New Phytologist

Citizen science data reveal ecological, historical and evolutionary factors shaping interactions between woody hosts and wood-inhabiting fungi

Jacob Heilmann-Clausen[1], Pietro K. Maruyama[1,2], Hans Henrik Bruun[3], Dimitar Dimitrov[4], Thomas Læssøe[1,3], Tobias Guldberg Frøslev[5] and Bo Dalsgaard[1]

[1]Centre for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, Copenhagen DK-2100, Denmark; [2]Departamento de Biologia Vegetal, Universidade Estadual de Campinas (UNICAMP), Cx. Postal 6109, Campinas, SP CEP: 13083-862, Brazil; [3]Department of Biology, University of Copenhagen, Copenhagen DK-2100, Denmark; [4]Natural History Museum, University of Oslo, PO Box 1172 Blindern, Oslo NO-0318, Norway; [5]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen DK-1350, Denmark

Author for correspondence:
Jacob Heilmann-Clausen
Tel: +45 51330991
Email: jheilmann-clausen@snm.ku.dk

Received: *27 April 2016*
Accepted: *8 August 2016*

New Phytologist (2016) **212:** 1072–1082
doi: 10.1111/nph.14194

Key words: ecological networks, endophytes, host selection, modularity, species richness, wood chemistry.

Summary

- Woody plants host diverse communities of associated organisms, including wood-inhabiting fungi. In this group, host effects on species richness and interaction network structure are not well understood, especially not at large geographical scales.
- We investigated ecological, historical and evolutionary determinants of fungal species richness and network modularity, that is, subcommunity structure, across woody hosts in Denmark, using a citizen science data set comprising > 80 000 records of > 1000 fungal species on 91 genera of woody plants.
- Fungal species richness was positively related to host size, wood pH, and the number of species in the host genus, with limited influence of host frequency and host history, that is, time since host establishment in the area. Modularity patterns were unaffected by host history, but largely reflected host phylogeny. Notably, fungal communities differed substantially between angiosperm and gymnosperm hosts.
- Host traits and evolutionary history appear to be more important than host frequency and recent history in structuring interactions between hosts and wood-inhabiting fungi. High wood acidity appears to act as a stress factor reducing fungal species richness, while large host size, providing increased niche diversity, enhances it. In some fungal groups that are known to interact with live host cells in the establishment phase, host selectivity is common, causing a modular community structure.

Introduction

Trees are fundamental components of forested ecosystems. As architectonically complex organisms, they provide habitats or food for a wide range of associated biota, including mycorrhizal fungi (Smith & Read, 2008), fungal pathogens and endophytes (Sieber, 2007), epiphytes (Barkman, 1958; Sáyago *et al.*, 2013), phytophagous insects (Southwood, 1961), pollinators (Ollerton *et al.*, 2011) and seed- and fruit-eating vertebrates (Jordano, 2013). Furthermore, dead leaves and wood support a wealth of organisms, not least saprotrophic fungi and insects (Boddy *et al.*, 2008; Stokland *et al.*, 2012). Traditionally, studies have focused on how trees as hosts affect species richness in associated biota. Most evidence points towards host size, range distribution and local abundance as the most important drivers (Brändle & Brandl, 2001; Miller, 2012; Kamiya *et al.*, 2014).

Within the last few decades, the focus has shifted towards analyzing interactions between hosts and associated biota not only as patterns of richness, but also as networks of species interactions (Jordano, 1987). The network approach enables a deeper analysis of community structure, such as how species partition interactions and form 'modules', that is, weakly connected subcommunities that are internally highly interlinked (e.g. Olesen *et al.*, 2007). Studies of biotic networks have shown that network structure may be driven by a mix of ecological, historical and co-evolutionary processes (e.g. Dalsgaard *et al.*, 2013; Martín González *et al.*, 2015). For instance, co-evolutionary processes may lead to modular network structures reflecting host or mutualist phylogeny (Rezende *et al.*, 2009; Donatti *et al.*, 2011; Bahram *et al.*, 2014; Martín González *et al.*, 2015), and ecological networks have been shown to be more modular in productive and historically climatically stable environments (Dalsgaard *et al.*, 2013). The network approach thus offers a way of examining and

capturing patterns of species interactions that is complementary to and more holistic than host specialization and species richness analyses.

Among plant-associated fungi, host specialization is pronounced and arguably a main driver of fungal biodiversity worldwide (Hawksworth, 2001), as well as in local fungal communities (e.g. Tedersoo *et al.*, 2008; Unterseher *et al.*, 2008). Fungi are involved in many types of interactions with plants, but most of the plant–fungal interaction data sets analyzed to date have focused on mycorrhizal fungi (e.g. Bahram *et al.*, 2014; Toju *et al.*, 2014). Some recent network studies have broadened this perspective by including endophytic and endolichenic fungi (e.g. Zhang & Yao, 2015; Chagnon *et al.*, 2016; Vincent *et al.*, 2016), but still many of the interactions between fungi and other organisms remain unexplored from a network perspective. This is true also for saprotrophic fungi, although they are the main agents of plant litter decomposition in terrestrial ecosystems world-wide (Boddy *et al.*, 2008). Dead wood constitutes the bulkiest type of plant litter, and provides a habitat for highly distinctive and species-rich fungal communities. The communities show varying degrees of host specialization, depending on forest type and local tree species richness (e.g. Gilbert & Sousa, 2002; Heilmann-Clausen *et al.*, 2005). However, apart from a well-known discrepancy between angiosperms and gymnosperms, knowledge of overall host selection patterns in wood-inhabiting fungi is surprisingly limited (Stokland *et al.*, 2012, p. 82). We thus have a limited insight into the factors driving patterns of species association between woody plants and wood-inhabiting fungi.

Wood-inhabiting fungi live in habitat-tracking metapopulations and depend on repeated successful recolonization of suitable hosts, which may happen after the host has died or while it is still alive (Boddy & Heilmann-Clausen, 2008). The latter is the case for heart rot fungi, wound parasites and endophytes with a latent decay strategy, which all interact with living host cells in the colonization phase. Previous observations have indicated host specialization to be most common in these groups, probably as a result of co-evolutionary processes (Boddy & Heilmann-Clausen, 2008). In parasitic or pathogenic insects and fungi that similarly interact with living host cells, host specialization has been found to be most prevalent on apparent, that is, abundant, and/or large hosts, probably because these have to invest more in defense systems to avoid attack from detrimental generalists (Brändle & Brandl, 2001; Parker *et al.*, 2015). In woody plants, a high lignin content, low pH and low macronutrient contents have been identified to reduce the activity of decay fungi (Weedon *et al.*, 2009; Freschet *et al.*, 2012). However, several wood-degrading enzymes are most efficient at low pH (Baldrian, 2008) and, based on laboratory experiments, it is traditionally assumed that wood-decay fungi generally have low pH optima (e.g. Kollmann & Côté, 1968; Zabel & Morell, 1992). Hence, the importance of these wood traits for fungal richness is unclear. Host size is likely to influence richness in associated fungi in several ways. Notably, large hosts provide greater microhabitat diversity and are more predictable habitats with higher longevity, both as alive and as dead hosts. Both factors are likely to affect species richness positively (Heilmann-Clausen & Christensen, 2004). Host frequency and time since establishment in an area are assumed to have similar effects, as prevalent hosts with a long local history are more likely to have established associations with – and sustain populations of – associated fungal taxa (cf. Brändle & Brandl, 2001; Heilmann-Clausen *et al.*, 2005).

