NEWS AND VIEWS

COMMENT

Does host plant richness explain diversity of ectomycorrhizal fungi? Re-evaluation of Gao et al. (2013) data sets reveals sampling effects

LEHO TEDERSOO,* MOHAMMAD BAHRAM* and IAN A. DICKIE†‡

*Natural History Museum and Institute of Ecology and Earth Sciences, University of Tartu, 14A Ravila, Tartu 50411, Estonia; †Landcare Research, PO Box 69040, Lincoln 7640, New Zealand; ‡Bio-Protection Research Centre, Lincoln University, PO Box 85084, Lincoln 7647, New Zealand

The generally positive relationship between biodiversity of groups of directly or indirectly interacting organisms is one of the most important ecological concepts (Gaston, 2000 Nature, 405, 220–227; Scherber C, Eisenhauer N, Weisser WW et al., 2010 Nature, 468, 553–556). In a recent issue of Molecular Ecology, Gao C, Shi N-N, Liu Y-X et al. (2013: 22, 3403–3414) reported that the richness of plants and ectomycorrhizal fungi is positively correlated both at local and at global scales. Here, we challenge these findings by re-analysis of data and ascribe the reported results to sampling effect and poor data compilation.

Keywords: biodiversity, chimeric sequences, erroneous statistics, metastudy, sampling design, species richness, taxonomic sampling effect

Received 26 July 2013; revised 3 September 2013; accepted 13 September 2013

In ecological surveys and experiments, sampling effects represent hidden treatments such as the contribution of influential species and spatial or temporal variables that are excluded from analysis, but nonetheless drive other predictors or the response variable directly (Wardle 1999). The influence of a particular species or other taxonomic unit is termed taxonomic sampling effect, which is ubiquitous in natural and experimental systems and often masks the effect of biodiversity per se (Cardinale et al. 2006). Therefore, careful selection of sampling sites and design of experimental units is warranted to avoid or statistically account for the sampling effect (Wardle 1999).

In their study, Gao et al. (2013) established 12 plots for ectomycorrhizal fungi (EcMF) diversity sampling in four levels of host tree richness (2–7 spp.; 2–5 genera). However, each of the four levels comprised predetermined host species – that is, only two particular confamilial species (Castanopsis eyrei and Lithocarpus glaber) were included in the lowest diversity treatment, while a few other fixed species were added to C. eyrei in mid-diversity treatments (Fig. 1). Thus, we argue that the relatively low EcMF richness in lowest host diversity plots may have resulted from poor performance of these tree species or genera, whereas greater EcMF richness in highest diversity plots may stem from the overall positive effect of species or genera that were only included in treatments of medium and high diversity. Ishida et al. (2007) demonstrated that confamilial tree species may substantially differ in richness of associated EcMF. To address this potential taxonomic sampling effect, we tested the effect of presence of each individual plant species to explain the recovered species richness of EcMF. One-way ANOVAs revealed that the presence of three tree species may have resulted in the reported richness effect (Pinus massoniana and Quercus serrata: \( F_{1,10} = 14.3, P = 0.004 \); Cyclobalanopsis glauca: \( F_{1,10} = 10.2, P = 0.009 \)). By using linear regression analyses, we also tested whether the relative contribution of each host species affected richness of EcMF. Based on log-ratio-transformed basal area of each host species (Table S1, Supporting Information; data kindly provided by authors; absent species were given a value of \( 10^{-5} \) m\(^2\)/ha that is an order of magnitude less than the least abundant species across all plots), contribution of P. massoniana and Q. serrata, respectively, explained 54.9% (\( F_{1,10} = 12.2; P = 0.006 \)) and 61.5% (\( F_{1,10} = 16.0; P = 0.003 \)) in EcMF richness in separate regression analyses. These analyses provide support to our concern that the presence or relative importance of particular host species or host genera (both genera were monospecific) may have driven the ‘host richness’ effect.

In addition to the issue with study design and statistics, we note that 57 (34.4%) and 17 (10.2%) sequences of the 66 deposited sequences were suspected of being chimeric (automatically detected using several chimera checkers and manually verified; Table S2, Supporting Information) and belonging to non-EcMF, respectively. However, this is a separate issue from sampling effects.

