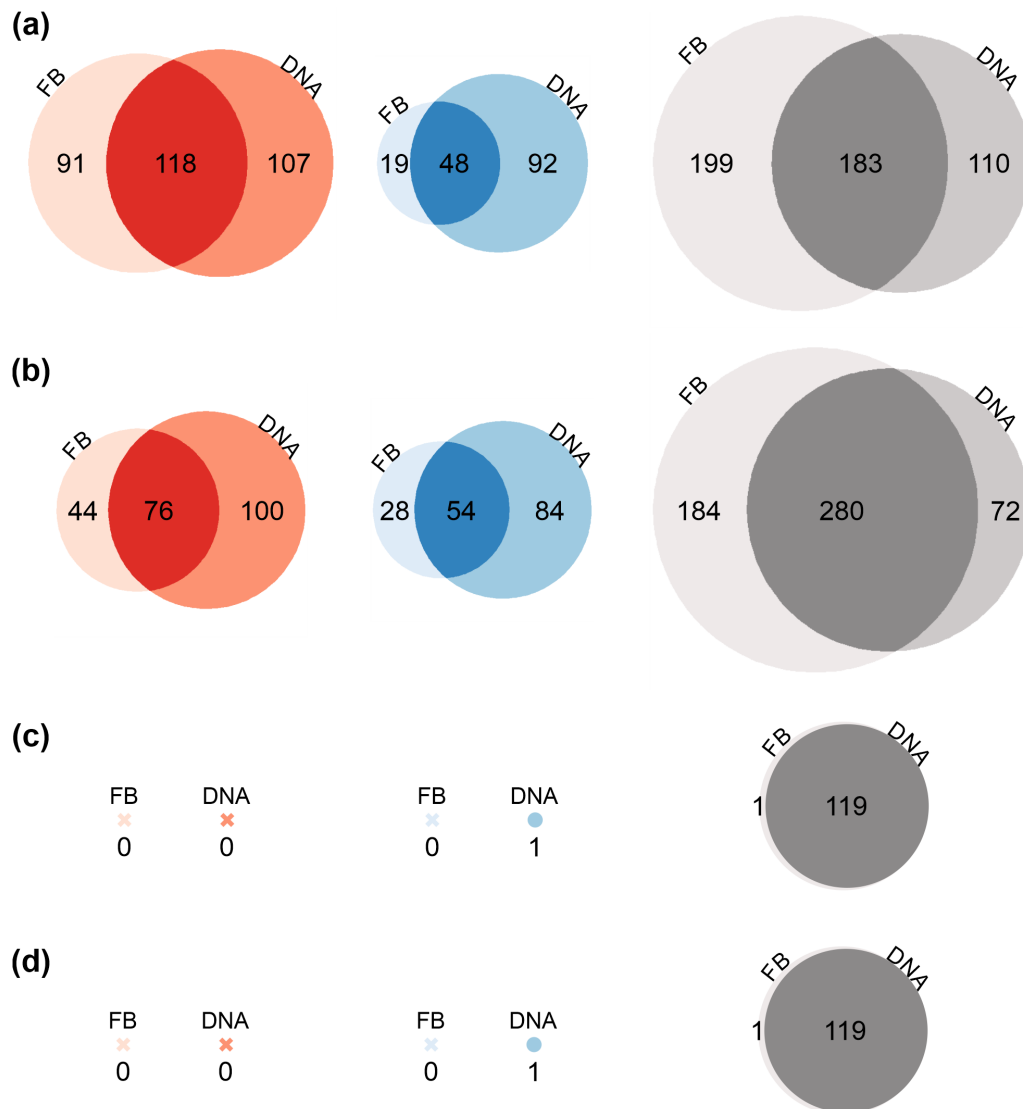


Figure 2. Euler diagrams showing the number of positively (red), negatively (blue) and neutrally (grey) associated species pairs estimated from the data based on fruit-body observations (FB) and DNA metabarcoding (DNA) (numbers on the left- and right-hand sides of the diagrams, respectively), and the number of species pairs recorded by both survey methods (numbers within the overlapping areas). The first and second rows of panels show (a) raw and (b) residual associations in the occurrence data, and the third and fourth rows of panels show (c) raw and (d) residual associations in the abundance data. The sizes of the circles correspond to the total number of associations detected by each survey method. If a data type detected no associations, this is shown with a cross.