The lack of knowledge of host selection patterns in wood-inhabiting fungi is partly attributable to a lack of quantitative host association data sets on large geographic scales. For plant pathogenic fungi, national inventory data are important sources, which have been used in a few studies on host richness and interaction patterns. Miller (2012) studied host richness patterns for plant pathogens across the USA, while Vacher *et al.* (2008, 2010) investigated interactions between woody hosts and their parasitic fungi in France. In both studies, a limited number of wood-inhabiting species damaging living tissues were included. For the majority of wood decomposing fungi, however, national inventory data are not available.

Citizen science offers an alternative approach for collecting large-scale and long-term data sets in ecology and environmental sciences, which may otherwise be prohibitively expensive (Silvertown, 2009; Tulloch *et al.*, 2013). For wood-inhabiting fungi and their host selection, data reported by amateur mycologists provide a rich data source useful for studying host specialization (Gange *et al.*, 2011).

Here, we used a Danish nationwide citizen science data set to explore host associations between 91 genera of woody host plants and 1085 species of associated wood-inhabiting fungi. We explored the importance of the woody host for fungal species richness and the associated network structure. Specifically, we tested the influence of host traits (i.e. architecture and wood chemistry), host frequency, time since establishment in Denmark and host phylogeny on fungal richness and modules within the countrywide network of host–fungal interactions. Analyses were conducted on the full data set and on data sets reduced and standardized to control for sampling effort. We expected fungal species richness per host genus to be related to wood chemistry, host size and time since host establishment. Further, we expected host–fungal networks to have a distinct modular structure reflecting host and fungal phylogeny which in turn reflects co-evolutionary adaptations between hosts and associated decay fungi establishing in living hosts.

Materials and Methods

Plant–fungal interaction network

Data on wood-inhabiting fungi were extracted from the Danish Fungal Atlas database (Danish Mycological Society, 2014), using 'bark' and 'wood' as search terms, and accepting only validated records. A total of 110 712 records were extracted, of which 83 637 remained after quality control, omitting records with uncertain host information and species known to be obligate ectomycorrhizal or bryophyte-associated, or to have their main habitat on humus or leaf litter (cf. Hansen & Knudsen, 1992–2000 and Knudsen & Vesterholt, 2012). Data were highly skewed, with *Fagus* being the host with most fungal records

(29 633 records for 467 fungal species), while 20 host genera had only one fungal species recorded.

The Danish Fungal Atlas was a 5-yr citizen science project running from 2009 to 2013 with the aim to collect data on the distribution and ecology of all fruit body-forming Basidiomycota in Denmark, but open also to records of Ascomycota and other fungal groups. The project was carried out as a collaboration between the Natural History Museum of Denmark, the Department of Biology at the University of Copenhagen, the Danish Mycological Society and MycoKey (http://www.mycokey.com). The project involved a rigorous validation process for both fungal species identification and host data. All fungal species were coded with requirements for validation (e.g. description of smell or taste, photograph or dried voucher specimen), and several thousand dried specimens were sent by volunteers to be validated or re-identified by professional experts. In total, > 400 active users contributed to the project, of which *c.* 100 can be described as core contributors, each supplying > 100 records. Plant host information was generally recorded only at the genus level by participating citizens, whereas fungi were recorded at species level, and these taxonomic levels were used in all analyses. In practice, only a limited number of host genera contain more than one or two species in Denmark, and thus, with the exception of the genera *Populus*, *Prunus* and *Salix*, effects of lumping species within genera can be considered negligible. The full list of recorded host genera and species is given in Supporting Information Table S1.

Based on the extracted data, we constructed a quantitative and a binary matrix of host plant−fungus interactions. The quantitative matrix summarized the number of records for each fungal species across all host genera in the data set, while the binary matrix scored presence–absence of these interactions. The matrices illustrate the network of interactions of the entire plant–fungus assemblage observed within Denmark.

Scoring of host traits

Data on host plant attributes were compiled from the literature, except for host frequency which was extracted from the database of the National Forest Inventory for Denmark (Table S1). The number of species per host genus was taken from Ødum (1968), while maximum height and maximum diameter at breast height (DBH; 1.3 m above ground) were mainly obtained from Møller & Staun (2001) and used as proxies for architectural complexity (in accordance with Brändle & Brandl, 2001). Wood physical and chemical trait data, such as wood density (oven-dried wood), lignin percentage and wood pH, were compiled from various sources (most records from Wagenführ & Scheiber, 1989; Brzeziecki & Kienast, 1994) and taken to directly characterize the environment of mycelia in dead wood. In the cases of host genera with more than one species present in Denmark, and if constituent species differed in their trait values, the most abundant species was taken to represent the whole genus. Data on maximum DBH, wood density, lignin and cellulose content and wood pH were missing for some host genera, and hence these traits were not included in all analyses.

Phylogenetic signal in host traits and comparative analyses

We extracted the topology for our set of host genera from the dated phylogeny of Zanne *et al.* (2013). This phylogeny was preferred over alternatives, for example, a phylogeny built using the PHYLOMATIC tool (http://phylodiversity.net/phylomatic/html/pm1.html), because it is based on an analysis of actual data for each of the tips and carries more accurate information on branch lengths. Only three host taxa present in our data set were not included in the phylogenetic tree (*Cytisus*, *Mahonia*, and *Symphoriocarpos*) and these were excluded from further analyses (the three host genera had 62, one and nine records of wood-inhabiting fungi in our data set, respectively, constituting < 0.1% of all records). The resulting ultrametric tree was used to test whether evolutionary history (phylogeny) could predict similarities in traits recorded for each genus. To test for phylogenetic signal, we calculated Pagel's lambda (λ) (Pagel, 1999) using the *pgls* function in the R package CAPER (v.0.5.2; Freckleton *et al.*, 2002). Pagel's λ is a scaling parameter of the phylogeny for the correlations between species, relative to the correlation expected under the Brownian motion model of evolution. Values vary from 0 to 1, where low values indicate a weak or no phylogenetic signal and 1 indicates a strong phylogenetic signal. There are several other metrics that can be used to measure phylogenetic signal based on different approaches (for a review and comparison of their performances, see Diniz-Filho *et al.*, 2012 and Münkemüller *et al.*, 2012). Here, we used Pagel's λ because it is a model-based approach and has been shown to suffer from a lower rate of type I error (and a low rate of type II error) for phylogenies of varying sizes (Münkemüller *et al.*, 2012).

For the fungal data set, we used a simple proxy of phylogeny, and scored the assumed taxonomic position of each species at phylum and order levels, based on information in Index Fungorum (http://www.indexfungorum.org) extracted on 7 July 2015. This approach was chosen because of the large amount of included taxa which remain to be studied phylogenetically, but also with the aim of yielding more general results that are easier to comprehend.