In their global metastudy, Gao et al. (2013) included host taxonomic richness information from 100 previously published studies to explain fungal richness. Evaluation of their data set and methods revealed multiple issues of concern in data compilation. First, it is not stated whether the authors used the number of plant species present or the number actually sampled. The values provided in their supplement revealed that both forms were used haphazardly, but the number of actually sampled hosts prevailed. Second, the number of EcMF species differed between the original...
The presence of tree species (columns) in plots (rows) of Gao et al. (2013). Different tree symbols represent different species. The Roman numerals in parentheses indicate tree diversity treatments I-IV according to Gao et al. (2013). reports and the compiled data set in 10% of records (in three cases more, in seven cases fewer species). This was not due to weeding out nonectomycorrhizal fungal species (as opposed to a similar data set in Tedersoo et al. 2012) but rather appears to reflect errors in compilation. Third, the data set included studies addressing ECNM diversity of pot‐grown, planted and naturally established seedlings as well as mature trees with no attempt to account for age differences. Age of vegetation is an important determinant of ECNM biodiversity at both local and global scales (Wallander et al. 2010; Tedersoo et al. 2012). Fourth, several studies were based on sources other than ectomycorrhizal root tips as claimed in their methods: fruit bodies (I) or root samples of a mycoheterotrophic Ericaceae species (I). Fifth, many distantly located sites were pooled into a single study (14), although 10‐km‐distance criterion was clearly stated. Sixth, many sites that comprised different host plant communities in terms of host species composition and richness were pooled (10). Among these, monospecific plantations of different hosts were pooled to represent a single study with multiple hosts (2). Conversely, on a few occasions, the authors used the same study twice (4) or reported data from the same multihost site many times referring to different monospecific communities (eight studies and three sites). These types of errors were so prevalent (more than 48% of the data rows contained one or more errors) as to undermine the validity of any subsequent analysis (Table S3, Supporting Information).

We also argue that, besides host richness, the authors ignored other important determinants such as sampling effort variables (e.g. sample volume, number of samples) that explained a large proportion of variation in a previous meta‐study, in which no residual effect of sampled and total host richness was found (Tedersoo et al. 2012). To account for differential sampling effort, we supplemented the data set with metadata about the number of samples and the volume of substrate examined. We contacted the original authors for information about the size of root samples or seedlings and age of vegetation if these data were missing in the publications. If there was no response, we estimated each seedling sample to be the equivalent of $10^{-3}$ m$^3$ of soil. This additional metadata revealed that in the compiled data set of Gao et al., sampling effort among studies varied two to four orders of magnitude. In particular, the number of samples and sampled soil volume ranged from 7 to 648 units and $2 \times 10^{-4}$ to 1.248 m$^3$, respectively. We corrected the global data set for obvious errors and excluded studies that (i) sampled fruit bodies and mycoheterotrophic plants; (ii) sampled predominately exotic trees in their non‐native range; (iii) comprised <15 actual samples; (iv) were treated twice; (v) constituted floristically different sites >10 km distant; vi) pooled monospecific plantations under the same ‘multihost’ study (Table S3, Supporting Information). When the data were sufficient and the above criteria were met, we separated certain studies by multiple sites or pooled several studies carried out at the same site. We re‐analysed the corrected data set of 75 sites considering tropical and temperate sites both separately or combined (following Gao et al. 2013). Age of vegetation (only seedlings <10 year old vs. older plants), sampling variables including their square‐root‐transformed values and host taxonomic richness at species, genus and family levels were used as predictors of ECNM species richness in general least‐squares (GLS) model selection and corrected Akaike Information criterion (AICc) following Tedersoo et al. (2012). Because seedlings tended to represent monospecific communities and they had significantly lower richness compared to other plants ($t = -2.58; P = 0.012$), we further excluded sites represented only by seedlings and re‐run the analysis using data from 67 sites.

Model selection revealed that host richness at any taxonomic level had no significant effect on ECNM species richness. In the best models, sampling variables explained a substantial proportion of the variation in ECNM richness in temperate ($n = 56; R^2_{adj} = 0.352$; sample volume: $t = -3.24, P = 0.002$; sample volume$^{0.5}: t = 3.69, P < 0.001$; number of samples: $t = -2.03, P = 0.047$; number of samples$^{0.5}: t = 2.41, P = 0.020$) and tropical ($n = 11; R^2_{adj} = 0.702$; sample volume: $t = -4.94, P = 0.002$; sample volume$^{0.5}: t = 4.96, P = 0.002$; number of samples: $t = 3.04, P = 0.019$) ecosystems and both combined ($n = 67; R^2_{adj} = 0.175$; sample volume: $t = -2.69, P = 0.009$; sample volume$^{0.5}: t = 3.11, P = 0.003$; number of samples: $t = -2.15, P = 0.035$; number of samples$^{0.5}: t = 2.44, P = 0.018$).
our re-analysis provides no evidence for the host taxonomic richness effect at local and global scales and implies that the inferred strong positive effect stems from sampling effect at the local scale and both sampling effect and incorrect data compilation and analysis at the global scale. We suggest that greater care must be taken to correct for sampling variables in comparing results across studies, either using these variables as covariates (Tedersoo et al. 2012) or using species richness extrapolation or rarefaction analysis (Taylor 2002; Dickie 2007). The latter methods, however, require the availability of species by sample distribution data and that samples are of comparable size. Minimum richness extrapolation further requires the estimator curve to stabilize, which is rarely the case even in most deeply sampled EcMF communities.