Compared to more classical species surveys, DNA metabarcoding methods yield more species-rich community datasets and represent more inconspicuous species (Creer et al. 2016). This is also the case for wood-inhabiting fungi (Allmér et al. 2006, Kubartová et al. 2012, Ovaskainen et al. 2013, Ottosson et al. 2015). Supporting these previous findings and our hypotheses, our results showed that the DNA-based survey detected more species than the survey based on fruit-body observations. However, with DNA-based methods not only more species are detected per log, but also each species is likely to be detected in a larger number of logs than with fruit-body surveys, providing stronger statistical power for inferring species-to-species associations.

Consequently, we estimated a larger absolute number of associations from data based on DNA than from data based on fruit-body observations. The result is linked to the observational error in fruit body-based surveys. Only a fraction of the species found as DNA are detected as fruit bodies since only the most abundant species as DNA produce fruit bodies and they appear with a time delay from the colonization (Allmér et al. 2006, Kubartová et al. 2012, Ovaskainen et al. 2013, Ottosson et al. 2015). Furthermore, most of the species produce ephemeral fruit bodies emerging at different times during the fruiting season (Halme and Kotiaho 2012, Abrego et al. 2016, Purhonen et al. 2017), fruit bodies of some fungal species are too small to be detected with naked eye

(Cannon and Sutton 2004, Abrego and Salcedo 2015) and some fungal individuals may never produce fruit bodies (Moore et al. 2008). Moreover, supporting our results, interspecific interactions in fungal communities take place mainly at the mycelial stage (Heilmann-Clausen and Boddy 2005, Fricker et al. 2008, Hiscox et al. 2018), which can be surveyed using the DNA-based methods only. Before the potential fruiting, species have to occupy space and resources within dead wood as mycelium and while doing so, they necessarily interact with other species (Boddy 2000, Boddy and Hiscox 2016).

There were also associations that were estimated only from the data based on fruit-body observations. Most likely these species pairs represent cases in which sequences have been misidentified in the DNA-based data. A large body of literature has demonstrated that when applying molecular species identification methods to data originating from high-throughput sequencing of environmental samples, species can easily be misidentified (Meyer and Paulay 2005, Lou and Golding 2012, Ficetola et al. 2015, Lahoz-Monfort et al. 2016). There are myriad reasons for this, such as sequencing errors, mislabeling errors in the reference databases, and the fact that some species are simply missing from the reference databases (Lou and Golding 2012, Somervuo et al. 2017). As a consequence, some species that actually occurred in sampling units might be missing from the data, while some absent species might be falsely recorded to be present. To take these issues into account as much as possible, we applied a probabilistic taxonomic placement method which accounts for these sources of uncertainty while estimating the reliability of taxonomic identifications (Somervuo et al. 2016, Abarenkov et al. 2018). Also, by including sequencing depth as a covariate in the models, we accounted for false absences due to limited observation effort (Ovaskainen et al. 2017). We also note that some species might remain undetected by the DNA-based survey also because saw dust samples are limited in volume (Allmér et al. 2006, Runnel et al. 2015), affecting the number of species detected (Ovaskainen et al. 2010b, Kubartová et al. 2012). Furthermore, the data applied in this study originates from the early phases of ITS metabarcoding, and thus it is not fully state-of-the-art in terms of the DNA-extraction and sequencing methods. More modern methods enable, for example, a higher sequencing depth which decreases the likelihood for missing species that are present in low DNA abundances. However, as here we focused only on the dominating part of the species community, we expect our results to be robust even if they can be somewhat conservative due to the noise generated by relatively modest sequencing depth.

The level of consistency between the species-to-species associations inferred from different data types depended on how the associations were estimated. By adding the environmental covariates in the model, the proportion of consistent associations increased compared to the model without the covariates. This supports the use of residual species-to-species associations as a hypothesis of species interactions, as

has been done in previous studies (Ovaskainen et al. 2010b, 2016, Ottosson et al. 2014, Abrego et al. 2017). The result also demonstrates that some species co-occur merely due to shared responses to certain abiotic conditions. In specific, the production of fungal fruit bodies depends strongly on the environmental conditions (Moore et al. 2008). Therefore, it is highly relevant to account for the effects of these conditions when inferring species-to-species associations from data based on fruit-body observations.