Fungal species richness analyses

The fungal data set was quantitative, but, given its origin in a voluntary citizen science project, the recorded fungal abundance partly reflects the activity of recorders, favoring fungi with conspicuous fruit bodies and widely distributed and common host genera. To account for this, we explored the association patterns based on the raw data as well as on data standardized for sampling effort. More specifically, we assessed the importance of plant hosts for fungal species richness using three different approaches. First, we analyzed the full quantitative matrix comprising all fungal records remaining after quality control (i.e. 83 637 records). Second, we analyzed the data standardizing for sampling effort. We did this by comparing similar numbers of fungal records for each host genus based on individual-based rarefaction and extrapolation (Colwell *et al.*, 2012). Both procedures produced unstable rankings when only a few records were

included in the analyses, and hence we omitted 66 host genera with < 100 fungal records to allow a robust comparison of host richness. After exploring several cut-off levels, we found the 100 records level to represent the best compromise between inclusion of host diversity and a representative sampling effort. The standardized data included 25 host genera, of which seven had between 100 and 300 records. For these, fungal species richness was extrapolated to 300 records using the function described by Colwell *et al.* (2012). Both rarefaction and extrapolations were performed in EstimateS 9.1 (Colwell, 2013). Third, we analyzed species richness based on a reduced data set, containing the same 25 host genera as included in the standardization procedure described above, but not involving rarefaction and extrapolation. This was done to test the effects of rarefaction separately from the effect of reduction in the number of host genera. Details of the data involved in the three procedures are summarized in Table 1. We also considered the possibility of using nonparametric species richness estimators (e.g. Unterseher *et al.*, 2008), but these appeared to provide unstable predictions, biased by sampling effort (Fig. S1). Hence, we preferred our more conservative approach.

Fungal species richness was modeled as a function of each host trait as a single predictor, and subsequently in a multiple regression model including all host traits with complete information (time since host establishment, host frequency, maximum height and number of species in genus). In all analyses, we modeled fungal species richness for all three data sets using phylogenetic least square models (PGLS), applying a log link function in the R package CAPER v.0.5.2 (Freckleton *et al.*, 2002).

Modularity analysis

To test for host selection patterns we used a network approach, focusing on the detection of modules within the host–fungal network data (Newman, 2004; Guimera & Amaral, 2005; Olesen *et al.*, 2007). Modules are defined as subunits of highly connected nodes within a network, and can be detected using an optimization algorithm that maximizes modularity (Guimera & Amaral, 2005; Marquitti *et al.*, 2014). We used simulated annealing as the optimization algorithm and calculated a modularity metric appropriate for a bipartite matrix (Barber, 2007; Marquitti *et al.*, 2014). The simulated annealing algorithm is stochastic and, hence, module arrangement may vary between runs. Thus, we retained the module conformation with the highest modularity value (Q) as the optimum after 30 independent runs (Marquitti *et al.*, 2014). The significance of the observed level of modularity was contrasted with two null models using a permutation test with 100 iterations (Marquitti *et al.*, 2014). The modules and the modularity metrics were computed using the software MODULAR (Marquitti *et al.*, 2014).

As in the analyses for species richness, modularity was analyzed based on three different approaches (Table 1). The full and reduced data sets were identical in the two analyses, but modularity was analyzed using binary data, in order to reduce bias from some fungal species groups (e.g. polypores) being much more sampled than others. In order to control for variable sampling effort among hosts, we constructed 10 standardized data sets, by randomly selecting 100 host–fungus interaction records from the full population of interactions for each host genus with at least 100 fungal records. Each of these was combined in a presence–absence matrix that was subjected to independent analysis of modularity.

The phylogenetic signal behind modularity in the hosts was tested using the function in R developed by Donatti *et al.* (2011) that implements the 'Fixed Tree, Character Randomly Reshuffled' model of Maddison & Slatkin (1991). In this approach, the minimum number of transitions along the phylogeny that results in the observed distribution of module identities is calculated. This is followed by a randomization of the module identities on the topology and an optimization of the number of transitions for the randomized data set. We repeated this randomization step 999 times and then compared the number of transitions in the observed data set with that in the randomized data. If the minimum number of transitions was higher than that in the observed data in at least 95% of the randomizations, a phylogenetic signal was considered significant.

Associations between host modules and host traits were explored using the Kruskal–Wallis test, while the phylogenetic signal (at phylum and order levels, nested within phyla) in fungal modules was explored using contingency table analyses. In these tests, orders represented by < 15 species were grouped as 'different' in order not to violate model assumptions. For the same reason, the fungal phylogenetic signal behind modules based on the standardized data sets was only tested at the phylum level. Kruskal–Wallis and contingency tests were computed in JMP v.12 (SAS Institute, Cary, NC, USA).

Results

Phylogenetic signal and sampling effort

Only two of the included host traits, maximum height and wood density, showed a significant phylogenetic signal based on Pagel's λ (Table 2). In the full data set, several host

Table 1 Overview of data sets used in the analyses for species richness and modularity

Data set	Full	Reduced	Standardized
Number of host genera	89	25	25
Number of fungal records	83.637/5.052[1]	82.739/4.603[1]	82.739/2.500[1]
Number of fungal species	1069	1044	1044/399–429[1]

[1]Values to the left show data properties in the data sets used to analyze species richness, while values to the right show data properties for the data sets used to analyze modularity.

variables were strongly correlated with sampling effort, and some collector bias remained in the reduced data set (omitting hosts with < 100 records), favoring hosts with large dimensions (maximum DBH) or high frequency in the landscape (Table 2). The number of submitted fungal records varied considerably among volunteers (from one to 7781 records), and for some host genera and fungal orders a majority of records stemmed from rather few volunteers. However, this was not judged to affect data reliability as hosts and orders burdened by unequal sampling were in all cases mainly recorded by volunteers with a very broad taxonomic scope in both dimensions (Fig. S2; Table S2).

Fungal species richness

In the full data set, fungal species richness was significantly positively related to host frequency, time since host establishment, maximum height and number of species in the host genus, but negatively related to lignin percentage, based on the phylogenetically controlled one-way regressions (Table 3). The models based on the reduced data set produced qualitatively similar results, except for a greater effect of time since host establishment, a smaller effect of host frequency, and no association with lignin percentage. By contrast, the models based on rarified richness estimates showed no effects of host frequency and history, but a positive effect of wood pH, number of species in host genus, and the two host size variables, that is, maximum height and DBH (Table 3). None of these variables were significantly correlated, and interaction terms were insignificant (results not shown). In the phylogenetically controlled multiple regression models for species richness, the amount of explained variation decreased steeply from 81% in the full data set to 12% in the standardized data set (Table 4). Host phylogeny contributed to explaining differences in species richness only in the full data set.