Across the world, climatic factors, disturbance and age of communities have much stronger effect compared with richness of hosts sampled and those present (Tedersoo et al. 2012). Because of insufficient data, global-scale analyses may also suffer from taxonomic sampling effect, because certain hosts such as *Gnetum* spp., Nyctaginaceae spp. and *Alnus* spp. exhibit strong specificity to certain mycobionts, and therefore, these plants associate with fewer EcMF species than other plant taxa in nearby habitats (Bechem & Alexander 2012; Hayward & Horton 2012; Pöhlme et al. 2013). In mixed plant communities, these taxa may also drive the view on host specificity, EcMF richness, etc., if these particular hosts are targeted in EcMF diversity studies.

Despite the issues we raise with the Gao et al.’s analysis, the key question remains: Does host plant richness or diversity per se influence EcMF richness? Based on other systems of interacting organisms aboveground and belowground, positive effects are expected, because specificity generates ecological niches (Cardinale et al. 2006; Dickie 2007). Plant species affect EcMF directly and indirectly through modification of soil properties, suggesting the importance of both genetic/physiological and environmental mechanisms for niche development (Morris et al. 2008). In general, communities of EcMF are related to communities of their hosts (Bahram et al. 2012). Whether EcMF communities along a resource or host diversity gradient differ in richness probably depends on the level of stress in the extremities of the gradient as well as the size and accessibility of a species pool adapted to these environments. Soil variables may affect both EcMF and host plant species composition, and therefore, soil effects could be also taken into consideration by using, for example, structural equation modelling. Such integrated analyses are thus far lacking in EcMF ecology (but see Antoninka et al. 2009 for arbuscular mycorrhiza).

Besides the identity of interacting host, neighbourhood of other potential host plants and nonhost plants may also affect EcMF in focal plants (Haskins & Gehring 2004; Jairus et al. 2011; Bogar & Kennedy 2013). Due to priority effects and competition among fungi, the presence of other hosts may hamper the development of typical EcMF communities and result in impoverished richness on roots of focal plants (Bogar & Kennedy 2013). This suggests that in mixed plant communities, competition between mycobionts of different plant hosts and allelopathy may have negative effects on overall EcMF richness. Such a detrimental neighbour effect is less expected when plant species grow in aggregated clumps, but in this scenario dispersal limitation may also negatively affect EcMF species richness (Peay et al. 2007). Whether the negative effects from competition and allelopathy or positive effects from facilitation and niche differentiation prevail (for simulation modelling, see Vellend 2008) remains to be studied. Because of the occurrence of such a wide variety of interactions between plants and potentially between fungi, the effect of mixing plant species on fungal community is probably prone to taxonomic sampling effect.

Whether or not richness of host plants affects local EcMF species richness remains an open question, which needs to be carefully determined by accounting for sampling effect and other potentially important variables (see also Kernaghan et al. 2003). The neighbouring tree effect on focal plant communities warrants further research to improve our understanding of plant mixture effects on soil processes and competition among microorganisms belowground. While these studies have utilized traditional Sanger sequencing, modern high-throughput sequencing methods provide a powerful tool to address diversity relationships among macro- and microorganisms at an unprecedented sampling depth. Although these methods suffer from various biases, artefacts and analytical errors (Dickie 2010; Tedersoo et al. 2010), careful sampling design coupled with bioinformatics and statistical analysis may allow these key ecological questions in microbial ecology to be addressed.

**Acknowledgements**

We thank the authors of many studies for providing data on sampling effort where it was not originally reported. Thanks to R.H. Nilsson for implementing semiautomatic chimera checking in PlutoF workbench (http://unite.ut.ee/temp/plutoF2/). We thank the reviewers, including for noting the effect of soil variables on plant host as well as EcMF communities. LT and MB receive financial support from New Zealand’s Ministry of Business, Innovation and Employment’s Science and Innovation Group.

**References**


© 2014 John Wiley & Sons Ltd


L.T. and I.A.D. checked data; M.B. analysed data; L.T. wrote the paper with discussion, editing and contributions from M.B. and I.A.D.

doi: 10.1111/mec.12660

Data accessibility

All data are available in Tables S1 and S3 (Supporting Information).

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Basal area contribution of ectomycorrhizal trees in study sites.

Table S2 Suspected chimeric and nonectomycorrhizal sequences.

Table S3 Corrected metadata of study sites.