We emphasize that while some of the residual associations detected in our study might represent biotic interactions, these can only be confirmed through experimental tests (Dormann et al. 2018). The detected positive residual species-to-species associations might reflect facilitative biotic interactions (Tiunov and Scheu 2005, Fukami et al. 2010) or parasitism (Niemelä et al. 1995), whereas the negative associations might reflect competitive interactions among wood-inhabiting fungi (Boddy 2000, Heilmann-Clausen and Boddy 2005). Species-to-species associations might also reflect community succession and priority effects (Ottosson et al. 2014, Norberg et al. 2019). For instance, the predecessor species *Fomitopsis rosea* showed a negative association with the successor *Phellolipilus nigrolimitatus* both in the fruit-body and DNA occurrence data, indicating that the later species replaces the former during species succession (Niemelä et al. 1995). Given the high predictive power of our model including environmental covariates, it is plausible that the detected residual species-to-species associations represent interspecific interactions. However, some of the associations most likely reflect a shared response to some missing environmental covariate that is relevant for the wood-inhabiting fungal species included in our study. For instance, we included decay stage of the logs as a proxy of the physiochemical properties of the wood. Yet, a direct measurement of physiochemical properties could have explained the occurrence of some species better. We also note that while competition is known to be the most common interaction type among wood-inhabiting fungi (Boddy 2000), we found more positive than negative species-to-species associations. This is partially because when inferring species-to-species associations using statistical methods, more data are needed for inferring negative than positive associations (Ovaskainen et al. 2016). In addition, also competing species might co-occur on a sampling unit level if they are spatially separated within a log, leading to a positive association at the log level (Ovaskainen et al. 2010a, Ottosson et al. 2014).

Finally, while a large number of residual species-to-species associations were found from occurrence data, only few associations were found from the abundance data. Namely, the fact that a given species was abundant did not affect the abundances of other species. This suggests that wood-inhabiting fungal interactions take place when the potentially interacting species are present, the interactions not depending on the abundance of interacting species. On the other hand, there might not have been enough statistical power to estimate associations from the abundance data as most of the



abundance data from fruit-body surveys were missing. Also, in the case of sequence data, read counts are affected by a range of factors and cannot be directly translated to abundance data (Deagle et al. 2019).

We conclude that even though fruit-body surveys focus only on the reproductive part of the wood-inhabiting fungal community, data derived from fruit-body and DNA-based surveys yield consistent association networks of wood-inhabiting fungi. This supports the robustness of our results and therefore, both methods should lead to comparable ecological inference on the association network structure of wood-inhabiting fungi. DNA-based surveys were more informative on the total number of interacting species but since unique interactions were also estimated from the fruit-body data, a combination of the methods would be ideal to obtain the most comprehensive insight into the network structure of wood-inhabiting fungi. Finally, our study demonstrates that accounting for the effects of relevant environmental covariates is of particular importance when inferring species-to-species associations in wood-inhabiting fungal communities.

## Data availability statement

The original data used in this study are published in Ovaskainen et al. (2013) and Mäkipää et al. (2017). The data formatted for this study as well as the R scripts for reproducing the results are available from the Dryad Digital Repository: <<http://dx.doi.org/10.5061/dryad.tmp4f4wm>> (Saine et al. 2020).

**Acknowledgements** – We thank all the people involved in collecting and compiling the dataset published in Ovaskainen et al. (2013) and Mäkipää et al. (2017) which we used for our study: H. Ali-Kovero, P. Auvinen, I. Eriksson, P. Halme, K. Juutilainen, L. Kajander, J. Kytömäki, O. Miettinen, H. Murdoch, R. Mäkipää, B. Nordén, E. Ottosson, T. Parkkima, L. Paulin, R. Penttilä, K. Savola, M. Suominen, D. Schigel, M. Tamminen and I. Viner.. **Funding** – This study was funded by the Research Foundation of the Univ. of Helsinki (SS), the Academy of Finland (grant no. 308651 to NA and grant 309581 to OO), and the Jane and Erkkö Aatos foundation (grant to OO).