Modularity

All networks were modular ($P_{null1} < 0.01$; $P_{null2} < 0.01$), with modularity values ranging from $Q = 0.29$ in the reduced data set to $Q = 0.30$ in the full data set to $0.41 < Q < 0.43$ in the 10 standardized matrices. Six modules were detected in the full data set, five in the reduced data set, and four to 10 in the standardized matrices. In all cases, modular patterns were significantly related to host phylogeny (Figs 1a, S3). The most consistent module across all matrices contained Pinaceae, but a phylogenetic signal was also evident within the angiosperms. In the full data set, Fagaceae, Betulaceae, Salicaceae and Rosaceae showed distinct clustering, but this was not evident in the standardized data sets, which generally showed unstable modular structures within the angiosperms, although with a tendency for consistent links within and between Betulaceae and Rosaceae, between *Acer* and *Ulmus*, and between *Aesculus* and *Fagus*. Relationships between modules and host traits were most significant in the full data set, with host frequency, maximum height and wood density showing the strongest associations. In the reduced data set, wood density was the only host trait significantly related to the modular structure, while four traits showed a modular signal in at least one of the standardized data sets (Table 5).

The fungal association with modules showed a significant phylogenetic signal in both the full and reduced data sets, at the levels of phylum and order, nested within phyla (Fig. 1b). Ascomycota were consistently underrepresented in the modules defined by gymnosperms. Within Ascomycota, the same modules consistently had no or very few representatives from the Diaporthales and Xylariales (Fig. 2a,b), while Helotiales and Pezizales (Fig. 2d,e) were relatively overrepresented. For other modules, phylogenetic patterns were less consistent; however, Hypocreales (Fig. 2c) were overrepresented in modules defined by family Fagaceae (full data), or the genus *Fagus* alone (reduced data set). Similarly, Diaporthales (Fig. 2a) appeared to be overrepresented in the modules defined by Rosaceae.

Table 2 Host variables scored in this study, and a summary of degrees of freedom (df), and tests for phylogenetic signal for each variable using Pagel's lambda (full data set) and for sampling effort based on Pearson correlation (both data sets)

	Full data set			Reduced data set[1]	
Variable	λ[2]	df[3]	Pearson *r* sampling effort[4]	df[3]	Pearson *r* sampling effort[4]
Host frequency	$0^{< 0.001,\ ns}$	85	0.62***	25	0.53**
Time since host establishment (yr)	$0^{< 0.001,\ ns}$	85	0.21	25	0.03
Number of species in genus	$0^{< 0.001,\ ns}$	85	0.18	25	−0.01
Maximum height (m)	$\mathbf{0.364}^{< 0.001,\ < 0.05}$	85	0.41***	25	0.31
Maximum DBH (cm)	$0^{< 0.001,\ ns}$	41	0.44**	23	0.54**
Wood density	$\mathbf{0.726}^{< 0.01,\ < 0.05}$	36	0.04	22	0.13
Lignin content (%)	†$0.194^{< 0.001,\ ns}$	24	−0.57**	18	−0.55*
Cellulose content (%)	$0^{< 0.001,\ ns}$	24	−0.13	18	−0.13
Wood pH	$0^{ns,\ ns}$	28	0.05	18	0.04

[1]The standardized data set has zero correlation with sampling effort, and the same number of df as in the reduced data set.
[2]Superscripts denote likelihood ratio tests for differences from 1 and 0; [ns]not significant. Cases where λ is significantly different from 0 (phylogentic signal is present) are marked in bold; cases where λ is > 0 but not significantly so are denoted with †.
[3]The degree of freedoms vary, reflecting incomplete trait information for several host genera.
[4]Significance levels are indicated as: ***, < 0.0001; **, 0.001 to < 0.01; *, 0.01 to < 0.05.

Table 3 Summary of one-way regression models for fungal species richness based on various host variables when controlling for phylogeny (*pgls*)

	Full data set				Reduced data set				Standardized data set			
Variable	F	R^2	λ^2	P^3	F	R^2	λ^2	P^3	F	R^2	λ^2	P^3
Host frequency	154.3	0.64	†0.11$^{<0.001,ns}$	<0.001***	4.8	0.14	0$^{<0.001,\ ns}$	<0.05**	2.3	0.04	0$^{<0.001,ns}$	ns
Time since host establishment	77.8	0.47	**0.59$^{<0.001,<0.05}$**	<0.001	14.5	0.36	**0.97$^{ns,<0.01}$**	<0.001	0.005	−0.03	*0.3$^{<0.001,ns}$	ns
Number of species in genus	37.7	0.3	†0.54$^{<0.001,ns}$	<0.001	1.2	0.01	0$^{<0.001,ns}$	ns	8.3	0.2	0$^{<0.001,ns}$	<0.01
Maximum height	79.1	0.48	**0.34$^{<0.001,<0.01}$**	<0.001***	11.4	0.3	0$^{<0.001,<\ ns}$	<0.01	8.1	0.19	0$^{<0.001,<\ ns}$	<0.01
Maximum DBH[1]	1.4	0.01	0.002$^{<0.001,ns}$	ns**	10.4	0.3	0$^{<0.01,ns}$	<0.01**	8.8	0.25	0$^{<0.001,ns}$	<0.01
Wood density[1]	1.4	0.01	**0.63$^{<0.01,<0.05}$**	ns	1.2	0.01	*0.60$^{<0.01,ns}$	ns	0.04	−0.04	*0.64$^{<0.01,ns}$	ns
Lignin content[1]	−9.9	0.27	†0.22$^{<0.01,ns}$	<0.01**	2.7	0.09	*0.23$^{<0.001,ns}$	ns*	2.8	0.09	0.08$^{<0.001,ns}$	ns
Cellulose content[1]	0.01	−0.04	0$^{<0.001,ns}$	ns	0.1	−0.05	0$^{<0.05,ns}$	ns	0.03	−0.05	0$^{<0.001,ns}$	ns
Wood pH[1]	2.1	0.04	1$^{ns,\ ns}$	ns	0.6	−0.02	0$^{<0.01,ns}$	ns	12.9	0.38	0$^{<0.05,ns}$	<0.01

DBH, diameter at breast height; ns, not significant.
[1]Variables with lower dfs because of incomplete data.
[2]Cases where λ is significantly different from 0 (phylogenetic signal is present) are marked in bold; cases where λ is > 0 but not significantly so are denoted with †.
[3]Superscripts indicate the Pearson correlation coefficient with sampling effort, with significance levels indicated as: ***, < 0.0001; **, 0.001 to < 0.01; *, 0.01 to < 0.05.

Within the Basidiomycota, the most consistent pattern was the strong overrepresentation of Boletales (Fig. 2h) within the modules defined by gymnosperms, which, in contrast, had very few representatives of the Corticiales (Fig. 2i).

Discussion

Fungal species richness

With data standardized for sampling effort, we found that fungal species richness was positively related to host size (maximum height and DBH), wood pH and the number of species in the host genus, whereas no effect of host frequency or time since host establishment was detected. In other words, fungal species richness increased with host size, but was lower for hosts with acidic wood and for those that are isolated taxonomically.

The positive effect of host size may reflect higher microhabitat diversity or more neutral colonization/extinction processes, that is, higher passive sampling and fewer stochastic extinctions on larger hosts (cf. Brändle & Brandl, 2001). We cannot effectively disentangle the relative contributions of these factors, but the surprising lack of signal from other neutral factors, that is, host frequency and time since host establishment, in the bias-controlled analysis suggests that higher microhabitat diversity is the main driver. This makes sense ecologically: all woody plants produce twigs and small stems, but only larger trees produce thick branches and trunks, supporting fungal species that are unable to thrive in smaller pieces of dead wood because of large mycelia or special microclimatic requirements (Heilmann-Clausen & Christensen, 2004; Abrego & Salcedo, 2013).

The positive relationship between wood pH and fungal richness was independent of phylogeny, indicating that acidic wood (e.g. *Quercus* and *Picea*), as opposed to more alkaline wood (e.g. *Fraxinus* and *Abies*), acts as an important filter limiting species from a large pool of unspecialized wood-associated fungi. It is well known that pH optima differ among wood-inhabiting fungi, with most species traditionally indicated to have optima at rather low pH values (e.g. Kollmann & Côté, 1968; Zabel & Morell, 1992). In the light of this, the overall positive effect of wood pH on species richness in this study is surprising, but may simply reflect the fact that most studies on wood pH preferences have dealt with a very limited number of decomposers known as pests in forests or unwanted degraders in timber, rather than full communities of fungi inhabiting dead wood in natural habitats. A recent study found initial wood pH to be a potentially strong positive predictor of decay rates in dead wood (Freschet *et al.*, 2012), supporting wood acidity as a stress factor limiting fungal decay and species richness.

Finally, the number of species in the host genus has been found to influence the richness of phytophagous insects (but see Neuvonen & Niemelä, 1981), but – to our knowledge – not previously that of fungi. The number of species in a genus can be used as a proxy for taxonomic isolation (cf. Brändle & Brandl, 2001), and the effect might reflect taxonomic isolation per se, that is, less sharing of fungal species with more distant relatives, or higher niche diversity within the actually sampled hosts. For

Table 4 Multiple regression models of fungal species richness with host frequency, time since host establishment, number of species per genus and maximum height as input variables and controlling for phylogeny (*pgls*)

	R-squared	λ	Residual standard error	*F*	*P* value
Full data set	0.81	**0.881**$^{< 0.001, < 0.001}$	0.077 (df = 81)	99.44	< 0.001
Reduced data set	0.62	0.791$^{ns, ns}$	0.029 (df = 20)	10.64	< 0.001
Standardized data set	0.12	0$^{ns, ns}$	0.019 (df = 26)	1.988	ns

df, degrees of freedom; ns, not significant. Cases where λ is significantly different from 0 are marked in bold.

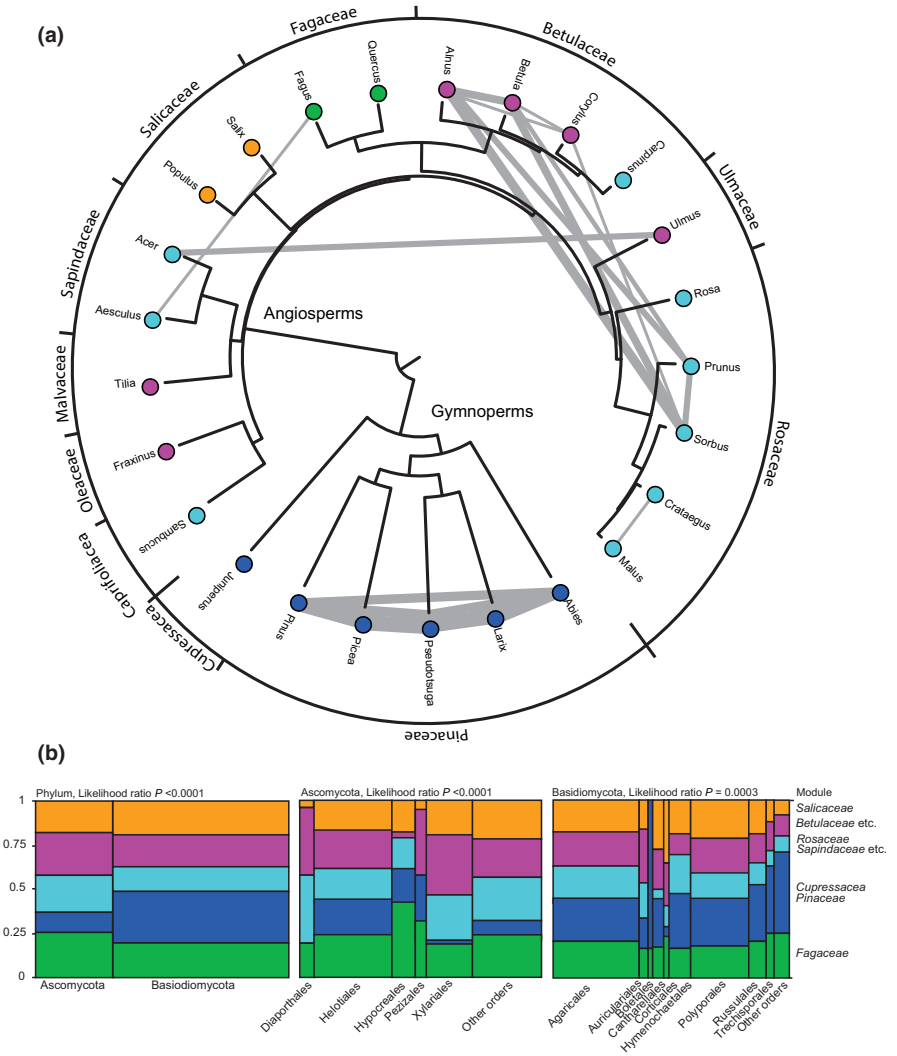

Fig. 1 (a) Phylogeny for the woody hosts included in the standardized data sets, with color codes reflecting the association with the five modules detected in the full data set, and the thickness of gray connecting lines illustrating the consistency of this modular configuration. The thickest lines indicate hosts occurring in the same module in 10 out of 10 random data sets, while the thinnest lines indicate hosts occurring in the same module in five out of 10 random data sets. Note the consistent position of the gymnosperm genera within the same module. (b) Mosaic plots showing the relative proportion of fungal species associated with the five host modules detected in the full data sets, at phylum level and order level nested within the two phyla Ascomycota and Basidiomycota. The width of columns is scaled to reflect the number of species across all modules that belong to each phylum or order, while the height of each tile in the columns is scaled to illustrate the proportion of species within each phylum/order that is associated with each module. The most important plant families in each host module are highlighted at the right.

instance, the host genera *Salix* and *Prunus*, represented by eight and eleven species in our data set, respectively, include both scrubs and smaller trees with variable ecology, which potentially could increase the fungal species pool at genus level.

Fungal community structure

In the analyses of fungal community structure, based on interaction modularity, a signal of host phylogeny was consistent in all data sets. Among the investigated traits, maximum size (DBH/height) and wood density were most consistently related to modularity, the latter being unaffected by sampling effort even in the full data set. As discussed below, these host traits were significantly related to host phylogeny. No signal of time since host establishment in Denmark on the modular structure was detected in any data set, which supports the finding by Vacher *et al.* (2010) that recently introduced woody hosts are easily integrated into existing host–fungus networks as a result of a high number of species with broad host selection.