## References

- Abarenkov, K. et al. 2018. Protax-fungi: a web-based tool for probabilistic taxonomic placement of fungal internal transcribed spacer sequences. – *New Phytol.* 220: 517–525.
- Abrego, N. and Salcedo, I. 2015. Taxonomic gap in wood-inhabiting fungi: identifying understudied groups by a systematic survey. – *Fungal Ecol.* 15: 82–85.
- Abrego, N. et al. 2016. Fruit body based inventories in wood-inhabiting fungi: should we replicate in space or time? – *Fungal Ecol.* 20: 225–232.
- Abrego, N. et al. 2017. Wood-inhabiting fungi with tight associations with other species have declined as a response to forest management. – *Oikos* 126: 269–275.
- Allmér, J. et al. 2006. Wood-inhabiting fungal communities in woody debris of Norway spruce (*Picea abies* (L.) Karst.), as reflected by sporocarps, mycelial isolations and T-RFLP identification. – *FEMS Microbiol. Ecol.* 55: 57–67.
- Araújo, M. B. and Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. – *Global Ecol. Biogeogr.* 16: 743–753.
- Barberán, A. et al. 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. – *ISME J.* 6: 343–351.
- Blackwell, M. 2011. The fungi: 1, 2, 3 ... 5.1 million species? – *Am. J. Bot.* 98: 426–438.
- Blois, J. L. et al. 2013. Climate change and the past, present and future of biotic interactions. – *Science* 341: 499–504.
- Boddy, L. 2000. Interspecific combative interactions between wood-decaying basidiomycetes. – *FEMS Microbiol. Ecol.* 31: 185–194.
- Boddy, L. 2001. Fungal community ecology and wood decomposition processes in angiosperms: from standing tree to complete decay of coarse woody debris. – *Ecol. Bull.* 49: 43–56.
- Boddy, L. and Hiscox, J. 2016. Fungal ecology: principles and mechanisms of colonization and competition by saprotrophic fungi. – *Microbiol. Spectr.* 4: 1–16.
- Boddy, L. et al. (eds) 2008. Ecology of saprotrophic basidiomycetes. – Elsevier.
- Bush, A. et al. 2017. Connecting Earth observation to high-throughput biodiversity data. – *Nat. Ecol. Evol.* 1: 0176.
- Cannon, P. F. and Sutton, B. C. 2004. Microfungi on wood and plant debris. – In: Mueller, G. M. et al. (eds), *Biodiversity of fungi: inventory and monitoring methods*. Elsevier Academic Press, pp. 217–239.
- Creer, S. et al. 2016. The ecologist's field guide to sequence-based identification of biodiversity. – *Methods Ecol. Evol.* 7: 1008–1018.
- Deagle, B. E. et al. 2019. Counting with DNA in metabarcoding studies: how should we convert sequence reads to dietary data? – *Mol. Ecol.* 28: 391–406.
- Dormann, C. F. et al. 2018. Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. – *Global Ecol. Biogeogr.* 27: 1004–1016.
- Ficetola, G. F. et al. 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. – *Mol. Ecol. Resour.* 15: 543–556.
- Fricker, M. D. et al. 2008. Mycelial networks: structure and dynamics. – In: Boddy, L. et al. (eds), *Ecology of saprotrophic basidiomycetes*. Elsevier, pp. 3–18.
- Frøslev, T. G. et al. 2019. Man against machine: do fungal fruitbodies and eDNA give similar biodiversity assessments across broad environmental gradients? – *Biol. Conserv.* 233: 201–212.
- Fukami, T. et al. 2010. Assembly history dictates ecosystem functioning: evidence from wood decomposer communities. – *Ecol. Lett.* 13: 675–684.
- Gardes, M. and Bruns, T. D. 1993. ITS primers with enhanced specificity for basidiomycetes – application to the identification of mycorrhizae and rusts. – *Mol. Ecol.* 2: 113–118.
- Gilman, S. E. et al. 2010. A framework for community interactions under climate change. – *Trends Ecol. Evol.* 25: 325–331.
- Guillera-Arroita, G. 2017. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. – *Ecography* 40: 281–295.