Considering host phylogeny, the clear split between angiosperms and gymnosperms was the most consistent pattern across all data sets. It reflects the major split in plant evolution,

Table 5 Summary of Kruskal–Wallis tests for independence between modular structure and host traits

Data set Variable	Full P^2	Reduced P^2	Standardized *P* (max)	 No[3]
Host frequency	0.0004***	0.21**		0
Time since host establishment	0.002	0.13		0
Number of species in genus	0.2	0.51		0
Maximum height	0.0004***	0.14	0.02	6
Maximum DBH[1]	0.04**	0.26**		0
Wood density[1]	0.0004	0.03	0.01	3
Lignin content[1]	0.18**	0.12*	0.03	1
Cellulose content[1]	0.46	0.49		0
Wood pH[1]	0.37	0.57	0.02	1

[1]Variables with lower dfs as a result of incomplete data.
[2]Superscripts indicate the Pearson correlation coefficient with sampling effort, with significance levels indicated as: ***, < 0.0001; **, 0.001 to < 0.01; *, 0.01 to < 0.05. (For details, see Table 1.)
[3]'No' denotes the number of random data sets (out of 10) that showed a significant trait signal.
DBH, diameter at breast height.

and is in accordance with previous observations (e.g. Küffer *et al.*, 2008; Stokland *et al.*, 2012), which also indicated Ascomycota to be most prominent in the decay of angiosperm wood, as found in our study. This could reflect filtering as a result of differences in wood density and chemistry among angiosperms and gymnosperms (Weedon *et al.*, 2009), but we did not find consistent effects of related wood traits on modular structure. Hence, biotic interactions facilitating co-evolutionary specialization appear to be more important, as previously suggested for heart rot fungi, parasites and endophytes that all interact with living host cells (Boddy & Heilmann-Clausen, 2008).

Among ascomycote endophytes, Sieber (2007) observed that Helotiales dominated communities in gymnosperms, while Diaporthales dominated in angiosperms, and linked this to the concurrent divergence of the relevant plant and ascomycote lineages in the Carboniferous period, some 300 million yr ago. Our results support this hypothesis for Diaporthales but also link the Xylariales, another order rich in fungal endophytes, tightly to the angiosperms. By contrast, our data indicate broader host selection in the Helotiales, which might reflect the greater taxon sampling in our study, or methodological differences (fruit body records versus mainly isolation of fungal cultures in Sieber (2007). Within the Basidiomycota, the close link between wood-decomposing Boletales and gymnosperms was previously emphasized by Binder & Hibbett (2006), who suggested that the specific brown rot type in the Boletales may have evolved as a response to the high lignin contents, special lignin types and secondary antibiotic compounds in the gymnosperms (cf. Weedon *et al.*, 2009).

Within the angiosperms, our results suggest Diaporthales to be especially prominent within the modules containing Betulaceae and most Rosaceae, while Hypocreales showed a high prevalence within Fagaceae. Whether this reflects co-evolutionary dynamics needs to be explored further, but it is worth noting that both orders are rich in endophytes and plant pathogens interacting with live host cells, while Hypocreales in addition includes many mycoparasites on wood decay fungi (Rossman *et al.*, 2007; Jaklitsch, 2009; Chaverri & Samuels, 2013).

Controlling for sampling effort

This study is the first to use citizen science data to explore host association patterns in complex systems combining high richness of hosts and their associated species. Our results demonstrate that such data have great potential, although sampling bias is a major issue that needs to be addressed, depending on the hypothesis to be tested. In line with previous studies (e.g. McCune *et al.*, 1997; Nielsen & Bascompte, 2007) we found the data most robust for addressing questions related to community structure (i.e. modularity). We identified host phylogeny to be the most important predictor of modularity in all analyses, but with more resolution within the angiosperms in the full data set. Our procedure to standardize for sampling effort in respect to modularity clearly gives more weight to commonly recorded species in the data set, because rare species are less likely to be picked in a random sample of 100 records. Furthermore, the omission of hosts with < 100 records reduced the coverage of genera in some host clades. In combination, these steps infer that hosts and fungi with few records contribute to a phylogenetically influenced modular structure within the angiosperms only in the full data set, while the data standardization results in modular structures with less detectable phylogenetic signal mainly driven by fungi and hosts with many records. Hence, we consider the results based on the full data set most informative for inferring community patterns.

In the analyses of species richness patterns, we found strongly contrasting effects of host traits, depending on whether we used the raw richness data or standardized data to account for sampling effort. In the full data set, a strong effect of host frequency and time since host establishment on species richness was evident, in line with several previous studies not controlling for potential bias related to sampling effort (e.g. Strong *et al.*, 1984; Newton & Haigh, 1998; Brändle & Brandl, 2001). However, the effect vanished when data were standardized for sampling effort, suggesting that local fungal richness is not higher on widespread and abundant hosts even if these, in theory, might support a larger regional species pool of host specialists. In a somewhat similar study on fungal pathogens on American plants, Miller (2012) standardized for sampling effort by using a citation index, which he found to be the strongest predictor for species richness. However, he still reported a significant positive effect of host range size, in contrast to our study. Nonetheless, this previous study highlights the need to take unequal sampling into account in host richness studies, and also to interpret results of previous studies that do not control for sampling bias in the right context.

Our study clearly supports the view that controlling for sampling effort is important when working with noisy data sets, as shown in previous research (e.g. Isaac *et al.*, 2014). Although data standardization is viable and allows the use of such data, it often reduces the effective size of data sets significantly. An alternative option that should be considered in future citizen science projects is therefore to design protocols that aim to reduce sampling biases

Fig. 2 Fruit bodies of the most commonly recorded species in each of nine fungal orders, which showed either a distinct preference for (or avoidance of) identified host modules (a–d, h, i) or included most species in the data set (f–g). See Fig. 1 for details of host associations. Photographs (a–e) represents orders in the phylum Ascomycota, while photographs (f–i) represents orders in phylum Basidiomycota. All photos © Jens H. Petersen, with permission.

in order to maximize the potential of collected data for answering specific questions (Silvertown, 2009). In projects focusing on species interactions, this could be implemented by using protocols standardizing sampling effort in time or volume per host.

Conclusions

In summary, our study indicates that woody hosts aggregate wood-inhabiting fungi from a large regional species pool containing both host generalists and specialists. Host phylogenies (and inherited traits) act as filters resulting in distinct fungal communities differing among host modules, most notably between angiosperm and gymnosperm hosts. Recent host history, that is, the time since host establishment in Denmark, was not found to influence modularity, indicating that recently introduced hosts are smoothly aggregated into existing host–fungus networks, as also reported for tree pathogens in France (Vacher *et al.*, 2010). Selectivity for modules was greatest for fungal orders rich in endophytes, for example Diaporthales and Xylariales, suggesting that interactions with living host cells are important in driving host selectivity, probably as a result of co-evolutionary processes. In contrast to community composition, species richness was not affected by host phylogeny but only by host size, wood pH and number of species per host genus. Hosts with acidic wood and small size supported species-poor communities compared with larger hosts with higher wood pH, suggesting both factors to act

as general filters limiting the capture of fungi from a large pool of host generalist fungi.

Acknowledgements

We warmly thank the hundreds of volunteers who collected the data used in this analysis and the Aage V. Jensen Naturfond for supporting the Danish Fungal Atlas, and the first author during the writing of this paper. Further, Jens H. Petersen and Jan Vesterholt (deceased) are thanked for their efforts in validating fungal records, including host data, in the Danish Fungal Atlas. Vivian Kvist Johannsen is thanked for information on host frequency from the National Forest Inventory for Denmark, and Camila Donatti is thanked for help with the analysis of phylogenetic signal in modules. P.K.M. thanks São Paulo Research Foundation for support (FAPESP proc. 2015/21457-4). Finally, J.H-C., P.K.M. and B.D. acknowledge the Danish National Research Foundation for funding for the Center for Macroecology, Evolution and Climate, grant no. DNRF96.

Author contributions

B.D., H.H.B., J.H-C. and P.K.M. planned and designed the research. J.H-C., T.G.F. and T.L. processed and validated fungal data. H.H.B. compiled plant trait data. D.D., J.H-C. and P.K.M. analyzed data. All authors contributed to writing of the manuscript, which was led by B.D. and J.H-C.

References

Abrego N, Salcedo I. 2013. Variety of woody debris as the factor influencing wood-inhabiting fungal richness and assemblages: is it a question of quantity or quality? *Forest Ecology and Management* **291**: 377–385.

Bahram M, Harend H, Tedersoo L. 2014. Network perspectives of ectomycorrhizal associations. *Fungal Ecology* **7**: 70–77.

Baldrian P. 2008. Enzymes of saptrotropcic basidiomycetes. In: Boddy L, Frankland JC, van West P, eds. *Ecology of saprotrophic basidiomycetes*. London, UK: Academic Press/Elsevier, 19–42.

Barber MJ. 2007. Modularity and community detection in bipartite networks. *Physical Review E* **76**: 066102.

Barkman JJ. 1958. *Phytosociology and ecology of cryptogamic epiphytes*. Assen, the Netherland: Van Gorcum.

Binder M, Hibbett DS. 2006. Molecular systematics and biological diversification of Boletales. *Mycologia* **98**: 971–983.

Boddy L, Frankland JC, van West P, eds. 2008. *Ecology of saprotrophic basidiomycetes*. London, UK: Academic Press/Elsevier.

Boddy L, Heilmann-Clausen J. 2008. Basidiomycete community development in temperate angiosperm wood. In: Boddy L, Frankland JC, van West P, eds. *Ecology of saprotrophic basidiomycetes*. London, UK: Academic Press/Elsevier, 211–237.

Brändle M, Brandl R. 2001. Species richness of insects and mites on trees: expanding Southwood. *Journal of Animal Ecology* **70**: 491–504.

Brzeziecki B, Kienast F. 1994. Classifying the life-history strategies of trees on the basis of the Grimian model. *Forest Ecology and Management* **69**: 167–187.

Chagnon P-L, U'Ren JM, Miadlikowska J, Lutzoni F, Arnold AE. 2016. Interaction type influences ecological network structure more than local abiotic conditions: evidence from endophytic and endolichenic fungi at a continental scale. *Oecologia* **180**: 181–191.

Chaverri P, Samuels GJ. 2013. Evolution of habitat preference and nutrition mode in a cosmopolitan fungal genus with evidence of interkingdom host jumps and major shifts in ecology. *Evolution* **7**: 2823–2837.

Colwell RK. 2013. *EstimateS, 9.1.0.* [WWW resource, launched 15 June 2013] URL http://viceroy.eeb.uconn.edu/estimates/.

Colwell RK, Chao A, Gotelli NJ, Lin S-Y, Mao CX, Chazdon RL, Longino JT. 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology* **5**: 3–21.

Dalsgaard B, Trøjelsgaard K, Martín González AM, Nogués-Bravo D, Ollerton J, Petanidou T, Sandel B, Schleuning M, Wang Z, Rahbek C *et al.* 2013. Historical climate change influences modularity and nestedness of pollination networks. *Ecography* **36**: 1331–1340.

Danish Mycological Society. 2014. *Danish fungal records database, contributed, edited and validated by Frøslev T, Heilmann-Clausen J, Lange C, Læssøe T, Petersen JH, Søchting U, Stjernegaard T, Vesterholt J.* [WWW document] URL http://www.svampeatlas.dk (accessed 15 May 2014).

Diniz-Filho JAF, Santos T, Rangel TF, Bini LM. 2012. A comparison of metrics for estimating phylogenetic signal under alternative evolutionary models. *Genetics and Molecular Biology* **35**: 673–679.

Donatti CI, Guimaraes PR, Galetti M, Pizo MA, Marquitti F, Dirzo R. 2011. Analysis of a hyper-diverse seed dispersal network: modularity and underlying mechanisms. *Ecology Letters* **14**: 773–781.

Freckleton RP, Harvey PH, Pagel M. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *American Naturalist* **160**: 712–726.

Freschet GT, Weedon JT, Aerts R, van Hal JR, Cornelissen JH. 2012. Interspecific differences in wood decay rates: insights from a new short-term method to study long-term wood decomposition. *Journal of Ecology* **100**: 161–170.

Gange AC, Gange EG, Mohammad AB, Boddy L. 2011. Host shifts in fungi caused by climate change? *Fungal Ecology* **4**: 184–190.

Gilbert GS, Sousa WP. 2002. Host specialization among wood-decay polypore fungi in a Caribbean mangrove forest. *Biotropica* **34**: 396–404.

Guimera R, Amaral LAN. 2005. Functional cartography of complex metabolic networks. *Nature* **433**: 895–900.

Hansen L, Knudsen H, eds. 1992–2000. *Nordic macromycetes, vol. I–III.* Copenhagen, Denmark: Nordsvamp.

Hawksworth DL. 2001. The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycological Research* **105**: 1422–1432.

Heilmann-Clausen J, Aude E, Christensen M. 2005. Cryptogam communities on decaying deciduous wood – does tree species diversity matter? *Biodiversity and Conservation* **14**: 2061–2078.

Heilmann-Clausen J, Christensen M. 2004. Does size matter? On the importance of various dead wood fractions for fungal diversity in Danish beech forests. *Forest Ecology and Management* **201**: 105–117.

Isaac NJB, van Strien AJ, August TA, de Zeeuw MP, Roy DB. 2014. Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution* **5**: 1052–1060.

Jaklitsch WM. 2009. European species of Hypocrea Part I. The green-spored species. *Studies in Mycology* **63**: 1–91.

Jordano P. 1987. Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. *American Naturalist* **129**: 657–677.

Jordano P. 2013. Fruits and frugivory. In: Gallagher RS, ed. *Seeds, the ecology of regeneration in plant communities*, 3rd edn. Wallingford, UK: CABI, 18–61.

Kamiya T, O'Dwyer K, Nakagawa S, Poulin R. 2014. What determines species richness of parasitic organisms? A meta-analysis across animal, plant and fungal hosts. *Biological Reviews* **89**: 123–134.

Knudsen H, Vesterholt J, eds. 2012. *Funga Nordica. Agaricoid, boletoid and cyphelloid genera.* Copenhagen, Denmark: Nordsvamp.

Kollmann FFP, Côté WA Jr. 1968. *Principles of wood science and technology.* Berlin Heidelberg, Germany, New York, NY, USA: Springer-Verlag.