- Halme, P. and Kotiaho, J. S. 2012. The importance of timing and number of surveys in fungal biodiversity research. – *Biodivers. Conserv.* 21: 205–219.
- Halme, P. et al. 2012. Monitoring fungal biodiversity – towards an integrated approach. – *Fungal Ecol.* 5: 750–758.
- Heilmann-Clausen, J. and Boddy, L. 2005. Inhibition and stimulation effects in communities of wood decay fungi: exudates from colonized wood influence growth by other species. – *Microb. Ecol.* 49: 399–406.
- Hiscox, J. et al. 2015a. Priority effects during fungal community establishment in beech wood. – *ISME J.* 9: 2246–2260.
- Hiscox, J. et al. 2015b. Antagonistic fungal interactions influence carbon dioxide evolution from decomposing wood. – *Fungal Ecol.* 14: 24–32.
- Hiscox, J. et al. 2016. Effects of pre-colonisation and temperature on interspecific fungal interactions in wood. – *Fungal Ecol.* 21: 32–42.
- Hiscox, J. et al. 2018. Fungus wars: basidiomycete battles in wood decay. – *Stud. Mycol.* 89: 117–124.
- Holmer, L. and Stenlid, J. 1997. Competitive hierarchies of wood decomposing basidiomycetes in artificial systems based on variables inoculum sizes. – *Oikos* 79: 77–84.
- Hottola, J. and Siitonen, J. 2008. Significance of woodland key habitats for polypore diversity and red-listed species in boreal forests. – *Biodivers. Conserv.* 17: 2559–2577.
- Kissling, W. D. et al. 2012. Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. – *J. Biogeogr.* 39: 2163–2178.
- Köljal, U. et al. 2013. Towards a unified paradigm for sequence-based identification of fungi. – *Mol. Ecol.* 22: 5271–5277.
- Kubartová, A. et al. 2012. Patterns of fungal communities among and within decaying logs, revealed by 454 sequencing. – *Mol. Ecol.* 21: 4514–4532.
- Lahoz-Monfort, J. J. et al. 2014. Imperfect detection impacts the performance of species distribution models. – *Global Ecol. Biogeogr.* 23: 504–515.
- Lahoz-Monfort, J. J. et al. 2016. Statistical approaches to account for false-positive errors in environmental DNA samples. – *Mol. Ecol. Resour.* 16: 673–685.
- Legendre, P. and Legendre, L. 2012. Numerical ecology. – Elsevier.
- Lou, M. and Golding, G. B. 2012. The effect of sampling from subdivided populations on species identification with DNA barcodes using a Bayesian statistical approach. – *Mol. Phylogenet. Evol.* 65: 765–773.
- Lumley, T. C. et al. 2000. Microscopic ascomycetes isolated from rotting wood in the boreal forest. – *Mycotaxon* 74: 395–414.
- Maynard, D. S. et al. 2018. Species associations overwhelm abiotic conditions to dictate the structure and function of wood-decay fungal communities. – *Ecology* 99: 801–811.
- Meyer, C. P. and Paulay, G. 2005. DNA barcoding: error rates based on comprehensive sampling. – *PLoS Biol.* 3: e42.
- Moore, D. et al. 2008. Fruit bodies: their production and development in relation to environment. – In: Boddy, L. et al. (eds), *Ecology of saprotrophic basidiomycetes*. Elsevier, pp. 79–103.
- Morales-Castilla, I. et al. 2015. Inferring biotic interactions from proxies. – *Trends Ecol. Evol.* 30: 347–356.
- Moruea-Holme, N. et al. 2016. A network approach for inferring species associations from co-occurrence data. – *Ecography* 39: 1139–1150.
- Mäkipää, R. et al. 2017. Interactions between soil- and dead wood-inhabiting fungal communities during the decay of Norway spruce logs. – *ISME J.* 11: 1964–1974.
- Niemelä, T. et al. 1995. Interactions of fungi at late stages of wood decomposition. – *Ann. Bot. Fenn.* 32: 141–152.
- Norberg, A. et al. 2019. Experimentally induced community assembly of polypores reveals the importance of both environmental filtering and assembly history. – *Fungal Ecol.* 41: 137–146.
- Ottosson, E. et al. 2014. Species associations during the succession of wood-inhabiting fungal communities. – *Fungal Ecol.* 11: 17–28.
- Ottosson, E. et al. 2015. Diverse ecological roles within fungal communities in decomposing logs of *Picea abies*. – *FEMS Microbiol. Ecol.* 91: 1–13.
- Ovaskainen, O. and Abrego, N. 2020. Joint species distribution modelling – with applications in R. – Cambridge Univ. Press.
- Ovaskainen, O. et al. 2010a. Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. – *Ecology* 91: 2514–2521.
- Ovaskainen, O. et al. 2010b. Identifying wood-inhabiting fungi with 454 sequencing – what is the probability that BLAST gives the correct species? – *Fungal Ecol.* 3: 274–283.
- Ovaskainen, O. et al. 2013. Data from: Combining high-throughput sequencing with fruit body surveys reveals contrasting life-history strategies in fungi. – *ISME J.* 7: 1696–1709.
- Ovaskainen, O. et al. 2016. Using latent variable models to identify large networks of species-to-species associations at different spatial scales. – *Methods Ecol. Evol.* 7: 549–555.
- Ovaskainen, O. et al. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. – *Ecol. Lett.* 20: 561–576.
- Peres-Neto, P. R. et al. 2001. Environmentally constrained null models: site suitability as occupancy criterion. – *Oikos* 93: 110–120.
- Pollock, L. J. et al. 2014. Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (JSDM). – *Methods Ecol. Evol.* 5: 397–406.
- Porter, T. M. et al. 2008. Fruiting body and soil rDNA sampling detects complementary assemblage of *Agaricomycotina* (Basidiomycota, Fungi) in a hemlock-dominated forest plot in southern Ontario. – *Mol. Ecol.* 17: 3037–3050.
- Pouska, V. et al. 2013. Co-occurrence patterns of wood-decaying fungi on *Picea abies* logs: does *Fomitopsis pinicola* influence the other species? – *Polish J. Ecol.* 61: 119–133.
- Purhonen, J. et al. 2017. Detailed information on fruiting phenology provides new insights on wood-inhabiting fungal detection. – *Fungal Ecol.* 27: 175–177.
- Renvall, P. 1995. Community structure and dynamics of wood-rotting Basidiomycetes on decomposing conifer trunks in northern Finland. – *Karstenia* 35: 1–51.
- Royal Botanic Gardens Kew et al. 2019. Index fungorum. – <www.indexfungorum.org/>.
- Runnel, K. et al. 2015. Surveying wood-inhabiting fungi: most molecularly detected polypore species form fruit-bodies within short distances. – *Fungal Ecol.* 18: 93–99.
- Saine, S. et al. 2020. Data from: Data collected by fruit body- and DNA-based survey methods yield consistent species-to-species association networks in wood-inhabiting fungal communities. – Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.tmpg4f4wm>.
- Somervuo, P. et al. 2016. Unbiased probabilistic taxonomic classification for DNA barcoding. – *Bioinformatics* 32: 2920–2927.