Küffer N, Gillet F, Senn-Irlet B, Aragno M, Job D. 2008. Ecological determinants of fungal diversity on dead wood in European forests. *Fungal Diversity* **30**: 83–95.

Maddison WP, Slatkin M. 1991. Null models for the number of evolutionary steps in a character on a phylogenetic tree. *Evolution* **45**: 1184–1197.

Marquitti FMD, Guimarães PR Jr, Pires MM, Bittencourt LF. 2014. MODULAR: software for the autonomous computation of modularity in large network sets. *Ecography* **37**: 222–225.

Martín González AM, Dalsgaard B, Nogués-Bravo D, Graham CH, Schleuning M, Maruyama PK, Abrahamczyk S, Alarcón R, Araujo AC, Araújo FP *et al.* 2015. The macroecology of phylogenetically structured hummingbird-plant networks. *Global Ecology and Biogeography* **24**: 1212–1224.

McCune B, Dey J, Peck J, Cassell D, Heiman K, Will-Wolf S, Neitlich P. 1997. Repeatability of community data: species richness versus gradient scores in large-scale lichen studies. *Bryologist* **100**: 40–46.

Miller ZJ. 2012. Fungal pathogen species richness: why do some plant species have more pathogens than others? *American Naturalist* **179**: 282–292.

Møller PF, Staun H. 2001. *Danmarks træer og buske*. Copenhagen, Denmark: Politikens Forlag.

Münkemüller T, Lavergne S, Bzeznik B, Dray S, Jombart T, Schiffers K, Thuiller W. 2012. How to measure and test phylogenetic signal. *Methods in Ecology and Evolution* **3**: 743–756.

Neuvonen S, Niemelä P. 1981. Species richness of Macrolepidoptera on Finnish deciduous trees and shrubs. *Oecologia* **51**: 364–370.

Newman MEJ. 2004. Detecting community structure in networks. *The European Physical Journal B-Condensed Matter and Complex Systems* **38**: 321–330.

Newton AC, Haigh JM. 1998. Diversity of ectomycorrhizal fungi in Britain: a test of the species-area relationship, and the role of host specificity. *New Phytologist* **13**: 619–627.

Nielsen A, Bascompte J. 2007. Ecological networks, nestedness and sampling effort. *Journal of Ecology* **95**: 1134–1141.

Ødum S. 1968. Udbredelsen af træer og buske i Danmark. Danmarks Topografisk-Botaniske Undersøgelse nr. 36. *Botanisk Tidsskrift* **64**: 1–118.

Olesen JM, Bascompte J, Dupont YL, Jordano P. 2007. The modularity of pollination networks. *Proceedings of the National Academy of Sciences, USA* **104**: 19891–19896.

Ollerton J, Winfree R, Tarrant S. 2011. How many flowering plants are pollinated by animals? *Oikos* **120**: 321–326.

Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* **401**: 877–884.

Parker IM, Saunders M, Bontrager M, Weitz AP, Hendricks R, Magarey R, Suiter K, Gilbert GS. 2015. Phylogenetic structure and host abundance drive disease pressure in communities. *Nature* **520**: 542–544.

Rezende E, Albert EM, Fortuna MA, Bascompte J. 2009. Compartments in a marine food web associated with phylogeny, body mass, and habitat structure. *Ecology Letters* **12**: 779–788.

Rossman AY, Farr DF, Castlebury LA. 2007. A review of the phylogeny and biology of the Diaporthales. *Mycoscience* **48**: 135–144.

Sáyago R, Lopezaraiza-Mikel M, Quesada M, Álvarez-Añorve MY, Cascante-Marín A, Bastida JM. 2013. Evaluating factors that predict the structure of a commensalistic epiphyte–phorophyte network. *Proceedings of the Royal Society of London B: Biological Sciences* **280**: 20122821.

Sieber TN. 2007. Endophytic fungi in forest trees: are they mutualists? *Fungal Biology Reviews* **21**: 75–89.

Silvertown J. 2009. A new dawn for citizen science. *Trends in Ecology and Evolution* **24**: 467–471.

Smith SE, Read D. 2008. *Mycorrhizal symbiosis*, 3rd edn. London, UK: Academic Press.

Southwood TRE. 1961. The number of species of insect associated with various trees. *Journal of Animal Ecology* **30**: 1–8.

Stokland JN, Siitonen J, Jonsson BG. 2012. *Biodiversity in dead wood*. Cambridge, UK: Cambridge University Press.

Strong DR, Lawton JH, Southwood TRE. 1984. *Insects on plants: community patterns and mechanisms*. Oxford, UK: Blackwell Scientific Publications.

Tedersoo L, Jairus T, Horton BM, Abarenkov K, Suvi T, Saar I, Kõljalg U. 2008. Strong host preference of ectomycorrhizal fungi in a Tasmanian wet sclerophyll forest as revealed by DNA barcoding and taxon-specific primers. *New Phytologist* **180**: 479–490.

Toju H, Guimarães PR, Olesen JM, Thompson JN. 2014. Assembly of complex plant–fungus networks. *Nature Communications* **5**: 5273.

Tulloch AIT, Possingham HP, Joseph LN, Szabo J, Martin TG. 2013. Realising the full potential of citizen science monitoring programs. *Biological Conservation* **165**: 128–138.

Unterseher M, Schnittler M, Dormann C, Sickert A. 2008. Application of species richness estimators for the assessment of fungal diversity. *FEMS Microbiology Letters* **282**: 205–213.

Vacher C, Daudin J-J, Piou D, Desprez-Loustau M-L. 2010. Ecological integration of alien species into a tree–parasitic fungus network. *Biological Invasions* **12**: 3249–3259.

Vacher C, Piou D, Desprez-Loustau M-L. 2008. Architecture of an antagonistic tree/fungus network: the asymmetric influence of past evolutionary history. *PLoS ONE* **3**: e1740.

Vincent JB, Weblen GD, May G. 2016. Host associations and beta diversity of fungal endophyte communities in New Guinea rainforest trees. *Molecular Ecology* **25**: 825–841.

Wagenführ R, Scheiber C. 1989. *Holzatlas*, 3rd edn. Leipzig, Germany: Fachbuchverlag.

Weedon JT, Cornwell WK, Cornelissen JH, Zanne AE, Wirth C, Coomes DA. 2009. Global meta-analysis of wood decomposition rates: a role for trait variation among tree species? *Ecology Letters* **12**: 45–56.

Zabel A, Morell JJ. 1992. *Wood microbiology: decay and its prevention*. San Diego, CA, USA: Academic Press.

Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB *et al.* 2013. Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**: 89–92.

Zhang T, Yao Y-F. 2015. Endophytic fungal communities associated with vascular plants in the high arctic zone are highly diverse and host-plant specific. *PLoS ONE* **10**: e0130051.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Rarefaction curve and Chao1 estimator of species richness for *Fagus*.

Fig. S2 Relationship between the number of fungal orders and the number of host genera recorded per volunteer.

Fig. S3 Diagnostic plots from *pgls* runs and phylograms for host genera.

Table S1 Host genera represented in the data set

Table S2 The number of records submitted by each individual volunteer, and their contributions to total coverage for each fungal order considered and for host genera with > 100 records

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.