- Somervuo, P. et al. 2017. Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding. – *Methods Ecol. Evol.* 8: 398–407.
- Steele, J. A. et al. 2011. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. – *ISME J.* 5: 1414–1425.
- Tedersoo, L. et al. 2014. Global diversity and geography of soil fungi. – *Science* 346: 1256688.
- Tikhonov, G. et al. 2020. Joint species distribution modelling with the R-package Hmsc. – *Methods Ecol. Evol.* 11: 442–447.
- Tiunov, A. V. and Scheu, S. 2005. Facilitative interactions rather than resource partitioning drive diversity–functioning relationships in laboratory fungal communities. – *Ecol. Lett.* 8: 618–625.
- Tyre, A. J. et al. 2003. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. – *Ecol. Appl.* 13: 1790–1801.
- Urban, M. C. et al. 2016. Improving the forecast for biodiversity under climate change. – *Science* 353: aad8466-1–aad8466-9.
- van der Linde, S. et al. 2012. Now you see it, now you don't: the challenge of detecting, monitoring and conserving ectomycorrhizal fungi. – *Fungal Ecol.* 5: 633–640.
- Warton, D. I. et al. 2015. So many variables: joint modeling in community ecology. – *Trends Ecol. Evol.* 30: 766–779.
- Weiss, S. et al. 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. – *ISME J.* 10: 1669–1681.
- White, T. J. et al. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. – In: Innis, M. A. et al. (eds), *PCR protocols: a guide to methods and applications*. Academic Press, pp. 315–322.
- Wirta, H. K. et al. 2014. Complementary molecular information changes our perception of food web structure. – *Proc. Natl Acad. Sci. USA* 111: 1885–1890.
- Wisz, M. S. et al. 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. – *Biol. Rev.* 88: 15–30.
- Yoccoz, N. G. et al. 2001. Monitoring of biological diversity in space and time. – *Trends Ecol. Evol.* 16: 446–453.
- Zhu, C. et al. 2019. Seeing is believing? Comparing plant–herbivore networks constructed by field co-occurrence and DNA barcoding methods for gaining insights into network structures. – *Ecol. Evol.* 9: 1764–1776.
- Zurell, D. et al. 2018. Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments? – *Ecography* 41: 1812–1819.